Inferring the phylogeny of problematic metazoan taxa using mitogenomic and phylogenomic data

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt am

Fachbereich Biologie/ Chemie

der Universität Osnabrück

von

Anja Golombek

Osnabrück 2018

“Wenn Arten aus anderen Arten durch unmerkbare kleine Abstufungen entstanden sind, warum sehen wir nicht überall unzählige Übergangsformen? Warum bietet nicht die ganze Natur ein Gewirr von Formen dar, statt daß die Arten, wie sie sich uns zeigen, wohl begrenzt sind?“

Charles Darwin

Table of contents

1. Abstract

In the past quarter of the century the understanding of evolutionary relationships between major metazoan taxa has changed. Morphology-based approaches had many successes but important aspects of the evolutionary tree often remained unclear. Phylogenetic analyses based on molecular data give implications about evolutionary relationships and events concerning metazoan taxa.

This thesis had a major contribution to clarify phylogenetic relationships within problematic metazoan taxa: Deuterostomia and Spiralia (especially Annelida and Platyzoa).

The phylogeny of platyzoan taxa is hard to define based on morphological characters. However, phylogenetic analyses based on phylogenomic data recovered Platyzoa as non-monophyletic and recoverd platyzoan taxa as paraphyletic with respect to other spiralians. This platyzoan paraphyly suggests that the last common ancestor of Spiralia was probably a simple-bodies organism and that more complex animals such as annelids evolved from such a simple organized ancestor.

Within annelids numerous meiobenthic annelid species exist and part of them inhabits the interstitium. Due to their apparently simple organization these interstitial annelid species (formerly called “Archiannelida”) have been regarded to be very close to the annelid stem species due to lack of many annelid-specific characters including absence of parapodia and chaetae in certain taxa. Phylogenetic analyses based on phylogenomic data suggest that interstitial annelid taxa are not part of the basal radiation of Annelida and inhabiting the interstitial realm and a simple body organization are not ancestral characters of Annelida.

During the last years mitogenomic data became more and more important as well resolving questions concerning metazoan phylogeny on various levels. Phylogenetic analyses based on mitochondrial data are often criticized because the resulting phylogenetic trees are often characterized by problematic long branches, lack of resolution at the base of the tree, or there is support for less reliable sister group.
These results of different investigations in this thesis based on mitogenomic data suggest that mitochondrial genomes have some value in phylogenetic analyses.

Phylogenetic analyses based on mitogenomic data recovered a robust phylogeny of deuterostomes with unequivocally support for the Notochordata hypothesis. Even the phylogenetic position of *Xenoturbella* could be resolved as sister to all deuterostomes (except for tunicates).

To resolve the phylogenetic affinities of the platyzoan taxon Gnathostomulida with respect to Syndermata, Platyhelminthes and Gastrotricha phylogenetic analyses of mitochondrial data support a sister group relationship of Gnathostomulida to Syndermata (Gnathifera hypothesis). However, the mitochondrial gene order data do neither support nor reject a closer relationship of Gnathostomulida to Syndermata, Gastrotricha, or Platyhelminthes. These observations give new insides into the high value of mitochondrial gene order as a phylogenetic marker.

Conflict in signal is a potential problem for tree reconstruction. Taxa affected by such biases are often misplaced in phylogenetic analyses.

The usage of mitogenomic data to resolve the phylogeny of terebelliform annelids recovered that the terebelliform taxon Trichobranchidae was caught in the symplesiomorphy trap. The investigations revealed that mitochondrial data of Alvinellidae and Ampharetidae exhibited strong compositional biases causing the misplacement of Trichobranchidae of Trichobranchidae and Terebellidae.

Furthermore, the phylogenetic position of the problematic taxon Diurodrilidae could be resolved using mitogenomic data. The mitochondrial genome sequence data was able to rescue Diurodrilidae from the placing with Platyzoa due to long-branch artifacts, and placed them within Annelida, instead.

The investigation and comparison of mitochondrial gene arrangements is a useful tool to resolve more issues concerning annelid evolution and phylogenetic relationships.

In annelids the mitochondrial gene order is highly conserved. However, investigations concerning the evolution of mitochondrial gene order of Syllidae and basal branching annelids such as Oweniidae, Magelonidae, Amphinomidae, and Chaetopteridae revealed that the mitochondrial gene order of these taxa is substantially different from the assumed annelid ground pattern which thus seems to be restricted to Pleistoannelida.

2. Introduction

The evolutionary origin as well as the phylogeny of higher metazoan taxa is still under debate although considerable progress has been made in the past 20 years. Metazoa (multicellular animals) represents a monophyletic group of highly diverse animals, which includes Bilateria, Cnidaria, Porifera, Ctenophores, and Placozoa.

Bilateria, the triploblastic animals, comprises the majority of metazoans and consists of three major clades: Deuterostomia, Spiralia (= Lophotrochozoa sensu lato), and Ecdysozoa (Halanych, 2004; Edgecombe et al., 2011), whereas the sister group taxa Spiralia and Ecdyzozoa form the monophyletic clade Protostomia (de Rosa et al., 1999; Halanych et al., 1995; Giribet et al., 2000; Mallatt and Winchell, 2002;). The term Spiralia is solely a synonym for Lophotrochozoa (Halanych, 2004). The term Lophotrochozoa describes the descendants of the last common ancestor of Mollusca and Annelida as well as the three lophophorate taxa (Brachiopoda, Phoronida, and Bryozoa) (Halanych, 2004), whereas the more comprehensive term Spiralia includes all animals with spiral cleavage (Edgecombe et al., 2011).

Spiralia includes a highly diverse number of animal phyla: Annelida, Brachiopoda, Bryozoa (= Ectoprocta), Cycliophora, Dicyemida, Entoprocta (= Kamptozoa), Gastrotricha, Gnathostomulida, Micrognathozoa, Mollusca, Nemertea, Orthonectida, Phoronida, Platyhelminthes, Syndermata, and possibly Chaetognatha (Kocot, 2016, and references therein). In 1995, Halanych et al. were the first to propose the concept of Lophotrochozoa. Based on 18S rDNA data they defined this taxon as the last common ancestor of annelids, mollusks, the three traditional lophophorate taxa (Brachiopoda, Phoronida, and Bryozoa), and all the descendants of that common ancestor. Later, this taxon has been expended including also groups such as Gastrotricha, Platyhelminthes, and Gnathostomulida (Aguinaldo et al., 1997; Kocot, 2016, and references therein). This taxon includes all phyla formerly united as Spiralia and thus several authors prefer the usage of Spiralia instead of Lophotrochozoa sensu lato (Hejnol et al., 2010; Dunn et al., 2014, Struck et al., 2015; Laumer et al., 2015).

The monophyly of Spiralia has been confirmed by numerous phylogenetic investigations based on different molecular data (e.g., Dunn et al., 2008; Halanych, 2004; Hausdorf et al., 2007; Nesnidal et al., 2013; Passamaneck and Halanych, 2006; Struck and Fisse, 2008; Struck et al., 2014). Spiralian taxa show a high diversity and plasticity in body plans as well as developmental, embryonic and morphological characters (Nielsen, 2012). Based on these

hypotheses, there are two major spiralian taxa: Platyzoa (see Edgecombe et al., 2011) and Lophotrochozoa (sensu stricto; Halanych, 2004; Hankeln et al., 2014) (= Trochozoa sensu Giribet et al., 2000; Roule, 1891; Rouse, 1999). Lophotrochozoa comprises at least annelids, the lophophorate taxa (Brachiopoda, Phoronida, and Bryozoa), mollusks, and those spiralian animals with a more complex morphology including a coelom (Halanych, 2004). In contrast Platyzoa comprises more simple appearing taxa such as Platyhelminthes (flatworms), Gastrotricha (hairy-backs), Rotifera (wheel animals), Acanthocephala (thorny-headed worms), and Gnathostomulida (jaw worms) (Cavalier-Smith, 1998). However, statistical support for Platyzoa was weak and their monophyly appeared uncertain.

Given the still disputed basal position of Xenacoelomorpha, the three clades Deuterostomia, Spiralia, and Ecdysozoa are sometimes united as Nephrozoa being the sister group of Xenacoelomorpha as highest ranked sister groups in the phylogenetic tree of Bilateria (Hejnol et al., 2009; Cannon et al., 2016). The name Nephrozoa refers to their probable morphological apomorphy, the existence of excretory organs, the nephridia. However, the position of *Xenoturbella* (Xenoturbellida) and Acoela (Acoelomorpha) is still questioned (Hejnol et al., 2009; Philippe et al., 2011b; Dunn et al., 2014).

Molecular data have profoundly changed the view of the bilaterian tree of life. One of the main questions concerning bilaterian phylogeny is still the on-going debate about the evolution of complexity in Bilateria. It was assumed that the last common ancestor of Deuterostomia, Ecdysozoa and Spiralia had a segmented and coelomate body organization resembling that of an annelid. Following this hypothesis, morphologically more simply organized taxa like Platyhelminthes or Nematoda must have been evolved by reductions (e.g., Brinkman and Philippe, 2008; Couso, 2009; De Robertis, 2008; Tomer et al., 2010). On the contrary, the traditional or morphologically based view is the evolution of Bilateria from a simple body organization towards more complex forms. In this hypothesis the last common ancestor of Bilateria resembles a platyhelminth-like animal without coelomic cavities and segmentation ("acoeloid-planuloid" hypothesis) (Hyman, 1951; Halanych, 2004; Hejnol et al., 2009). To resolve this question, it is necessary to unravel the phylogenetic relationships within Bilateria (Halanych, 2004; Edgecombe et al., 2011).

This thesis had a major contribution to clarify phylogenetic relationships within deuterostomes and Spiralia (especially annelid taxa) as well as to clarify the questionable monophyly of Platyzoa. Whereas, it is generally accepted that annelids are part of Spiralia and Lophotrochozoa (Halanych, 2004; Kocot, 2016, and references therein), there are still

numerous phylogenetic questions concerning the position of certain groups within Annelida such as interstitial polychaete taxa (e.g., Diurodrilidae, Dinophilidae, Protodrilidae, Polygordiidae, Saccocirridae, etc.), formerly united as “Archiannelida” (e.g., Westheide, 1985; 1987). Numerous analyses concerning questions of bilaterian phylogeny are based on phylogenomic data (e.g., Dunn et al, 2008; Hausdorf et al., 2007; Hejnol et al., 2009; Nesnidal et al., 2010; Paps et al., 2009a; Struck et al., 2011). However, during the last few years mitogenomic data became more and more important as well resolving questions concerning bilaterian phylogeny on various levels (see Wägele and Bartolomaeus, 2014; and references therein).

3. Mitochondrial genomes as phylogenetic marker(s)

In the past decades (since the 1980's) the amount of molecular data (phylogenomic data) used in phylogenetic reconstructions has been steadily increased (e.g., Dunn et al., 2008; Dordel et al., 2010; Giribet et al., 2000; Hausdorf et al., 2007; Hejnol et al., 2009; Philippe et al., 2009, 2011a, 2011b; Struck et al., 2011, 2015). In comparison to phylogenomic datasets, mitochondrial genome data (mitogenomics) allow a larger taxon sampling, whereas mitogenomic data include a much smaller amount of sequence information. Also, the importance of other characters such as gene order has been considered for phylogenetic evaluation, because closely related species often share identical unchanged gene orders and gene rearrangements rarely occur independently in different lineages (Boore, 1999; Boore and Brown, 1994). Since the 2000's, the amount of analyses based on mitogenomic data increased and the value of mitochondrial sequence data as phylogenetic marker was/is under controverse discussions (e.g., Galtier et al., 2009; Rubinoff and Holland, 2005; and references theirein). The sequences of complete mitochondrial genomes can be compared on nucleotide and amino-acid level. Especially mitochondrial protein coding sequence data show valuable phylogenetic signal (e.g., Bourlat et al., 2008).

Complete mitochondrial genomes with respect to genome architecture, gene content/arrangement, exon/intron structure, genetic code and secondary structure of rRNA and tRNA genes became more and more import to infer phylogenetic relationships of different taxa ranging from population to phylum level (e.g., Avise, 2000; Boore and Brown, 1998; Boore, 2006; Vallès and Boore, 2006; Wägele and Bartolomaeus, 2014).

Mitochondrial genomes have been determined in all metazoan taxa and contain a conserved complement of genes ("housekeeping genes"), with exception of the animal phylum Loricifera which shows a lack of mitochondria (Danovaro et al., 2010).

Mitochondrial genomes are small (15,000 bp – 20,000 bp), extrachromosomal molecules, which are nearly circular (Fig.1), with exception of some sponges (Lavrov et al., 2013), and cnidarian classes (Cubozoa, Skyphozoa and Hydrozoa), which have linear mtDNA chromosomes (Bridge et al., 1992). Mitochondrial genomes typically have only 37 genes ("housekeeping genes"), which encode 13 protein-coding subunits (*cox1-3*, *nad1-6*, *nad4L*, *atp6*, *atp8*, and *cytb*) for components of the respiratory chain, two ribosomal RNAs (*rrnS* and *rrnL*) of the mitochondrial ribosome, and a set of 22 – 24 tRNAs which are necessary for the translation of the proteins encoded by mitochondrial DNA (Wolstenholme, 1992; Boore, 1999;

Vallès and Boore, 2006) (Fig.1). There are some exceptional taxa such as Nematoda, some bivalves, and some cnidarians which show a different gene content (e.g., Boore, 1999, 2006; Mwinyi et al., 2010; Osca et al., 2014; Papillon et al., 2004; Steinauer et al., 2005; Valles and Boore, 2006).

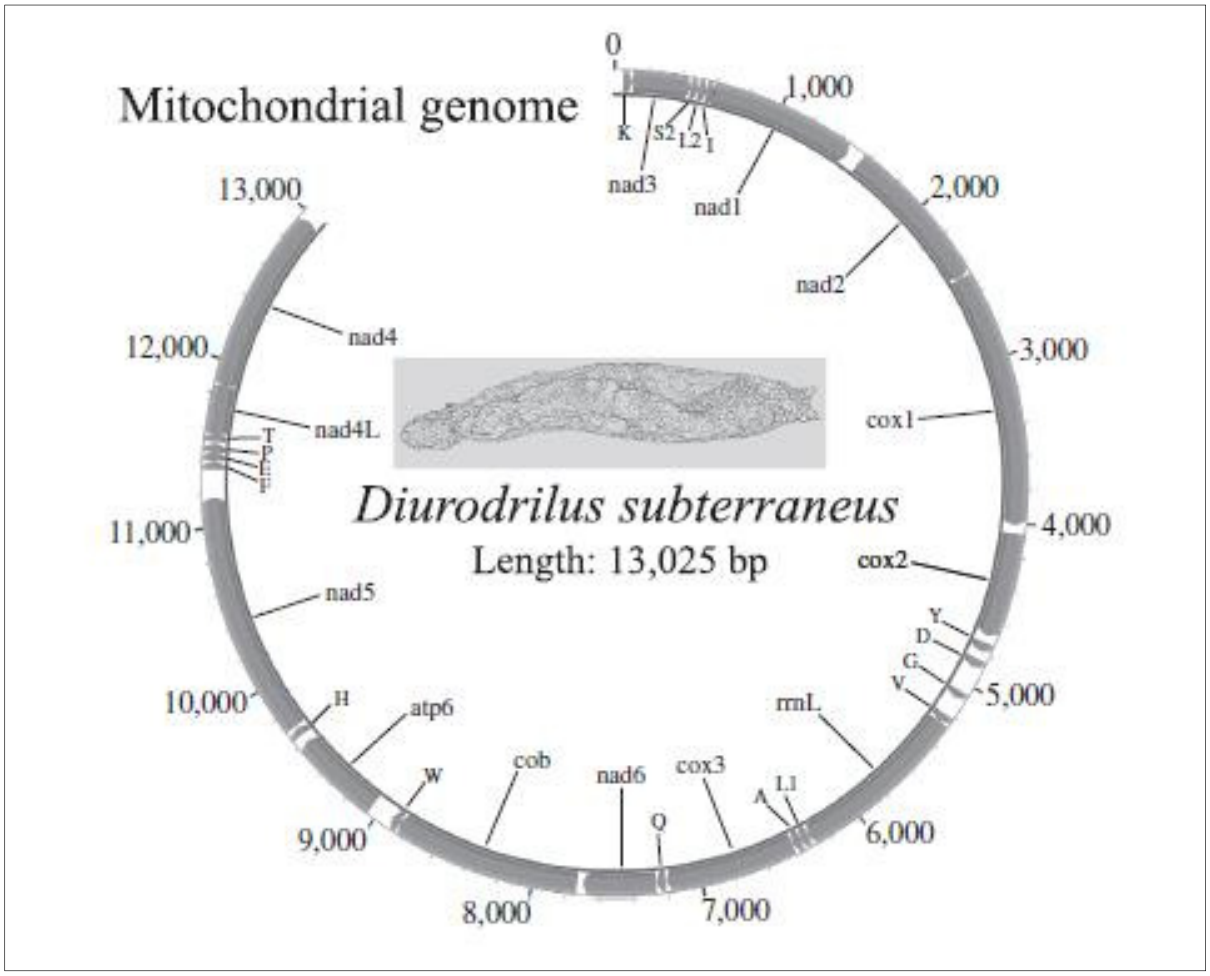


Fig.1: The circular structure and the gene content of mitochonadrial genomes using the example of the mitochondrial genome of *Diurodrilus subterraneus* (from Golombek et al., 2013). The shown gene order is not general for metazoan taxa. nad1-6, 4L = NADH dehydrogenase subunits 1-6 and 4L; cox1-3 = cytochrome oxidase subunits 1-3; cob = cytochrome b; atp6 = ATP synthase subunit; rrnS = small ribosomal RNA (also known as 12S); rrnL = large ribosomal subunit (also known as 16S); single letters = mitochondrial tRNAs coding for alanine (A), aspartic acid (D), glutamic acid €, phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L1 (CUA), L2(UUA)), proline (P), glutamine (Q), serine (S2(UCA)), threonine (T), valine (V), tryptophane (W), tyrosine (Y).

In certain taxa much larger mitochondrial genomes have been found, but these are the product of duplications of part of the mtDNA rather than a variation in gene content (see Boore, 1999 and references therein). The variations in gene content are rarely due to changes in the complement of protein-coding genes. Mostly, changes in the content of tRNA genes are observed (e.g., Golombek et al., 2015). Furthermore, animal mitochondrial genomes possess

only limited intergenic sequences aside from one large non-coding region, which contains controlling elements for replication and transcription (Clary and Wolstenholme, 1984; Shadel and Clayton, 1997).

Phylogenetic analyses based on mitochondrial data are often criticized because the resulting phylogenetic trees are often characterized by problematic long branches (e.g., Nematoda, Platyhelminthes), or there is support for less reliable sister group relationships (e.g., Hassanin et al., 2005; Steinauer at al., 2005; Yokobori et al., 2008; Mwinyi et al., 2010). Despite these problems mitochondrial genomes were successfully used at different levels of phylogenetic reconstructions using protein sequence analyses, gene order rearrangements, and features of the genetic code and secondary structure of tRNA's (e.g., Bleidorn et al., 2007; Perseke et al., 2007; Stach et al., 2010).

3.1. Criteria of molecular marker

The suitability of mitochondrial genomes as molecular markers for phylogenetic analyses can be evaluated based on a set of specific criteria (Cruickshank, 2000). **(I)** At first, the orthology criterion, which should be fulfilled. That means, speciation events cause changes between gene sequences and not gene duplication events (Altenhoff and Dessimoz, 2012). **(II)** Second, the relevant genes should be present in all taxa under study, also known as "housekeeping genes". **(III)** Third, selection should only be a stabilizing factor on marker genes to avoid that phylogenetic signal is obscured by positive or negative selection. **(IV)** Fourth, character state frequencies (nucleotides or amino acids) and substitution rates of molecular markers are constant. **(V)** Fifth, conserved and variable parts have to be present in the alignment. The conserved parts are important for the construction of PCR primer sets useful for many species. The variable sites provide an adequate amount of phylogenetic signal.

Mitochondrial genomes fulfill most of these criteria. However, a problem is the detection of non-functional nuclear copies of mitochondrial sequences (numts), whose identification appears to be problematic (Bensasson et al., 2001; Bernt et al., 2013a).

4. Results

The following articles are all part of this cumulative thesis. In this chapter each article is summarized showing the phylogenetic background, the main results and conclusions. The complete published articles are included in chapter 5. Published articles. Additionally, the supplementary data of each study (if available) is added on CD which you can find at the end of the thesis.

4.1. "A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny" (Bernt et al., 2013b)

The state-of-the-art of animal phylogenetics was the analysis of large multilocus datasets, derived from whole genomes or large-scale EST approaches, also called "phylogenomics" (e.g., Dunn et al., 2008; Hausdorf et al., 2007; Hejnol et al., 2009; Philippe et al., 2009; Pick et al, 2010). In these analyses the "new animal phylogeny" was largely confirmed: Bilateria are subdivided into Protostomia and Deuterostomia, the former being subdivided in Spiralia (Lophotrochozoa sensu lato) and Ecdysozoa. However, the internal phylogeny of Spiralia, Ecdysozoa, and Deuterostomia, as well as the basal relationships between the non-bilaterian taxa and Bilateria were still inconsistent comparing different phylogenomic studies (e.g., Hejnol et al., 2009; Nosenko et al., 2013; Philippe et al., 2009; Philippe et al., 2011a; Pick et al., 2010).

A general problem in these large data sets with many taxa is missing or incomplete data or sequence data which has no phylogenetic signal to resolve issues concerning bilaterian phylogeny at all. In the last 20 years, the applicability of mitochondrial genome sequence data as phylogenetic marker was shown in many phylogenetic analyses concerning evaluation of phylogenetic relationships of different taxa ranging from population to phylum level (e.g., Avise, 2000; Boore et al., 1995; Golombek et al., 2013; Mwinyi et al., 2010; Perseke et al, 2013; Stach, Braband and Podsiadlowski, 2010; Wägele and Bartolomaeus, 2014; Wey-Fabrizius et al., 2013; Zhong et al., 2011).

However, deep metazoan phylogeny based on mitochondrial data was often under debate because of high substitution rates of nucleotides, large differences in amino acid substitution

rates between taxa, and biases in nucleotide frequencies (Hassanin et al., 2005; Perna and Kocher, 1995).

The present study is the most comprehensive analysis of bilaterian phylogeny based on mitochondrial genome data representing a profound sample of the phylogenetic as well as sequence diversity. The analyzed data set comprises more than 650 mitochondrial genomes mostly comprising invertebrate species as well as a selection of a few vertebrate species, and outgroup taxa from fungi and protists (see supplementary material). The analyses are based on metazoan mitochondrial genome sequences in RefSeq (Pruitt et al., 2007). Additionally, the new optimized automated annotation pipeline MITOS was used to re-annotate all sequences to overcome annotation errors which are known to appear in NCBI RefSeq entries of mitochondrial genomes (Bernt et al., 2013c). The phylogenetic reconstruction is exclusively based on protein coding genes on amino acid level.

The analysis reveals unbalanced phylogenetic trees with large differences in branch lengths and a lack of supported resolution for most basal nodes (Fig. 2). The limited taxon sampling of mitochondrial genomes for especially these basal groups (e.g., Hydrozoa, Scyphozoa, Hexactinellida) clearly biases the analyses in this part of the tree. Furthermore, there is an increase of branch lengths among bilaterian taxa compared to non-bilaterian and outgroup taxa (Fig. 2). The phylogenetic analysis shows support (high bootstrap values) for traditional phyla of the lophotrochozan part of the tree, e.g., Brachiopoda, Nemertea, Annelida (including Sipuncula and Echiura), Entoprocta, and Bryozoa, but the results failed to give support for undisputed high-ranking taxa like Mollusca, Insecta, and Arthropoda (Fig. 2). Furthermore, extreme long branches are observed for Nematoda, Platyhelminthes, Syndermata, and some other taxa (Fig. 2). However, the phylogeny of the deuterostome clade in this analysis is largely congruent with phylogenetic analyses based on nuclear genome data sets, except for the position of tunicates which show the longest branch among deuterostomes (Fig. 2).

In conclusion, the phylogenetic analysis of mitochondrial genome data of Bilateria lacks resolution at the base of the tree as well as for ecdysozoan and spiralian interrelationships. However, some parts of the tree show reasonable branching pattern and good bootstrap support for even deep splits such as the monophyly of deuterostomes. These results suggest that mitochondrial genomes have some value in phylogenetic analyses, even if there are no extreme differences with respect to gene order, nucleotide frequency, or strand bias among the studied taxa.

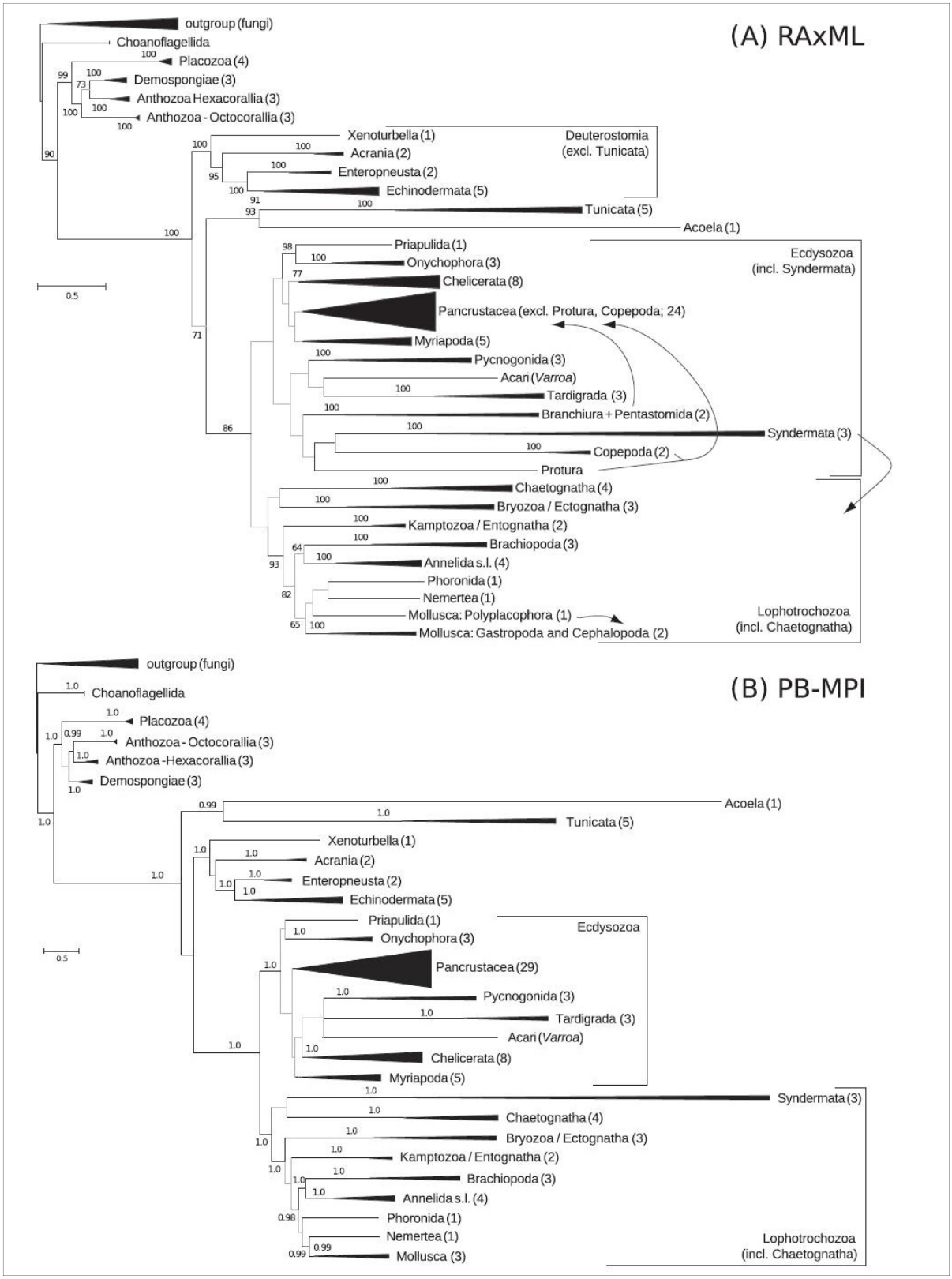


Fig. 2: Phylogenetic tree obtained with an alignment of amino acid sequences from the dataset reduced to 100 taxa (from Bernt et al., 2013b). (A) Best tree from RAxML analysis with bootstrap percentages (>50%). Differences to the tree shown in subfigure B are highlighted by arrows. (B) Consensus tree from six independent chains of PhyloBayes-MPI. Bayesian posterior probabilities are given when >0.95. In both trees the numbers in brackets after each taxon name refer to the number of species representing this taxon. The light gray parts of the trees are insufficiently supported parts.

To identify the sources of misleading phylogenetic signals it is necessary to understand and to consider the influence of nucleotide abundance, strand bias, and mitochondrial substitution rates between metazoan taxa on phylogenetic reconstructions. There is a correlation of gene rearrangements and substitution rates promoting long branches and problems in sequence-based analyses (Xu et al., 2006).

4.2. "The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny" (Perseke et al., 2013)

Deuterostomia are one of the three major lineages of Bilateria and are the sister group of Protostomia comprising Spiralia (Lophotrochozoa sensu lato) and Ecdysozoa (Eernisse and Peterson, 2004; Philippe et al., 2005; Telford et al., 2005). Within monophyletic Deuterostomia, there are diverse morphological forms ranging from solitary to colonial body plans accompanied by free-living or sessile life styles. Deuterostome taxa are recognized based on shared embryonic developmental patterns such as radial cleavage, gastrulation occurring at the vegetal pole, blastopore becoming the anus and the mouth being a secondary formation (deuterostomy), as well as the formation of the coelom by enterocoely (summarized in Swalla, 2006).

Traditionally, three monophyletic groups are recognized: Chordata comprising Craniota (Vertebrata), Cephalochordata (Acrania) and Tunicata (Urochordata); Hemichordata consisting of Enteropneusta and Pterobranchia; and Echinodermata. The phylogenetic relationships of these taxa to each other have been hardly discussed for a long time.

The traditional view of chordate phylogeny suggests Cephalochordata as sister group to Craniota because of the possession of a notochord, and several other characters (Notochordata hypothesis) (Garstang, 1928; Stach, 2008). Another hypothesis suggests a closer relationship of Tunicata and Craniota by the discovery of a migratory cell population in both taxa, which is missing in Cephalochordata (Olfactores hypothesis) (Stach, 2008 and references therein). Whereas, molecular data could not resolve this issue so far, because the analyses give support for Notochordata hypothesis (e.g., Bourlat et al., 2009; Cameron et al, 2000;), and Olfactores hypothesis (e.g., Delsuc et al., 2006; Stach et al., 2010).

Based on morphological data the phylogenetic position of Echinodermata and Hemichordata with respect to each other and Chordata are uncertain (see Winchell et al., 2002 and references therein). There are two hypotheses: a close relationship of Enteropneusta to Chordata (Cyrtotreta hypothesis) (Schaeffer, 1987) or monophyletic Hemichordata with a sister group relationship to Echinodermata (Ambulacraria hypothesis) (Metschnikoff, 1881). Analyses based on different molecular data sets give unequivocally support for the Ambulacraria hypothesis (e.g., Cameron et al., 2000; Perseke et al., 2011; Swalla and Smith; Winchell et al., 2002).

For a long time, Chaetognatha and the lophophorate taxa (Brachiopoda, Phoronoida, and Bryozoa) have been considered of being closely related and nested within Deuterostomia as well (e.g., Barnes, 1968; Zimmer, 1973). However, molecular studies suggested chaetognaths and lophophorates within Protostomia (see Halanych, 2004 and references therein; Marletaz et al., 2006 and references therein; Matus et al., 2006). Monophyly of the latter was not supported by these early molecular studies.

In addition, further taxa, with partly less complex morphological features, such as *Xenoturbella* and Acoelomorpha (Acoela and Neodermatida) have been included within Deuterostomia to assess the evolution of simple and complex morphological features of Bilateria (e.g., Adoutte et al., 2000; Hejnol et al. 2009; Philippe et al., 2011b; Swalla and Smith, 2008). Based on morphological data several phylogenetic positions of *Xenoturbella* have been discussed: a close relationship to Acoelomorpha (see Nielsen, 2010 and references therein), to Enteropneusta and Holothuroidea (Echinodermata) (Reisinger, 1960), to Mollusca (Isrealsson, 1997), or a basal bilaterian position (Ehlers and Sopott-Ehlers, 1997). Molecular analyses yielded a sister group relationship of *Xenoturbella* and Acoelomorpha as sister to all other Bilateria (Hejnol et al., 2009) as well as a position of *Xenoturbella* within Deuterostomia (e.g., Bourlat et al., 2009; Perseke et al., 2007, Philippe et al., 2011b).

The phylogenetic relationships of these three major lineages have been revealed in many molecular analyses, but with week support only (e.g., see Hejnol et al., 2009; Philippe et al. 2011b; Swalla and Smith, 2008 and references therein). Resolving their relationships to each other could give implications for body plan and life style evolution within Deuterostomia and Bilateria (see Cameron et al., 2000 and references therein).

To resolve these issues, we addressed a mitogenomic approach comprising more than 300 complete deuterostome mitochondrial genomes. All complete mitochondrial genomes were retrieved from the NCBI data base, whereas additionally the new optimized automated

annotation pipeline MITOS was used to re-annotate all sequences to overcome annotation errors (Bernt et al., 2013c).

The analyses of this study recovered the chordate groups Craniota, Cephalochordata and Tunicata with maximal support as monophyla (Fig. 3). Cephalochordata were placed as sister group to Craniota with strong support in all analyses substantiating the Notochordata hypothesis (Fig. 3).

However, the placement of long-branched Tunicata was highly unstable depending on the taxon sampling and was affected by long-branch artifacts. As a consequence Tunicata was always related to other long branches in the data sets (Fig. 3). Because of the difficulties concerning the placement of Tunicata the monophyly of morphologically well-supported Chordata could not be recovered. Previous analyses based on mitochondrial sequence data by Singh et al. (2009) and Stach et al. (2010) were able to recover a monophyletic clade Chordata and give support for the Olfactores hypothesis, but again with low bootstrap support. In contrast to our analyses, the taxon sampling in these studies was much smaller, containing a few selected species.

As a result based on mitochondrial data, the phylogenetic position of the long-branched Tunicata remained uncertain. Even their mitochondrial gene order did not provide evidence for their phylogenetic placement as it was the case for example in another long-branching problematic taxon Myzostomida (Bleidorn et al., 2007). However, our analyses provided a good resolution of phylogenetic relationships within Tunicata with highest support values (Fig. 3), which also have been found in previous mitochondrial genome analyses using smaller taxon sampling (Singh et al., 2009; Stach et al., 2010). In all analyses Stoliodobranchiata was sister to all other tunicates, Thaliacea was placed within Phlebobranchiata, and monophyletic Aplousobranchiata has been recovered.

Nearly all analyses of this study supported the Ambulacraria hypothesis: the sister group relationship of Hemichordata (represented by Enteropneusta) and Echinodermata with high support values (Fig. 3). Furthermore, the monophyly of Echinodermata was supported by different levels of mitochondrial genome information (sequence analyses, genetic code, and gene order). The well supported monophyly of Ambulacraria and Echinodermata suggests a solitary lifestyle of the last common ancestor of Deuterostomia. Furthermore, these results suggested convergent origins of colonial lifestyles within Deuterostomia. However, more mitochondrial sequence data are necessary to resolve the relationship and discussed monophyly

of the hemichordate subgroups Pterobranchia and Enteropneustato within Ambulacraria, and finally to unravel the evolution of coloniality within deuterostomes (Swalla and Smith, 2008).

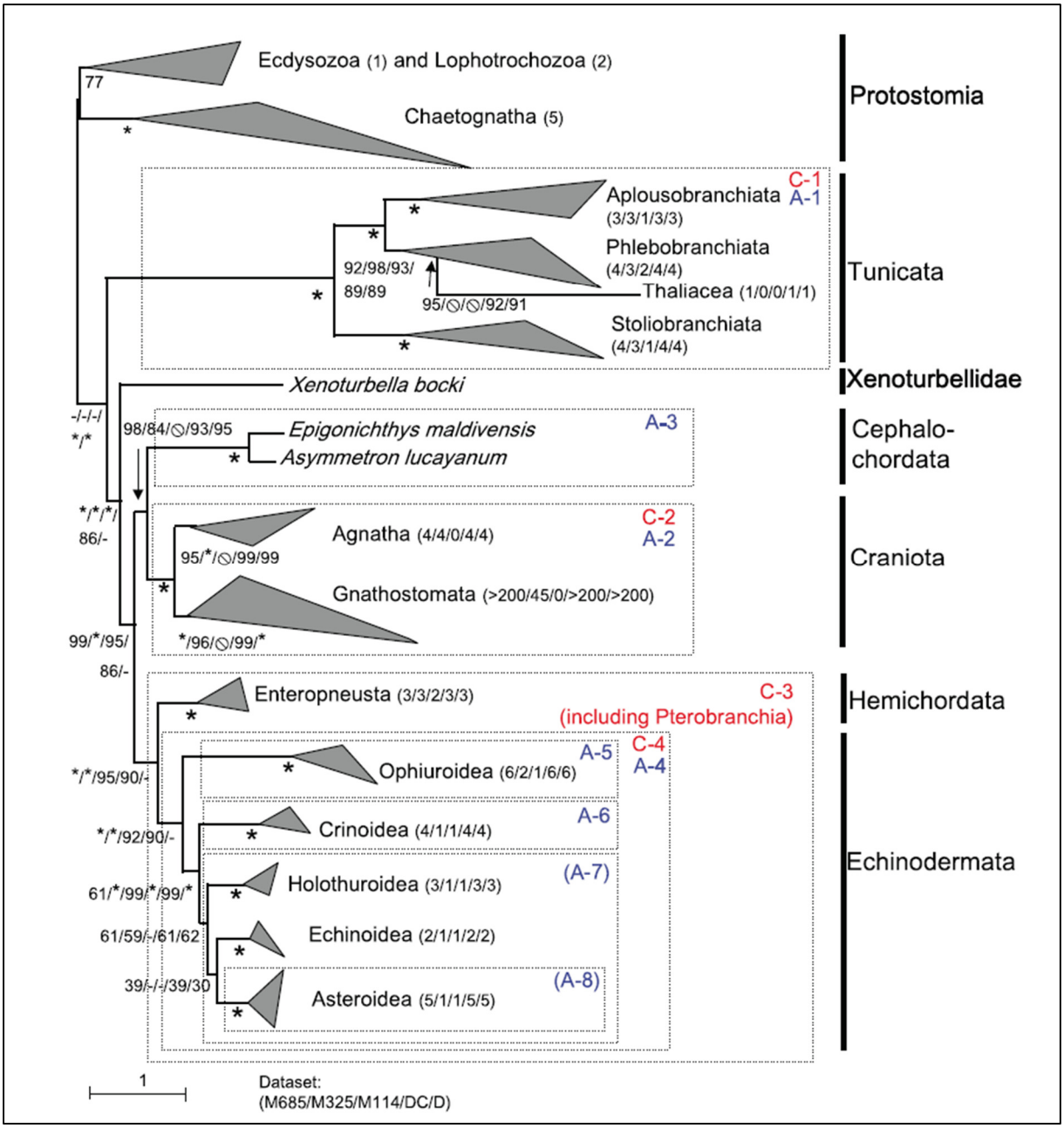


Fig. 3: Maximum likelihood tree of the dataset DC (from Perseke et al., 2013). The numbers on the nodes show support values from all analyses in the order of dataset M685, M325, M114, DC and D. An asterisk indicates highest support values obtained in all datasets (bold) or in an individual analysis. Species are collapsed to groups. The variability of the group is indicated by a triangle for the shortest and longest branch length. The numbers in brackets behind the groups show the number of analyzed species in the same order as shown for the support values. Furthermore, mt features of gene arrangements (‘‘A’’) and genetic codes (‘‘C’’) supporting a group are shown. The discontinuous lines frame the groups and the features.

Our analyses recovered the five echinoderm subgroups Crinoidea, Ophiuroidea, Asteroidea, Echinoidea and Holothuroidea as monophyla with highest support values (Fig. 3). Furthermore, the gene orders of mitochondrial genomes substantiate the monophyly of Ophiuroidea, Crinoidea and Asteroidea. Ophiuroidea are placed as sister group to all other echinoderms with high support values (Fig. 3). Within the clade of the remaining four taxa Crinoidea are placed as sister group to a clade of Asteroidea, Echinoidea, and Holothuroidea (Fig. 3). Additionally, the monophyly of this clade is supported by highly similar gene orders, but the relationships between these three taxa could not be resolved. The gene orders of Asteroidea, Echinoidea, and Holothuroidea are similar to the craniote arrangements (see Boore, 1999), but recent analyses show that the similarity most likely arose by convergence (Perseke et al. 2010). The results of the phylogenetic relationships of the echinoderm subgroups are similar to previous analyses of mitochondrial sequences, but in strong contrast to the Eleutherozoa hypothesis suggesting Crinoidea as sister group to all other echinoderms (e.g., Mallatt and Winchell, 2007).

Even the affiliation of *Xenoturbella* could be resolved in our analyses. *Xenoturbella* was always placed as sister to all deuterostomes (except for tunicates) with robust support values (Fig. 3). This close relationship was substantiated by the mitochondrial gene order which was similar to those of the hemichordate *Balanoglossus* and of chordates. Unfortunately, Acoelomorpha were not included in these analyses. Therefore, an affiliation to Deuterostomia cannot be ascertained as well as their possible phylogenetic relationship to *Xenoturbella*.
The simple bauplan of *Xenoturbella* and the basal position within Deuterostomia indicates that the last common ancestor of Deuterostomia did not exhibit a high degree of morphological complexity.

4.3. **"Platyzoan paraphyly based on phylogenomic data supports a nonacoelomate ancestry of spiralia"** (Struck et al., 2014a)

In 1998, Cavalier-Smith proposed Platyzoa as a taxon consisting of Platyhelminthes and Acanthognatha, comprising Gastrotricha, "Rotifera", Acanthocephala, and Gnathostomulida. The platyzoan monophyly as well as phylogenetic relationships between these platyzoan subgroups, and the internal phylogenies of them are highly discussed. Some authors regarded Platyzoa as sister group taxon to Spiralia (= Lophotrochozoa sensu stricto = Trochozoa (Giribet

el al., 2000; Roule, 1891; Rouse, 1999)) (see Edgecombe et al., 2011; Halanych, 2004), whereas others placed the platyzoan taxa within Lophotrochozoa, thus rendering Spiralia synonymous with Lophotrochozoa (see Halanych, 2004).

Moreover, additional taxa have been discussed to be platyzoan taxa as well: Micrognathozoa, Cycliophora, and Entoprocta (e.g., Halanych, 2004; Kristensen and Funch, 2000). Micrognathozoa show anatomical similarities to Gnathostomulida and "Rotifera", two jaw-bearing platyzoan taxa (e.g., Kristensen and Funch, 2000, Paps et al., 2009a; Worsaae and Rouse, 2008). On the other hand, Cycliophora and Entoprocta are possibly closely related to other spiralian taxa such as Ectoprocta (e.g., Hausdorf et al., 2007; Hejnol et al., 2009; Struck and Fisse, 2008).

Platyzoan taxa are hard to define morphologically. Typically, they are characterized by a small body size, a flat unsegmented worm-shaped body, which is usually ciliated, the lack of a vascular system (acoelomate body cavity), and a direct development. However, unique morphological autapomorphies supporting monophyletic Platyzoa were not found (Giribet, 2008).
Morphological-cladistic and total-evidence analyses find platyzoan taxa either as non-monophyletic assemblages within Spiralia or even Protostomia, or with only weak support as monophyletic Platyzoa (e.g.; Giribet et al., 2000; Peterson and Eernisse, 2001; Zrzavý, 2003; Zrzavý et al., 1998; and references therein). Furthermore, phylogenetic analyses based on single genes could not resolve this question as well (e.g.; Halanych, 2004; Paps et al. 2009a; Passamaneck and Halanych, 2006; and references therein). However, phylogenetic analyses based on large-scale molecular data sets were able to recover monophyletic Platyzoa with more or less strong support and eventually due to long-branch attraction (e.g.; Halanych, 2004; Hausdorf et al., 2007; Hejnol et al., 2009; Helmkampf et al., 2008; Struck and Fisse, 2008; Witek, et al., 2009). In contrast, other analyses based on large scale data sets found paraphyletic assemblages of platyzoan taxa (e.g., Hausdorf et al., 2007; Witek et al., 2008; 2009) or polyphyletic positions within Protostomia (e.g., Helmkampf, Bruchhaus, and Hausdorf, 2008; Nesnidal et al., 2010; Struck and Fisse, 2008).

The putative monophyly of Platyzoa and the phylogenetic position of platyzoan taxa within Spiralia are still contentious. The placement of platyzoan taxa appears unstable, probably due to low data and taxa coverage (Edgecombe et al., 2011). Furthermore, in many analyses platyzoan taxa are characterized by long branches, which could cause wrong phylogenetic groupings by long-branch attraction (Edgecombe et al., 2011).

The aim in the study by Struck et al. (2014a) was to overcome these problems by now employing a phylogenomic approach with an increased number of species within Platyzoa and increased species-specific sequence coverage in data sets of up to 82,162 amino acid positions. The idea was that an increased number of taxa and data should yield to a better resolution of platyzoan or spiralian phylogeny, respectively, with strong support values. To get a robust phylogeny with strong support values it is important to disentangle phylogenetic signal from misleading effects such as long-branch attraction (long branch score).

We generated transcriptome sequence data for ten platyzoan and two nemertean species using second-generation sequencing technology and a modified RNA amplification method (see supplementary table S1). The complementation of these data was done with genomic or transcriptomic data of 53 other spiralian and ecdysozoan species, including other representatives of Platyzoa (see supplementary table S2). The taxon coverage of Platyzoa was increased by 3.5-fold. For specific platyzoan taxa such as Gastrotricha and Syndermata the taxon coverage increased 5-fold, in comparison to previous large-scale analyses of spiralian phylogeny (Dunn et al., 2008; Hejnol et al., 2009). After orthology assignment using HaMStR (Ebersberger et al, 2009), the data were screened for sequence redundancy (Kvist and Siddall, 2013), potentially paralogous sequences (Struck, 2013) and contamination (Struck, 2013). The result was a data pruning of about 7% of sequence data (see supplementary tables S3 – S8).

The phylogenomic analyses of this study did not recover monophyletic Platyzoa. Rather, Platyzoa appeared as a paraphyletic assemblage with respect to other Spiralia. Platyhelminthes and Gastrotricha formed a monophyletic clade (named Rouphozoa) as sister group of the monophyletic Lophotrochozoa (sensu stricto) (Fig. 4). Furthermore, Syndermata and Gnathostomulida grouped together in a sister group relationship, which is congruent with the Gnathifera hypothesis (Ahlrichs, 1997; Hausdorf et al., 2010; Herlyn and Ehlers, 1997; Witek et al., 2009). Gastrotrich and gnathostomulid species were identified as most unstable species within the sampled platyzoans (leaf stability index). Furthermore, due to the exclusion of Gnathifera, Rouphozoa and Lophotrochozoa sensu stricto represent a monophyletic group here named Platytrochozoa (Fig. 4).

Platyzoan paraphyly provides strong support for the view that the last common ancestor of Spiralia was a small organism exhibiting a simple organization without coelom, segmentation, and complex brain structures, which very probably inhabited the marine interstitial realm (“acoeloid-planuloid” hyopothesis). This implies that Bilateria evolved most likely from a simple organized ancestor to more complex descendants independently within the three major

bilaterian clades Deuterostomia, Ecdysozoa, and Spiralia. However, it cannot be ruled out that several independent events of secondary reduction, such as miniaturization or progenesis within these taxa lead to loss of morphological complexity, as it is known e.g., to have occurred within annelids and arthropods (Bleidorn, 2007; Jenner, 2004a).

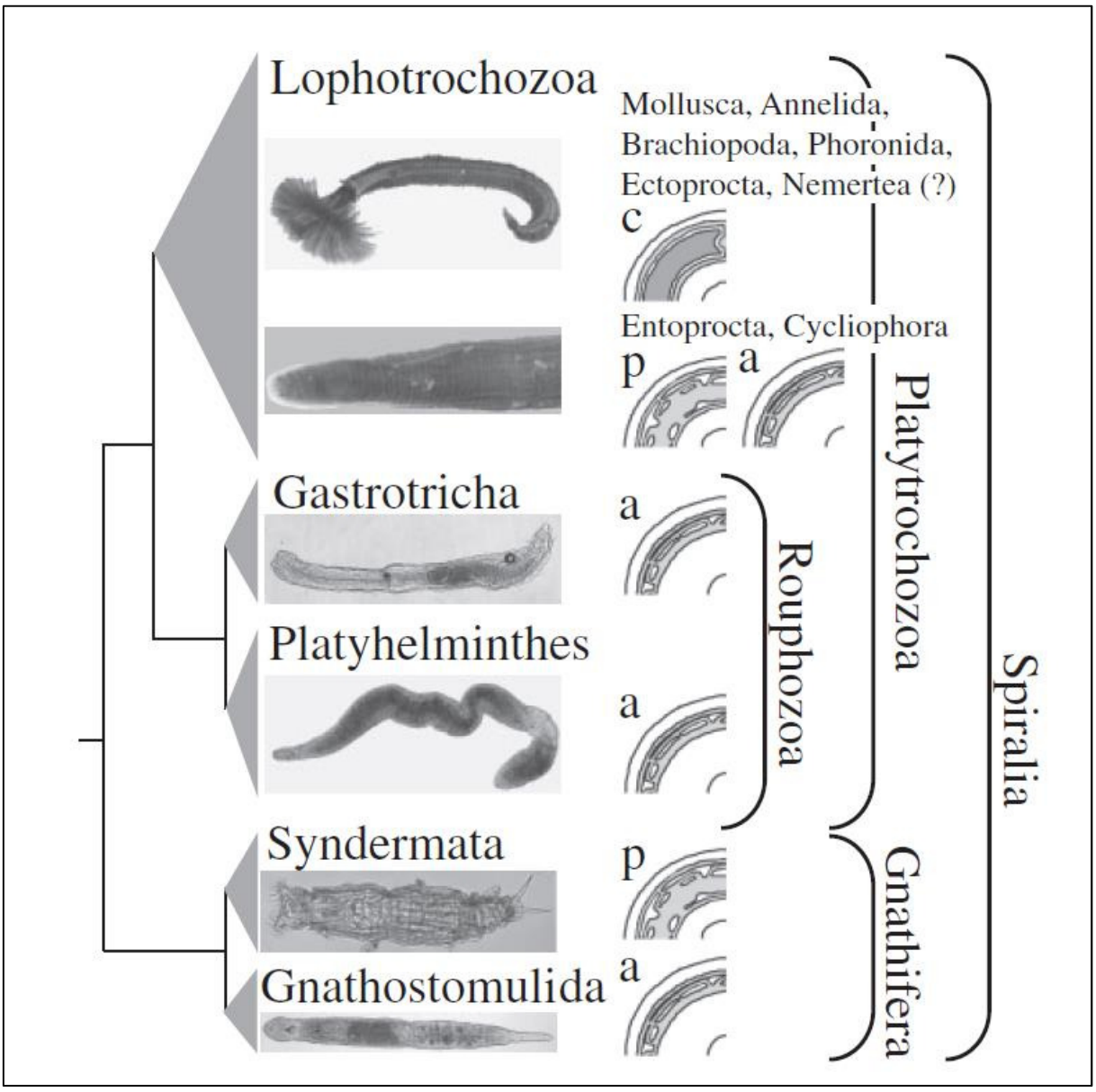


Fig. 4: Proposed phylogeny of Spiralia based on phylogenomic analyses (from Struck et al., 2014a). Drawings depict the acoelomate (=a), pseudocoelomate (=p), and coelomate (=c) body organization. Picture of *Rotaria neptunoida* (Syndermata) was courtesy of Michael Plewka. (?) means that it is still discussed if the lateral vessels of the nemertean circulatory system are homologous to coelomic cavities of other lophotrochozoan taxa (Turbeville 1986). Higher taxonomic unites and names are indicated.

4.4. "Elucidating the phylogenetic position of Gnathostomulida and first mitochondrial genomes of Gnathostomulida, Gastrotricha and Polycladida (Platyhelminthes)" (Golombek et al., 2015)

Gnathostomulida is a small taxon of small marine worms with a slender and dorsoventrally flattened body (Brusca and Brusca, 2003) inhabiting the marine interstitium (Nielsen, 2012). Gnathostomulida are characterized by the presence of a specific ventral pharyngeal jaw apparatus (Herlyn and Ehlers, 1997; Kristensen and Nørrevang, 1977; Sørensen et a., 2003, Sterrer et al., 1985). Gnathostomulids have monociliated epithelial cells rather than multiciliated, which is exceptional for bilaterian organisms and only shared with certain Gastrotricha and the basal taxon Oweniidae (Nielsen, 2012; Sterrer et al., 1985). Since their discovery in 1956 by Ax their systematic position has been debated and the first phylogenetic hypotheses were brought forward in the 1980th (Ax, 1985; Sterrer et al., 1985).

Based on morphological data there are three hypotheses discussing the phylogeny of Gnathostomulida. The first hypothesis is the Plathelminthomorpha hypothesis going back to Ax (1985) and describing a close relationship of Gnathostomulida to Platyhelminthes (e.g., Giribet et al.,2000 and references therein; Peterson an Eernisse, 2001 and references therein). The second hypothesise is the Monokonta/Neotrichozoa hypothesis placing Gnathostomulida as sister group to Gastrotricha (Cavalier-Smith, 1998; Zrzavy et al., 2001). The third one is the Gnathifera hypothesis proposing a closer relationship of gnathostomulids to Syndermata (e.g., Herlyn and Ehlers, 1997; Nielsen, 2012; Zrzavý, 2003 and references therein). These latter ideas were first brought into discussion by Sterrer at al. (1985).

First molecular analyses based on single or a few genes were inconclusive. These analyses found Gnathostomulida as sister to Syndermata (e.g., Zrzavý, 2003), to Gastrotricha (e.g., Zrzavý et al., 2001), within Platyzoa (e.g., Giribet et al., 2000) or as sister to all other spiralian taxa (e.g., Papset al., 2009b). First phylogenomic studies concerning bilaterian phylogeny placed Gnathostomulida as sister to Acoela or Syndermata (Dunn et al., 2008; Hejnol et al., 2009). The bootstrap support was low and Gnathostomulida was the most unstable taxa in these analyses. Recent phylogenomic studies focusing on platyzoan phylogeny, especially by using specific data sets and analytical methods, found a closer relationship of Gnathostomulida to Syndermata (Gnathifera hypothesis) (Struck et al., 2014a), but also positions close to Platyhelminthes (Plathelminthomorpha hypothesis) or Gastrotricha (Monokonta/Neotrichozoa

hypothesis) were found and could not be rejected (Wey-Fabrizius et al., 2014; Witek et al., 2009).

To resolve the phylogenetic affinities of Gnathostomulida with respect to Syndermata, Platyhelminthes and Gastrotricha we determined the first data of complete or nearly complete mitochondrial genomes for two gnathostomulids (*Gnathostomula paradoxa* and *Gnathostomula armata*), one gastrotrich (*Lepidodermella squamata*) and one polyclad species (*Stylochoplana maculata*) as a representative of free-living Platyhelminthes.

In our analyses, mitochondrial data support a sister group relationship of Gnathostomulida to Syndermata (Gnathifera hypothesis), with strong bootstrap support in each analysis (Fig. 5). Interestingly these analyses neither reconstructed a close relationship of Gnathostomulida to Gastrotricha (Monokonta/Neotrichozoa hypothesis) nor to Platyhelminthes (Plathelminthomorpha hypothesis), rejecting two of the morphology-based hypotheses. Furthermore, the monophyly of Gnathostomulida, Platyhelmihtes, Syndermata, Annelida, Nemertea, Entoprocta, Brachiopoda, and Ectoprocta, but not of Mollusca, was generally found with robust support values (Fig. 5).

Previous studies (e.g., Struck et al., 2014a; Zhong et al., 2011) showed that Gnathostomulida, Syndermata or Platyhelminthes, as well as mitochondrial sequence data in general, can be affected by misleading biases such as taxon instability, branch length heterogeneity and base composition heterogeneity. Sensitivity analyses showed that the position of Gnathostomulida is consistent across the different analyses and furthermore, the support values for the Gnathifera hypothesis increased.

In contrast to mitochondrial sequence data, the mitochondrial gene order data do neither support nor reject a closer relationship of Gnathostomulida to Syndermata, Gastrotricha, or Platyhelminthes. The gene order of the two gnathostomulid species is the same with exception of the position of some tRNA's. The gene order of Gnathostomulida is not similar to one of the known gene orders of Syndermata, Gastrotrich or Platyhelminthes. The gene order of the polyclad *Stylochoplana maculata* is different from the known orders of Platyhelminthes as well. These observations give new insides into the high value of mitochondrial gene order as a phylogenetic marker.

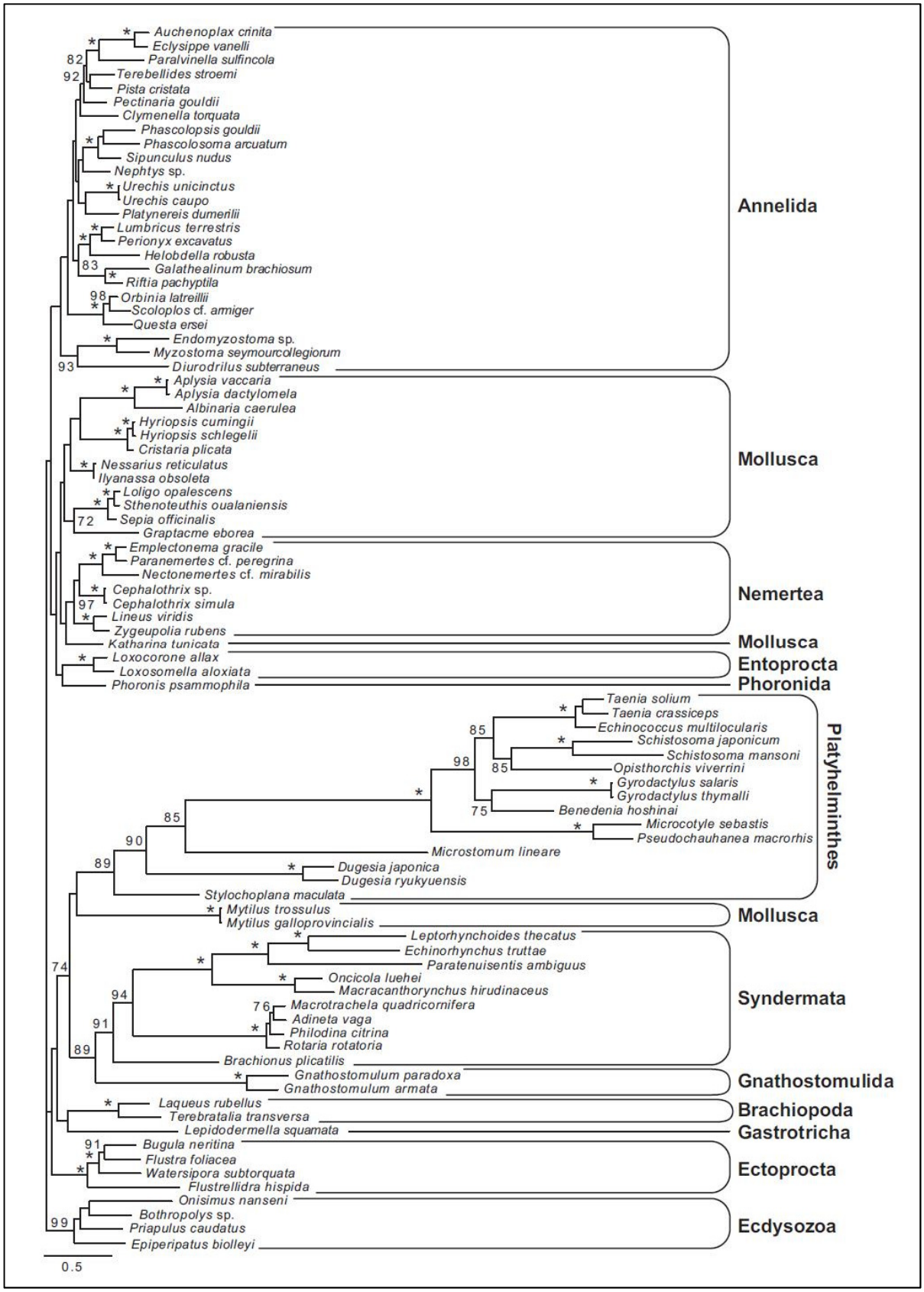


Fig. 5: Phylogram of the non-partitioned Maximum likelihood analysis with all taxa (from Golombek et al., 2015). Only bootstrap values ≥70 are shown. An asterisk (bold) indicates bootstrap value of 100. Higher taxonomic units are indicated.

Interestingly the Gnathifera hypothesis is the only one which is supported by highly specific and complex morphological synapomorphies as well. These features are an extended non-contractile region of the pharyngeal musculature (Zrzavý, 2003), as well as the location of the protonephridial lumen within the protonephridial cells (Haszprunar, 1996; Zrzavý, 2003). However, the most significant and name-giving character supporting the Gnathifera hypothesis is the complicated jaw apparatuses, which can be found in gnathostomulids and the rotifers. Our analyses strongly support the view that these complex jaw apparatuses are homologous as suggested previously in numerous morphological studies (e.g., Ahlrichs, 1997; Herlyn and Ehlers, 1997; Jenner, 2004b and references therein; Sørensen, 2003). The homology of the jaw apparatuses of Gnathostomulida and Syndermata is further substantiated by morphological similarities to the jaw apparatus of Micrognathozoa (Sørensen, 2003), whose phylogenetic position is still under controversial debate.

4.5. Phylogeny of Annelida

Annelids were still part of discussions about major transitions in animal evolution concerning the development of segmentation, evolution of the nervous system, origins and diversification of larval types, as well as the transitions to a terrestrial lifestyle (Purschke, 1999; Rouse, 1999; Seaver 2003; Jekely et al. 2008). Annelida is one of three major animal groups possessing true segmentation. The general assumption is, that the understanding of annelid body-plan evolution is essential to clarify the evolution of Bilateria (Raible et al., 2005; Rivera and Weisblat, 2009; Tessmar-Raible and Arendt, 2003). Furthermore, it has been proposed that the bilaterian ancestor resembled an annelid. Because of this, a robust annelid phylogeny is crucial to understand animal body-plan evolution, and to reveal the bilaterian ground pattern.

Annelida, the segmented worms, is a highly diverse animal group within Spiralia (Lophotrochozoa sensu lato) with over 21,000 described species. Annelids inhabit diverse terrestrial, aquatic and marine habitats. They are highly abundant in marine environments and constitute the dominant macrofauna from the intertidal zone down to the deep sea. This group shows a high diversity in body forms, developmental pattern, life modes, feeding and reproductive strategies (Rouse and Pleijel, 2001). Many aspects of annelid phylogeny are poorly understood.

Traditionally, Annelida has been classified into Polychaeta and Clitellata (Grube, 1850; Fauchald and Rouse, 1997; Rouse and Pleijel, 2001 and references thererein). The monophyly of Clitellata is very well supported by analyses based on morphological and molecular data and provided robust phylogenetic hypotheses within this taxon (Erséus, 2005). However, in molecular analyses the monophyly of Polychaeta lacked support as well as the resolution of polychaete relationships was difficult (McHugh, 2000; 2005). Polychaetes have been classified into nearly 80 families, which are generally supported as monophyletic, but there was only weak support for more inclusive nodes (Rouse and Pleijel, 2001; McHugh, 2005; Struck et al., 2008). Therefore, the phylogeny of polychaetes was still under hardly debate until recently.

Until the middle of the last century polychaetes were classified as either Errantia or Sedentaria, based on their morphology and their mode of living (Fauvel, 1923; 1927; de Quatrefages, 1866). These names reflect their mode of living as either more vagile (errant), or as more sessile (sedentary). But this classification only based on the presence or absence of distinct body regions (de Quatrefages, 1866). For this reason, this systematization was rejected by most authors, because this grouping was regarded to reflect primary the mode of life rather than their evolutionary history and was useful only for practical purposes (Dales 1963; 1977; Fauchald, 1977).

Based on morphological cladistics analyses Polychaeta were regarded as monophyletic, and proposed constisting of two major clades, Scolecida and Palpata, whereas Palpata was further divided into Canalipalpata and Aciculata (Rouse and Fauchald, 1997; Rouse, 1999; 2000). However, the monophyly of Scolicida, Paplpata, Aciculata, and Canalipalpata was still questioned (Bartolomaeus, 1998; Bartolomaeus et al., 2005; Hausen, 2005; Westheide et al., 1999). Also, molecular analyses did never recover the proposed phylogeny by Rouse and Fauchald (1997) (e.g., Hall et al., 2004; McHugh, 2000; 2005; Struck and Purschke, 2005). More recently it was shown that three so-called phyla of non-segmented groups might represent annelid subtaxa: Pogonophora (now Siboglinidae), Echiura and Sipuncula (see Struck et al., 2014b and references therein). This would indicate a higher variability of segementation than previously thought including various levels of reduction until a more or less complete loss.

Based on early phylogenetic analyses there were two major hypotheses describing the evolution of annelids. The first scenario suggests the evolution of annelids from an earthworm-like ancestor with a more or less simple external organization to a more complex form, in which the stem species of Annelida lacked head appendages, parapodia and pygidial cirri, and had only simple chaetae as an adaption to a burrowing life style (Clark, 1969; Westheide, 1997).

Morphological support for this scenario was given by cladistics analyses of Rouse (1999; 2000), and Rouse and Fauchald (1997).

The second scenario suggests the evolution of annelids from a more complex errant and epibenthic ancestor with well-developed head appendages, parapodia and chaetae to more simple forms, evolved due to reductions and losses (e.g., Storch, 1968; Westheide, 1997). This hypothesis is supported by various investigations based on morphological as well as on molecular data, which found Echiura, Siboglinidae, and Clitellata within Polychaeta, and Myzostomida and Sipuncula within the annelid radiation (McHugh, 1997; Struck et al., 2008; Bleidorn et al., 2009b; Zrzavy et al., 2009; Helm et al., 2012).

In the last decade, a lot of different molecular approaches have been done to resolve the issues of annelid phylogeny and evolution (see Struck et al., 2014b, and references therein): First, a targeted-gene approach was applied using complete mitochondrial genomes, nuclear genes such as 18S and 28S rRNA, protein-coding genes like aldolase, ATP synthase b chain, catalase, or methionine adenosyltransferase, myosin II heavy chain. Second, the determination of expressed sequence tags (EST) libraries of various annelid taxa (phylogenomic approaches), and third, the usage of gene structure marker systems such as introns with probably phylogenetic signal. All these different approaches allowed more robust conclusions about annelid phylogeny and evolution (see Struck et al., 2014b, and references therein). These analyses revealed that both evolutionary scenarios did not sufficiently describe the most probable course of annelid evolution and a new ground pattern reconstruction has been put forward (Struck et al., 2011, 2015; Weigert et al. 2014). This may be seen as a modified version of the second scenario and a complete rejection of the first hypotheses based on an earthworm-like ancestor.

First of all, these studies clearly showed that the usage of targeted-gene approaches is not the method of choice to resolve annelid phylogeny and evolution. Struck et al. (2008), showed that to increase the amount of sequence information will more likely recover annelid phylogeny. Additionally, Dordel et al. (2010) suggested that amino acid data are better suited than nucleotide data because of the higher number of possible character states. Furthermore, Dordel et al. (2010) demonstrated in comparison to targeted-gene approaches, the potential of a phylogenomic approach. The analyses of this study favored a placement of Sipuncula within Annelida, and significantly rejected a sister group relationship of Sipuncula and Annelida.

In 2011, Struck et al. provided the, at that time, best supported annelid phylogeny based on phylogenomic data, with a comprehensive taxon sampling including also Chaetopterida, Myzostomida, Echiura, Sipuncula, and Clitellata. In this analysis Chaetopterida, Myzostomida

and Sipuncula were part of the basal annelid radiation, and the remaining annelids split into two well-supported major clades which have been also recovered in analyses before with exception of the position of Amphinomida (Struck et al., 2008; Zrzavý et al., 2009). These two clades are very similar to previously proposed groups of "Polychaeta": Errantia and Sedentaria (Fauvel, 1923; 1927; de Quatrefages, 1866). Traditionally, Errantia consists of Amphinomida, Phyllodocida, and Eunicida. Orbiniidae were seen basal in Sedentaria as a kind of intermediate stage between Errantia and Sedentaria. Traditionally, Sedentaria comprises all remaining polychaete groups. Struck et al. (2011) resurrected the names Errantia and Sedentaria, together forming the higher-level clade Pleistoannelida (Struck, 2011). Furthermore, Weigert et al. (2014) recovered most nodes found by Struck et al. (2011) and provided robust support for a basal branching of Oweniidae, Magelonidae, Chaetopteridae, Sipuncula, and Amphinomidae at the base of the annelid tree. This study led to a slightly modified view of annelid phylogeny in placing the errant forms Amphinomida among the basal branching groups and removing Orbiniida from Errantia and found them basal in Sedentaria. This general picture of annelid phylogeny seems now to be stable and has been confirmed in several subsequent analyses (see below).

Phylogenomic approaches have clear advantages in comparison to analyses based on targeted genes (see Jeffroy et al., 2006). However, large-scale phylogenomic analyses may fail to resolve phylogenetic relationships in the presence of mutational saturation (Jeffroy et al., 2006), or lineage-specific divergence rates (Rasmussen and Kellis, 2007). For the resolution of deep organismal divergences, alternative molecular markers are needed. The analysis of mitochondrial sequence data and the gene order provide valuable information to resolve phylogenetic relationships at different levels (Vallès and Boore, 2006). Mitochondrial genomes have been shown to be suitable for younger divergences (Bernt et al., 2013b). Especially, the gene order of mitochondrial genomes has successfully been used as phylogenetic marker in molecular systematics (e.g., Bleidorn et al., 2007; Golombek et al., 2013).

4.5.1. "Detecting the symplesiomorphic trap: a multigene phylogenetic analysis of terebelliform annelids" (Zhong et al., 2011)

Conflict in signal is a potential problem for tree reconstruction. Analyses of different molecular data such as nuclear and mitochondrial data, do not always result in congruent phylogenetic reconstructions (e.g., Galtier et al., 2009). There are a lot of molecular evolutionary events which can cause inferred gene trees to differ from species trees: gene duplication, horizontal gene transfer, heterotachy, gene extinction, long-branch attraction, substitution rates, saturation and model misspecifications. Thus, taxa affected by such biases are often misplaced in phylogenetic analyses (e.g., Bleidorn et al., 2007; Golombek et al., 2013). In addition, it has been shown that such biases may also influence the placement of unbiased taxa (Zhong et al., 2011). This phenomenon is called symplesiomorphy trap, which has been characterized as a special class of long-branch attraction (Wägele, 2005; Wägele and Mayer, 2007). The symplesiomorphy trap describes a paraphyletic assemblage of taxa grouping together as monophyletic. This putative monophyly based on the possession of symplesiomorphic characters, which are mistakenly supposed to be apomorphic. The symplesiomorphy trap is known for morphological data. For example, based on morphological data Clitellata were placed as sister group to Polychaeta due to the lack of typical polychaete characters such as parapodia and nuchal organs (Rouse and Fauchald, 1997), whereas molecular data clearly place Clitellata within Polychaeta (e.g., Dordel et al., 2010; Struck et al., 2007; Struck et al., 2011). These obviously plesiomorphic characters such as nuchal organs, palps, parapodia etc. were mistaken as polychaete apomorphies not recognizing their loss in the last common ancestor of Clitellata. The symplesiomorphy trap is not restricted to morphological data, and can also apply to molecular data (Wägele, 2005). The detection of the trap occurs more or less indirectly by revealing chraracteristic signatures in the data and excluding possibilities of incongruence, because compositional bias or increased substitution rates do not affect the misplaced taxa directly. Furthermore, the monophyly for misplaced taxa is not artificial because the monophyly is based on apomorphies for a higher taxonomic unit. At last, the 'true' phylogeny has to be known to detect the symplesiomorphy trap.

Terebelliformia (sensu Rouse and Pleijel, 2001) are marine tube-dwelling annelids comprising Alvinellidae, Ampharetidae, Terebellidae, Trichobranchidae and Pectinariidae (see Rouse and Pleijel, 2001 and references therein). The phylogenetic position of Terebelliformia within Annelida and relationships among these five 'families' remain debatable. However, the

monophyly of Terebelliformia has been recovered (Colgan et al., 2001, Struck et al., 2007) with Trichobranchidae as a 'subfamily' of Terebillidae (e.g., Rouse and Pleijel, 2001), or a recognized 'family' (Glasby et al., 2004; Fauchald and Rouse, 1997). Phylogenetic analyses based on mitochondrial sequence data strongly supported a Trichobranchidae/Terebellidae clade (Zhong et al., 2008).

In this study, mitochondrial genome sequence data, data from nuclear 18S rDNA, 28S rDNA and elongation factor-1α genes have been used to resolve the phylogeny of Terebelliformia. Three different data sets were analyzed: a mitochondrial data set, a nuclear data data set, and a combined data set consisting of mitochondrial and nuclear genes. The phylogenetic trees from the analyses of the mitochondrial and the combined data set were nearly the same with differences only occurring in outgroup relationships (Fig. 6B, C). In these analyses the monophyly of Terebelliformia is significantly supported. Trichobranchidae was placed as sister to Terebellidae with strong support values in all analyses of these two data sets (TriTer hypothesis) (Figs. 6B, C). Furthermore, Pectinariidae branched off first within terebelliforms representing the basal lineage in Terebelliformia, and Alvinellidae was recovered as sister to Ampharetidae in all analyses of both data sets (Fig. 6B, C). Topology testing significantly rejected a sister group relationship of Terebellidae to a clade of Alvinellidae/Ampharetidae (TerAA hypothesis) in both data sets, and significantly rejected a sister group relationship of Trichobranchidae to a clade of Alvinellidae/Ampharetidae (TriAA hypothesis). However, it has to be mentioned that topology testing of the combined data set did not significantly reject the TriAA hypothesis.

The phylogenetic trees from the nuclear data set were different (Fig. 6A). The monophyly of Terebelliformia was not recovered due to the placement of *Pectinaria gouldi* as sister to the sipunculid *Phascolopsis gouldi*, but only with weak support. The four remaining taxa formed a clade with strong nodal support. Furthermore, a well-supported sister group relationship of Alvinellidae and Ampharetidae was recovered as observed in the phylogenetic trees from mitochondrial and combined data set. Based on nuclear data, a sister group relationship of Trichobranchidae to a clade of Alvinellidae/Ampharetidae was well supported (TriAA hypothesis), whereas topology testing significantly rejected the alternative TriTer and TerAA hypotheses.

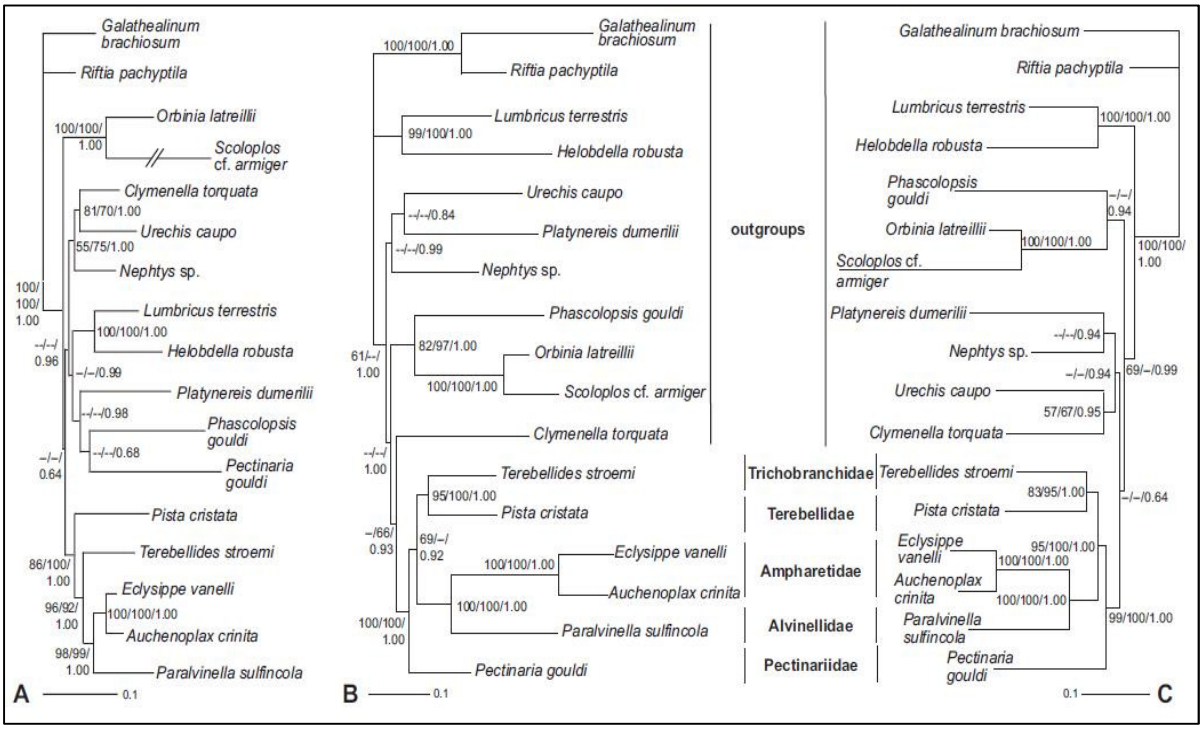


Fig. 6: Phylogenetic reconstructions using nuclear, mitochondrial and combined nucleotide datasets with 17 taxa (from Zhong et al., 2011). (A) Nuclear Maximum likelihood tree. The branch of *Scoloplos cf. armiger* was reduced by 75%. (B) Mitochondrial Maximum likelihood tree. (C) Combined Maximum likelihood tree. All trees represent identical topologies regarding terebelliform relationships for both Maximum likelihood and partitioned Bayesian inference. The nodal support values are given at branches in the order: non-partitioned Maximum likelihood bootstrap, partitioned Maximum likelihood bootstrap and Posterior probability of the Bayesian inference. A dash indicates < 50%.

These different results indicate that there is incongruence between mitochondrial and nuclear data. To pinpoint the source of these incongruence removal and addition of taxa, spectral analyses, detection of compositional biases, models of non-stationary sequence evolution, and recording of characters, were done.

The placement of Trichobranchidae and Terebellidae in both mitochondrial and combined analyses was influenced by biases in nucleotide frequencies of mitochondrial sequence data.

However, the taxa themselves did not exhibit compositional biases. Quite the contrary, the biases in Alvinellidae and Ampharetidae influenced the misplacement of Trichobranchidae and Terebellidae and can be related to the "symplesiomorphy trap". The investigations revealed that mitochondrial data of Alvinellidae and Ampharetidae exhibited strong compositional biases causing the misplacement of Trichobranchidae. These results indicate that probably the TriAA hypothesis (Trichobranchidae placed as sister to a clade of Alviellidae/Ampharetidae) supported in the nuclear data sets might be the 'correct' (most probable) hypothesis concerning the phylogeny of terebelliform annelids.

To ameliorate the effect of the symplesiomorphy trap the best stragtegy is to increase the taxon sampling of the in-group (Wägele and Mayer, 2007). But, increasing the taxon sampling might not always be the best possibility. In this study, the most effective strategy was RY coding to reduce the effects of compositional biases within pyrimidine and purines. But strong support for the TriAA hypothesis with either mitochondrial or combined data was not recovered. Moreover, the phylogenetic signal in all data sets was substantially decreased by RY coding. Even the exclusion of biased taxa had no noticeable effect. The influence of the symplesiomorphy trap could not be completely ameliorated.

4.5.2. "Mitochondrial genomes to the rescue – Diurodrilidae in the myzostomid trap" (Golombek et al., 2013)

Diurodrilidae is a monogeneric taxon of Spiralia only comprising the genus *Diurodrilus* (Remane, 1925), which includes six described species, but there are eventually four more species, which have not being described yet (Paxton, 2000; Riser, 1984; Worsaae and Kristensen, 2005; Worsaae and Rouse, 2008).

Diurodrilus is a meiofaunal interstitial taxon, which occurs worldwide from intertidal sandy beaches to subtidal habitats. This taxon is characterized by small size (length from 250 up to 500 µm, and width of 50-80 µm) and a dorsoventrally flattened body, which consists of a prostomium, a peristomium, seemingly five weakly recognizable trunk segments and a pygidium given an annelid affinity (Worsaae and Kristensen, 2005). Morphological evidence for the monophyly of the family Diurodrilidae is given by single epidermal cells (ciliophores) which are located ventrally on head and trunk (Worsaae and Kristensen, 2005). Other typical morphologic features of the genus *Diurodrilus* are the forked, toe-like pygidium which carries duo-gland systems of adhesive glands for the attachment and detachment to sand grains, adhesive head glands, dorsal plates, unusual spermatozoa, and a ventral muscular pharynx with large central glands (Kristensen and Niilonen, 1982; Worsaae and Kristensen, 2005; Worsaae and Rouse, 2008).

This interstitial taxon has always been considered as part of the annelid radiation, either as basal or derived taxon. Remane (1925) placed *Diurodrilus* with the annelid taxon Dinophilidae as part of 'Archiannelida' (see below). But, there was no morphological character which supported

the placement of *Diurodrilus* within Dinophilidae, and a new family Diurodrilidae (Kristensen and Niilonen, 1982) and order Diurodrilida (Westheide, 1985) was erected within Annelida. However, *Diurodrilus* has none of the typical annelid characters such as head appendages, parapodia, chaetae, nuchal organs, or segmentation. Also, comprehensive analyses of the nervous system and the musculature of Diurodrilidae were not able to reveal evidences for an ancestral segmentation (Worsaae and Rouse, 2008), as it had been possible for the non-segmented annelid taxa Echiura and Sipuncula (Hessling, 2002; 2003; Hessling and Westheide, 2002; Kristof et al., 2008; Purschke et al., 2000).

In 2008, Worsaae and Rouse did a comprehensive study of *Diurodrilus* based on morphological and molecular data (18S rDNA and 28S rDNA) to address the hypotheses that *Diurodrilus* is paedomorphic and an annelid taxon. Morphological data of this study could not find any close affinity of *Diurodrilus* to any annelid taxon and an exclusion from annelids was proposed suggesting that, *Diurodrilus* is more closely related to other protostome taxa.
However, like annelid larvae *Diurodrilus* lacks every common adult annelid character and could represent an extreme case of paedomorphosis, though obvious retained larval characters are missing. Also, the molecular analyses of combined 18S rDNA and 28S rDNA data did not place *Diurodrilus* within Annelida. Instead of that, the molecular data suggested a closer relationship to platyzoan taxa albeit with weak support. *Diurodrilus* was found either unresolved in a basal trichotomy, or in a basal position as sister to Micrognathozoa. Worsaae and Rouse (2008) suggested removing Diurodrilidae from Annelida and to place them as *incertae sedis* within Spiralia with affinities to platyzoan taxa. Furthermore, in the molecular analyses of Worsaae and Rouse (2008) the platyzoan taxa as well as *Diurodrilus* exhibit strongly increased substitution rates in both genes, what is known to result in the artificial attraction of long-branched taxa (for review see Bergsten, 2005). Thus the close phylogenetic association of Diurodrilidae and platyzoan taxa found by these authors might be due to long branch attraction rather than true phylogenetic relationships.

To address the uncertain phylogeny of Diurodrilidae we determined nearly the complete mitochondrial genome of *Diurodrilus subterraneus* as well as complete 18S and 28S sequence data. Furthermore, we also determined 18S and 28S sequence data of several platyzoan taxa (three gastrotrichs, one rotifer, one gnathostomulid) to reveal the affinities to platyzoan taxa. Additionally, we also included published 18S and 28S data of numerous annelid taxa which were not included by Worsaae and Rouse (2008) to achieve a better resolution in the analyses.

The phylogenetic analyses of 18S rDNA and 28S rRNA data recovered neither the monophyly of Annelida nor of Platyzoa. However, *Diurodrilus subterraneus* was sister *to Diurodrilus* sp. and form the clade Diurodrilidae with highest bootstrap support. Diurodrilidae was placed with the annelid taxon Terebelliformia as part of the large clade of annelid taxa but only with low bootstrap support. To get a better resolution of these phylogenetic reconstructions individual analyses of 18S and 28S data were done placing Diurodrilidae within a larger clade of Annelida (low bootstrap support) but never close to or even within any platyzoan taxon.

The phylogenetic analyses of mitochondrial genome sequence data recovered monophyletic Annelida with robust bootstrap support and placed *Diurodrilus subterraneus* as part of Annelida with as sister to Myzostomida with strong bootstrap support. Myzostomida exhibited long branches in these analyses. To test for long-branch attraction of *Diurodrilus* to Myzostomida, we excluded Myzostomida. These analyses placed *Diurodrilus* within monophyletic Annelida with significant bootstrap support as sister to Orbiinidae. However, the bootstrap support for the latter placement was low. In accordance with Worsaae and Rouse (2008) there was no support for a closer relationship to Dinophilidae.

The analyses of the mitochondrial gene order of *Diurodrilus subterraneus* showed that the gene order of *D. subterraneus* is very similar to the gene order observed in the majority of Annelida (Fig. 7). Even though, using mitochondrial gene order *Diurodrilus subterraneus* could not be placed confidentially in either Sedentaria or Errantia (Struck, 2011; Struck et al., 2011). Furhtermore, the mitochondrial gene order of the platyzoan taxa substantially differs from the one observed in *D. subterraneus* (Fig. 7).

Based on these phylogenetic analyses as well as the analyses of the mitochondrial gene order, the exclusion of Diurodrilidae from Annelida and a basal placement within Spiralia, as assumed by Worsaae and Rouse (2008), can be rejected without doubt. In the nuclear rRNA analyses of Worsaae and Rouse 2008 Diurodrilidae exhibited a long branch as it known for platyzoan taxa. This indicates that the phylogenetic affiliation of Platyzoa and Diurodrilidae most likely is caused by long-branch attraction. A similar case is also known for Myzostomida, a taxon which

also shows highly increased substitution rates (for discussion see Eeckhaut et al., 2000; Zrzavy et al., 1998; Zrzavy et al., 2001).

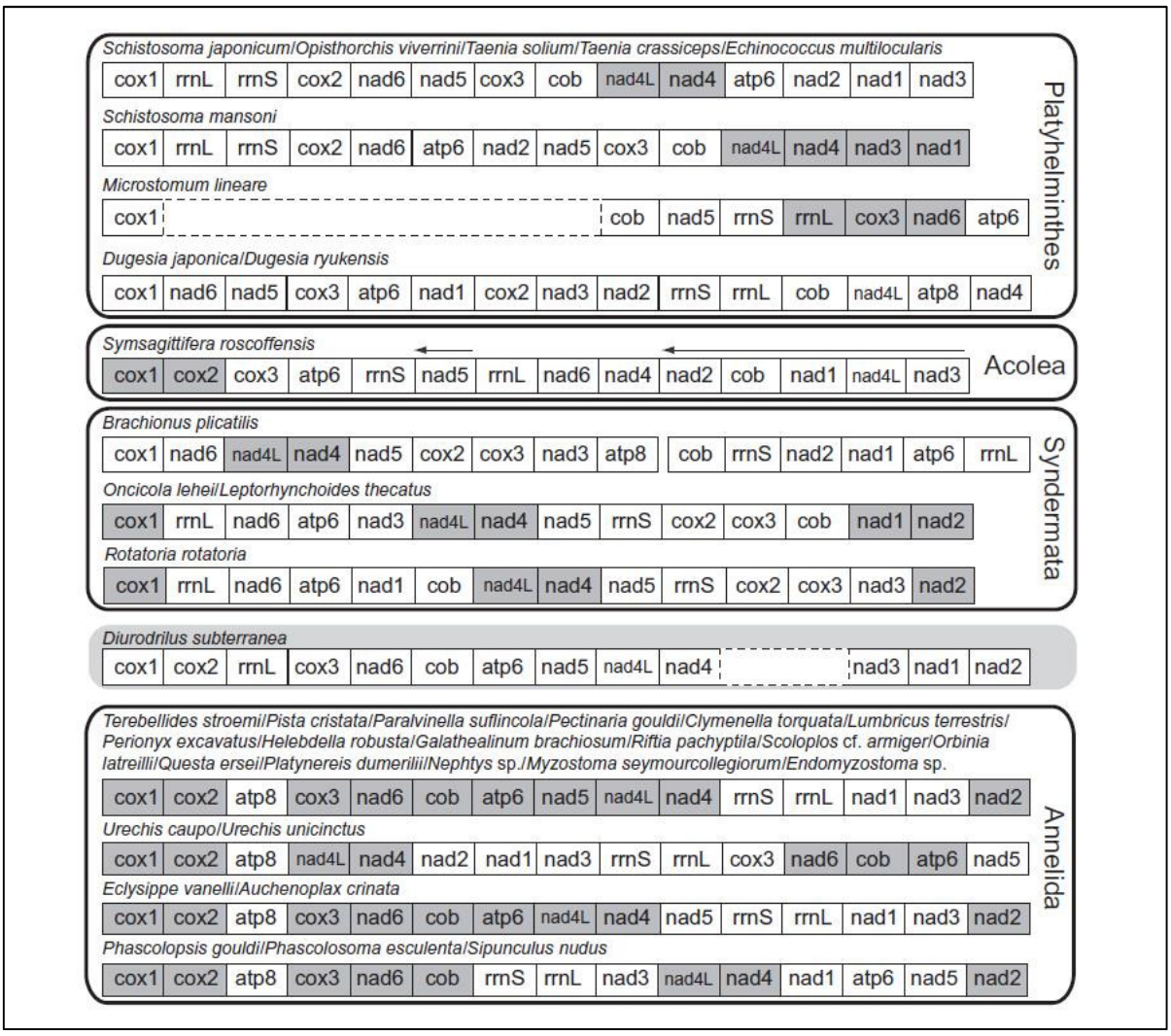


Fig. 7: Comparison of the mitochondrial protein-coding and rRNA gene order of Platyhelminthes, Acoela, Syndermata and Annelida with the gene order of *Diurodrilus subterraneus* (from Golombek et al., 2013). Blocks of identical gene order in *Diurodrilus subterraneus* and with another of the gene orders are highlighted in grey. nad1-6, 4L = NADH dehydrogenase subunits 1-6 and 4L; cox1-3 = cytochrome oxidase subunits 1-3; cob = cytochrome b; atp6 = ATP synthase subunit; rrnS = small ribosomal RNA (also known as 12S); rrnL = large ribosomal subunit (also known as 16S).

However, there are several comprehensive morphological studies which support a placement of Myzostomida within Annelida (Bleidorn, 2008; Bleidorn et al., 2007; 2009a; Eeckhaut and Lanterbecq, 2005; Helm et al., 2012; Lanterbecq et al., 2008; Müller and Westheide, 2000). As shown here for *Diurodrilus*, in most molecular studies Myzostomida was trapped within Platyzoa due to long-branch artifacts. The usage of more slowly evolving molecular data such as mitochondrial genomes rescued Myzostomida within Annelida (also called myzostomid

trap) (Bleidorn et al., 2007; 2009b). In this study, the mitochondrial genome sequence data was able to rescue Diurodrilidae from the placing with Platyzoa due to long-branch artifacts, and again placed them within Annelida, instead.

Progenetic evolution could explain the lack of annelid key characters such as head appendages, chaetae, parapodia, segmentation and nuchal organs in Diurodrilidae as it is already known for several other annelid taxa, e.g. Dinophilidae (e.g., Struck, 2006; Westheide, 1987). The possibility of a progenetic origin of Diurodrilidae from juvenile stages of a larger annelid ancestor has to be investigated to resolve the phylogenetic relationship of Diurodrilidae within Annelida. Moreover, loss of segmentation is observed for other taxa as well, e.g., Sipuncula and Echiura, which have formerly been placed as phyla separate outside from Annelida (Fauchald and Rouse, 1997; Halanych, 2004), whereas molecular analyses included both within the annelid radiation with strong support (e.g., Dordel et al., 2010; Struck et al., 2007, 2011).

4.5.3. **"Evolution of mitochondrial gene order in Annelida"** (Weigert et al., 2016)

The investigation and comparison of mitochondrial gene arrangements is a useful tool to resolve more issues concerning annelid evolution and phylogenetic relationships. Species which are closely related to each other often share identical unchanged gene orders and gene rearrangements rarely occur independently in different lineages (Boore, 1999; Boore and Brown, 1994).

Mitochondrial genomes of Spiralia (Lophotrochozoa sensu lato) show a high variability concerning gene number, gene arrangements, transcription from only one or both strands, repetitive and intergenetic regions, and unusual modes of inheritance (Boore, 1999; Vallès and Boore, 2006). In annelids the mitochondrial gene order is highly conserved with exception of Sipuncula, Echiura, Ampharetidae, Diurodrilidae and Eunicidae (e.g., Bleidorn et al., 2006; Golombek et al., 2013; Li et al., 2016; Vallès and Boore, 2006). Furthermore, the mitochondrial gene order of basal branching annelids such as Oweniidae, Magelonidae, Amphinomidae, and Chaetopteridae is substantially different from the assumed ground pattern which thus seems to be restricted to Pleistoannelida (Weigert et al., 2016). Therefore, it should be taken as the ground pattern of Pleistoannelida rather than for Annelida as a whole. It is known that gene rearrangements in Annelida have occurred less often than in other Spiralia (e.g., Jennings and

Halanych, 2005; Osca et al., 2014). The reconstruction of a putative mitochondrial ground pattern of annelids remains difficult.

The majority of annelids are classified into two major clades: Errantia and Sedentaria, together forming Pleistoannelida (Struck, 2011; Struck et al., 2011). At the basal part of the annelid tree Oweniidae and Magelonidae, Chaetopteridae as well as Sipuncula and, Amphinomidae were found (Weigert et al., 2014). For Sedentaria and Errantia complete mitochondrial genomes are available exhibiting a highly conserved gene order in most cases (e.g., Golombek et al., 2013). To get more insights into the evolution of mitochondrial gene order arrangements in annelids and to determine the mitochondrial gene order ground patteren of Annelida a higher coverage of mitochondrial genome data of the basal branching lineages is needed. Only two complete mitochondrial genomes of Sipunculida are available so far.

In this study the first complete mitochondrial genome sequences for all remaining basal branching annelids were generated and described: *Owenia fusiformis* (Oweniidae), *Magelona mirabilis* (Magelonidae), *Eurythoe complanata* (Amphinomidae), *Chaetopterus variopedatus* and *Phyllochaetopterus* sp. (Chaetopteridae).

The mitochondrial gene orders of the investigated basal branching annelids are significantly different from each other and are substantially different from the pattern found in Pleistoannelida (Fig. 8). Based on this, the assumption of a highly conserved gene order in all annelids has to be rejected, because it is restricted to Pleistoannelida representing the ground pattern of this group. In annelids, mitochondrial genes are generally transcribed only from one strand However, in *Owenia fusiformis* and *Magelona mirabilis* the two tRNAs encoding for proline (tRNA-P) and threonine (tRNA-T) are located on the opposite strand (‘-‘strand). With regard to the mitochondrial gene order pattern in the last common ancestor of annelids two hypotheses can be raised: (1) the last common ancestor of annelids had lost the transcription signal on the other strand and all genes are locatedon one strand, or (2) the last common ancestor of annelids had the transcription signal on both strands and lost the signal on one strand in the lineage leading to the remaining annelid taxa, whereas Oweniidae and Magelonidae would be sister group to this clade. Consequently, these have been united as Palaeoannelida by Weigert and Bleidorn (2016). This and the outgroup comparison imply, that the last common ancestor of annelids possessed the same mitochondrial pattern as observed in Oweniidae and Magelonidae meaning that the strand usage and inversion of both tRNAs is a plesiomorphic condition in annelid rather than a synapomorphic feature for Oweniidae and Magelonidae. To resolve this issue investigations concerning the mitochondrial gene order in other spiralian taxa

are necessary to find out if this pattern represents an ancestral condition for Annelida.
Another interesting feature, which could not be observed in annelids before, is the conjecture of the ATP8 to ATP6 gene found in *Magelona mirabilis* (Fig. 8). This is usually a common order in almost all animal mitochondrial genomes with exception of certain spiralian phyla such as Mollusca, Brachiopoda, Nemertea, and Phoronida where this gene boundary is disrupted (see Boore, 2006 and references therein). In Platyhelminthes, Acoelomorpha and Acanthocephala ATP8 (e.g., Mwinyi et al., 2010; Vallès andBoore, 2006), and in Chaetognatha both genes are missing (Papillon et al., 2004). Based on the phylogenetic position of Magelonidae, the ATP8-ATP6 conjecture could be part of the mitochondrial ground pattern.
It also has to be mentioned, that among all basal branching annelids, the mitochondrial gene order of *M. mirabilis* is most similar to the putative ground pattern of Spiralia (Bernt et al., 2013b).

Additionally, phylogenetic analyses based on mitochondrial sequence data were performed analysing two data sets: (1) a data set including all annelid taxa of which all 13 protein-coding genes were available, and (2) a data set including also annelid taxa with partial mitochondrial genomes. Maximum likelihood estimations (ML) and Bayesian inference (BI) were done. All analyses of both data sets could not recover the monophyly of Annelida due to the grouping of *Owenia fusiformis* with Nemertea. Furthermore, despite the position of Oweniidae, all analyses of complete mitochondrial genomes recovered well-supported monophyletic Errantia, Sedentaria and Pleistoannelida, as well as the basal branching annelids in agreement with previous phylogenomic studies (Struck et al., 2011; Weigert et al., 2014). However, all analyses of the data set including also partial mitochondrial genomes, could not resolve annelid relationships as described above, mainly due to the inclusion of fast evolving taxa like Myzostomida and Diurodrilidae which are characterized by long branches).
Although using mitochondrial sequence data to reconstruct ancient annelid relationships has its limits, comparing mitochondrial gene order still seems to be a promising tool for phylogenetic implications. Mitochondrial gene order can be used as an additional phylogentic marker especially if morphological or molecular data are controversial or lacking (e.g., Golombek et al., 2013; Mwinyi et al., 2009).

To understand the evolutionary relationships within Annelida the phylogeny of early branching annelids is important especially with regard to other spiralian taxa. Also, several spiralian taxa show a high variability in mitochondrial gene rearrangements at least in their basal radiation (e.g., Boore, 1999; Vallès and Boore, 2006). All of this could give new implications about a

putative mitochondrial ground pattern of the last common ancestor of Spiralia and Bilateria respectively.

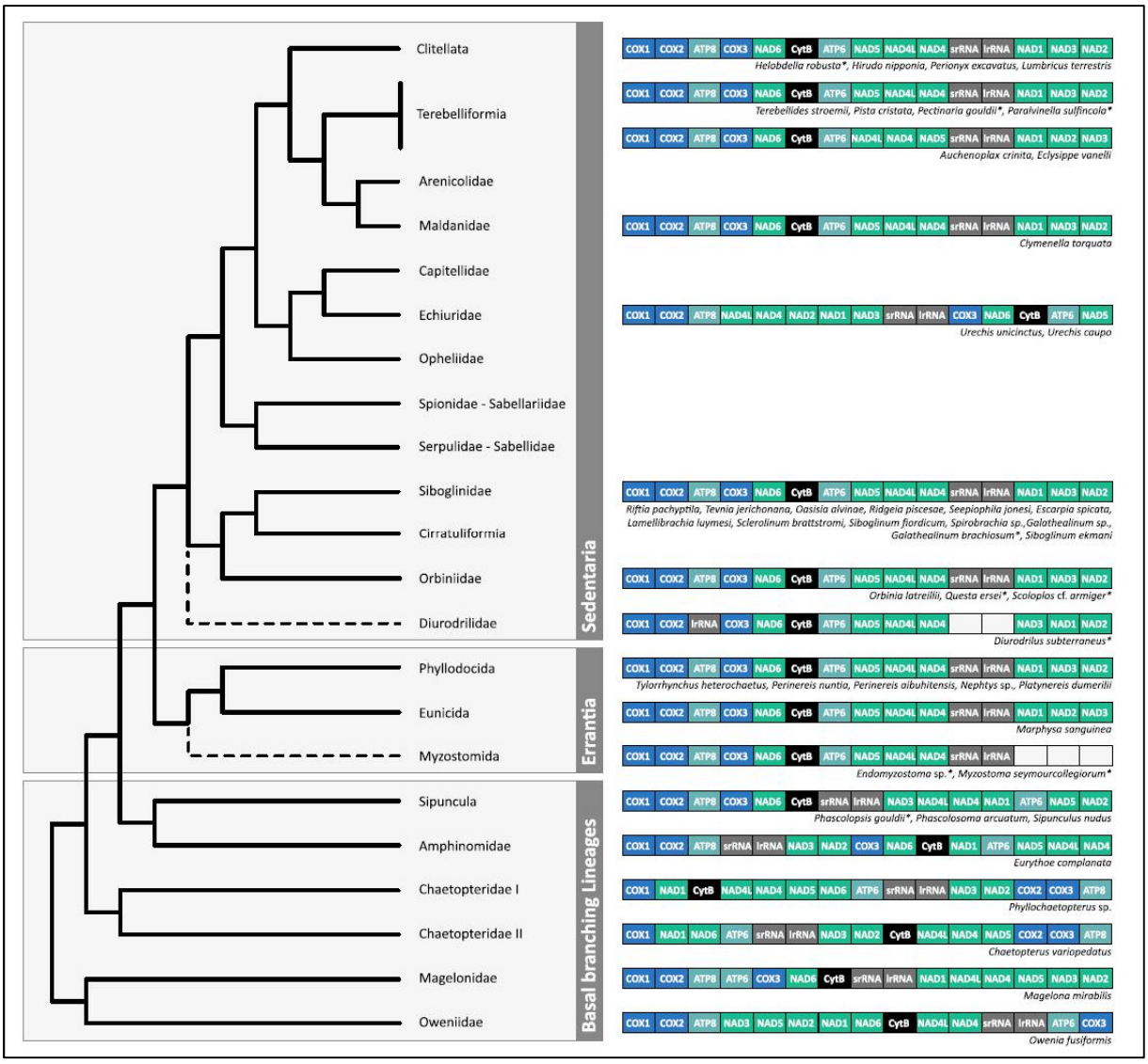


Fig. 8: Relationships within Annelida and different mitochondrial gene order of each taxon (from Weigert et al., 2016). The annelid phylogeny is depicted based on Weigert et al. (2014) and sister group relationships of families which are not represented in Weigert et al. (2014) are obtained from Struck et al. (2007) (Maldanidae and Ampharetidae) and Golombek et al. (2013) (Diurodrilidae). An uncertain phylogenetic position is indicated by dashed lines. Taxa with partial mitochondrial genomes are marked with an asterisk and missing genes within the mitochondrial gene order are indicated with gray boxes. nad1-6, 4L = NADH dehydrogenase subunits 1-6 and 4L; cox1-3 = cytochrome oxidase subunits 1-3; cob = cytochrome b; atp6 = ATP synthase subunit; srRNA = small ribosomal RNA (also known as 12S); lrRNA = large ribosomal subunit (also known as 16S).

4.5.4. "Syllidae mitochondrial gene order is unusually variable in Annelida" (Aguado et al., 2016)

As mentioned above, Annelida is a highly diverse animal phylum with over 21,000 described species (Rouse and Pleijel, 2001). Within this animal phylum there are annelid taxa with a high number of species such as, e.g., Terebelliformia (> 800 species) or Syllidae (> 700 species).

Syllidae (Phyllodocida, Errantia) represents one of the largest groups within Annelida inhabiting practically all marine benthic realms (Aguado et al., 2015a). The phylogeny within Syllidae was analyses in various morphological and molecular studies. Initially, phylogenetic relationships of Syllidae were studied including only few taxa (e.g., Nygren, 1999; Nygren and Sundberg, 2003), or focusing on specific groups (e.g., Licher, 1999; Nygren, 2004). All analyses of Aguado et al. (2012, 2015a, 2015b) recovered the Syllidae as monophyletic group divided in two major clades: a clade corresponding to the subfamiliy Anoplosyllinae and the other clade comprising the remaining syllids. Interestingly, within the second clade the "traditional" subfamilies such as Autolytinae, Exogoninae, Syllinae, and Eusyllinae were recoverd as monophyletic. However not all genera could be assigned to one of these subfamilies and remained as independent groups (Aguado et al., 2012).

For Syllidae, the available molecular information is restricted to three genes: *18S*, *rrnL* and *cox1*. So far, only two mitochondrial genomes of syllids (*Ramisyllis multicaudata* and *Trypanobia cryptica*) have been investigated (Aguado et al., 2015a). Interestingly, the mitochondrial gene order of these two syllids was completely different from the proposed ground pattern hypothesized for Pleistoannelida (Aguado et al., 2015a).

To clarify whether Syllidae in general or only the two investigated members of *Ramisyllis* and *Trypanobia* show these aberrant mitochondrial gene order, the complete mitochondrial genomes of five additional syllids species have been generated, described and analyzed in this study: *Streptosyllis* sp. (Anoplosyllinae), *Eusyllis blomstrandi* (Eusyllinae), *Myrianida brachycephala* (Autolytinae), *Typosyllis antoni* and *Typosyllis* sp. (Syllidae). Furthermore, based on these data it was hoped to get new insights into mitochondrial gene order evolution and possible divergence scenarios within Syllidae.

In all analyses of the mitochondrial genomes of the five syllids, all 13 protein-coding genes, two ribosomal RNAs and 22 tRNAs could be detected. The mt genomes of syllids are AT-rich and all genes are transcribed from one strand.

In addition, mt genomes of syllids show a pattern of codon usage bias which is similar to other annelids (e.g., Bleidorn et al., 2006; Jennings and Halanych, 2005; Zhong et al., 2011). Even tRNAs mostly possess the common cloverleaf structure. Comparing mitogenomic features of the analyzed syllids, such as genome length, skewness, codon bias, and especially gene content, the most similarities to the rest of annelids have been found in *Eusyllis blomstrandi* and *Myrianida brachycephala* (Fig. 9). The remaining syllid mt genomes of *Streptosyllis* sp., *R. multicaudata*, *T. cryptica*, *T. antoni* and *Typosyllis* sp. are the most dissimilar ones (Fig. 9).

Additionally, phylogenetic analyses have been performed to get more evidence for phylogenetic relationships within Syllidae. In this study the phylogenetic analyses are in agreement with previous phylogenetic analyses (Aguado et al., 2012, 2015a, 2015b) based on only three genes: *18S*, *rrnL* and *cox1*. Syllidae have been recovered as well-supported monophyletic group (Fig. 9). Also Anoplosyllinae was recovered as part of Syllinae usually as sister group to the remaining syllid taxa or as part of a basal polytomy (Fig. 9).

In this study, only the mitochondrial gene orders of *Eusyllis blomstrandi* (Euysllinae) and *Myrianida brachycephala* (Autolytinae) show similarities to the ground pattern found in Pleistoannelida (Fig. 9). In contrast, the gene orders of *Streptosyllis* sp. (Anoplosyllinae) and especially in the Syllinae show extremely different patterns to that found in Pleistoannelida (Figs. 9). These results imply that the gene order of Pleistoannelida could be more diverse than expected.

Syllinae represents the most complex group within Syllidae, and their investigated mt genomes do not show a clear common pattern. Also closely related species to each other (*Typosyllis* sp. and *T. antoni*, respectively *R. multicaudata* and *T. cryptica*) show more gene rearrangements as usually found across Pleistoannelida at all. This indicates that the gene order might have changed at least two times within Syllidae: within Syllinae and in *Streptosyllis* sp.

Furthermore, these results show that the mitochondrial gene order is phylogenetically informative in this clade. Although diverse mitochondrial gene orders are also known within tunicates, molluscs and brachiopods (e.g., Luo et al., 2015; Stach et al., 2010; Yuan et al., 2012), but the question remains why the evolution of mitochondrial genomes is highly dynamic within this Syllidae.

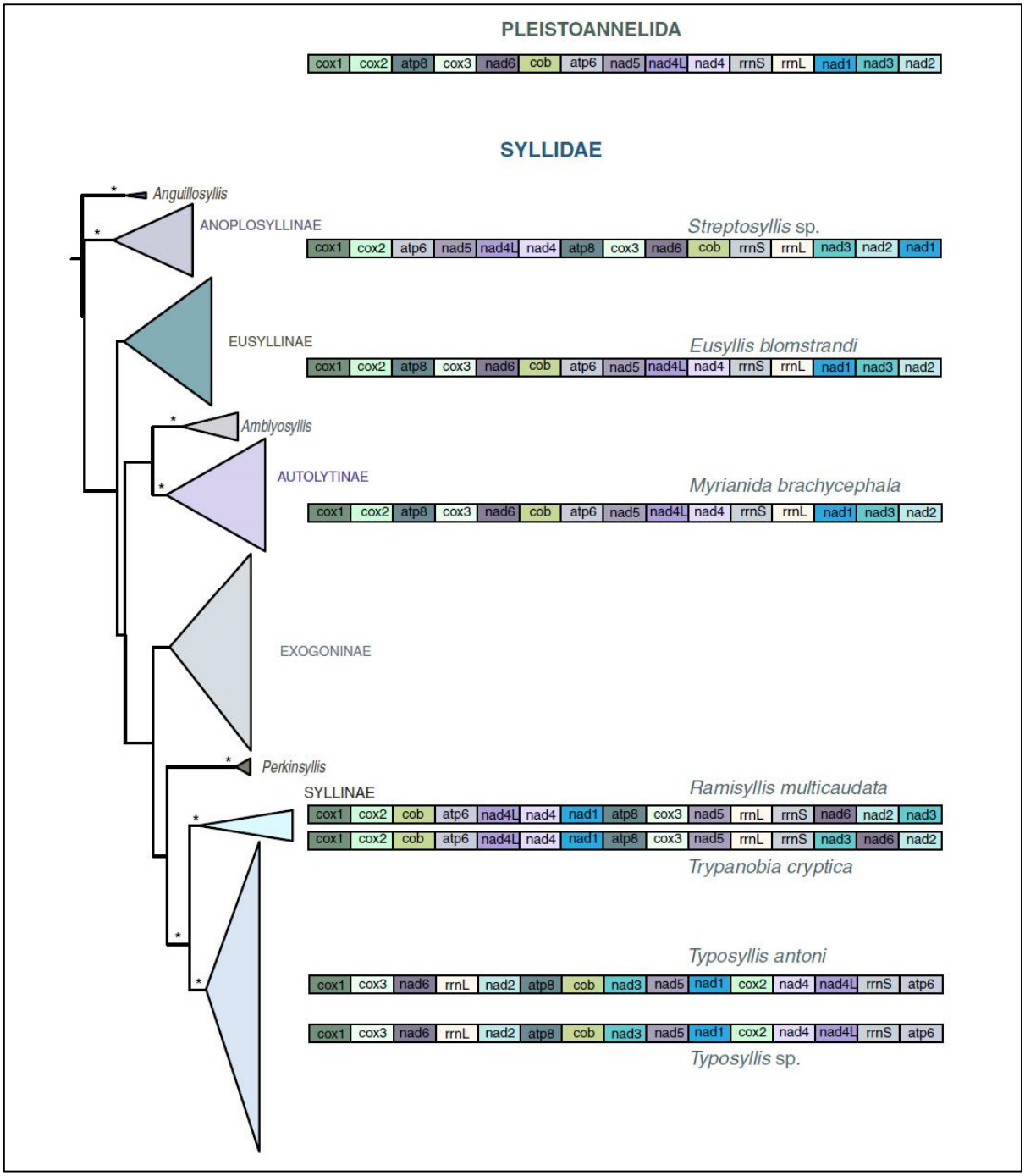


Fig. 9: Gene order in Syllidae (from Aguado et al., 2016). Mitochondrial genomes represented without tRNAs. The phylogenetic relationships based on Aguado et al. (2015a) and the gene order in Pleistoannelida based on Weigert et al. (2016). An asterisk indicates well supported clades. nad1-6, 4L = NADH dehydrogenase subunits 1-6 and 4L; cox1-3 = cytochrome oxidase subunits 1-3; cob = cytochrome b; atp6 = ATP synthase subunit; rrnS = small ribosomal RNA (also known as 12S); rrnL = large ribosomal subunit (also known as 16S).

4.5.5. "The evolution of annelids reveals two adaptive routes of the interstitial realm" (Struck et al., 2015)

Besides many macrofauna species numerous meiobenthic annelid species also exist, generally only a few millimeters in length. Part of them inhabits the interstitium, i.e., the space between the sand grains (Giere, 2009; Noodt, 1974). Whereas some species are more or less miniaturized examples of their larger relatives such as certain Syllidae, other species exhibit a very different morphology resembling larvae or juveniles. Due to their apparently simple organization these interstitial annelid species have been regarded to be very close to the annelid stem species due to lack of many annelid-specific characters including absence of parapodia and chaetae in certain taxa (Hermans, 1969). The taxa united as "Archiannelida" generally comprise Dinophilidae, Diurodrilidae, Nerillidae, Polygordiidae, Protodrilidae and Protodriloididae. Later on, it was generally accepted that they constitute a polyphyletic assemblage of highly derived but secondarily simplified annelid taxa adapted to their interstitial realm, which evolved independently and were only grouped together because of a secondary simplified body plan (e.g., Fauchald, 1974; Purschke, 1985a, b; Purschke and Jouin, 1988; Rouse and Fauchald, 1997; Struck et al., 2002; 2005; 2008; Tzetlin and Purschke, 2005; Westheide, 1985; 1987; 1990; Worsaae and Kristensen, 2005; Worsaae and Rouse, 2008).
The "Archiannelida"-concept exhibit that the annelid taxa mentioned above represent ancestral conditions of Annelida (Hermans, 1969; Struck, 2012). As already discussed above, Weigert et al. (2014) placed Oweniidae and Magelonidae, two non-interstitial macrobenthic taxa, at the base of the annelid tree. However, the question remained how these animals evolved, whether they are monophyletic or polyphyletic and which are their sister groups.

There are different evolutionary scenarios which may explain the existence of interstitial animals. **First**, the interstitial realm is seen as the ancestral habitat of bilaterians as shown by extant gastrotrichs and gnathostomulids (Hejnol et al., 2009; Struck et al., 2014a). This implies that the last common ancestor of Bilateria must have been a small species. **Second**, the interstitial taxa evolved by miniaturization from larger ancestors. The hypothesis is that larger adult ancestors, with an infaunal or epibenthic life cycle, colonized the interstitium via step-by-step decrease in body size (Westheide, 1987). The **third** scenario describes that interstitial taxa may have evolved from larger ancestors by paedomorphosis (Weistheide, 1987; Struck, 2006). Paedomorphosis, is the retention of ancestral larval or juvenile stages of descendents and can arise either by a retardation of somatic development (neotony) or by an acceleration of the

sexual maturation (progenesis) (Gould, 1977). Moreover, paedomorphosis (especially progenesis) has been regarded as the major evolutionary process responsible for the permanent colonization of interstitial habitats by meiofauna organisms in general (Struck, 2006; Worsaae and Kristensen, 2005; Westheide, 1987).

Resolving the phylogenetic relationships of the former archiannelid taxa might help to understand the evolution of annelids and to reconstruct the ground pattern of the last common ancestor of annelids. The general assumption for all former archiannelid taxa is that progenesis has to be assumed as evolutionary process (e.g., Eibye-Jacobsen and Kristensen, 1994; Golombek et al., 2013; Struck, 2006; Worsaae and Kristensen, 2005; Westheide, 1987). Assuming progenesis based on morphological data alone entails the risk of one-dimensional phylogenetic reconstructions (Struck, 2006), because characters of adult stages are exclusively compared with larval or juvenile characters and not with adult characters of other taxa by what progenesis has to be subsequently concluded (Wiens et al., 2005). Certain phylogenetic analyses based on molecular data placed 'archiannelid' taxa as part of the basal annelid radiation (Struck, 2012; Struck et al., 2008). Albeit the nodal support for this phylogenetic placement was only weak, these results brought back the possibility of an interstitial ancestry of Annelida.

To investigate, whether interstitial annelids show a putative ancestral condition of Annelida, or whether these interstitial taxa evolved by miniaturization or progenesis, a phylogeneomic approach was conducted including 12 interstitial 'archiannelid' species from Dinophilidae, Diurodrilidae, Nerillidae, Protodrilida, Polygordiidae, and *Apharyngtus* as well as two additional interstitial annelid species from Parergodrilidae.

The phylogenetic reconstructions in this study significantly supported the phylogenetic relationships within Annelida which already have been found in previous analyses (e.g., Struck et al., 2011; Weigert et al., 2014) (Fig. 10). The only exception in this study is the placement of Cirratuliformia/Siboglindae as sister to Sabellida/Spionida (Fig. 10), whereas in Weigert et al. (2014) the clade of Cirratuliformia/Siboglindae is sister to Orbiniidae.

The analysis of this study, placed the interstitial annelid taxa into two major groups with significant nodal support (Fig. 10). The first group consists of monophyletic Protodrilida and Polygordiidae which were placed in Errantia as sister group to the remaining Errantia (Eunicida and Phyllodocida). This is in contrast to morphology-based hypotheses placing neither Protodrilida nor Polygordiidae closely related to any errant taxon, respectively suggesting progenesis (Struck, 2006). The second group consists of *Apharyngtus*, Dinophilidae,

Diurodrilidae, and Nerillidae which were placed together with Orbiniidae and Parergodrilidae as Orbiniida within Sedentaria. This is also in contrast to the morphological-based assumption of a closer relationship of Dinophilidae to Errantia or its subtaxon Eunicida, also assuming progenesis (Struck, 2006, 2012).

Additionally, we checked if reconstruction artifacts such as paralogous sequences, cross-contamination, branch-length heterogenetiy, overall evolutionary rate, amino acid composition, compositional heterogenetiy, and shared missing data, affected the tree topology and the placement of interstitial taxa (e.g., Struck, 2013; Struck et al., 2014). All phylogenetic analyses based on these data sets recovered the same results with respect to the phylogeny of interstitial annelid taxa with strong nodal support.

Across different analyses, the relationships within the two clades including interstitial taxa were also stable. The monophyly of Protodrilida and Polygordiidae was recovered with strong bootsrap support. Furthermore, Parergodrilidae is always sister to Orbiniidae also with maximal support. *Apharyngtus* was placed as sister to Diurodrilidae, both forming the sister group to a clade of Orbiniidae/Parergodrilidae. Dinophilidae and Nerillidae have always been placed as sister to these four taxa. These results suggest that interstitial annelid taxa are not part of the basal radiation of lineages. Furthermore, these results indicate that inhabiting the interstitial realm and a simple body organization are not ancestral characters of Annelida, and that the "Archiannelida"-concept has to be rejected with certainty confirming previous studies (Struck, 2012; Westheide, 1987).

Based on close phylogenetic relationships of Nerillidae, Dinophilidae, Diurodrilide, and *Apharyngtus* to Orbiniidae and Parergodrilidae a new clade named Orbiniida was formed (Fig. 10). Comparing these four interstitial taxa with adult and juvenile stages of Orbiniidae show that these taxa possess morphological features which also can be found in larval and juvenile stages of Orbiniidae, which temporarily inhabit the interstitial realm (Rouse and Pleijel, 2001; Struck et al., 2005). The ancestral state reconstructions revealed that the last common ancestor of Orbiniida is more similar to the last common acestor of Sedentaria and Orbiniidae. Furthermore, within Orbiniidae several independent progenetic events are known (Bleidorn et al., 2009c). This indicates that Nerillidae, Dinophilidae, Diurodrilidae, and *Apharyngtus* probably evolved independently by progenesis from a larger ancestor, whereas strong evidence for Nerillidae is lacking (Westheide, 2008; Worsaae, 2005). Eventually, an alternative phylogenetic position for Nerillidae has to be assumed (unpublished data).

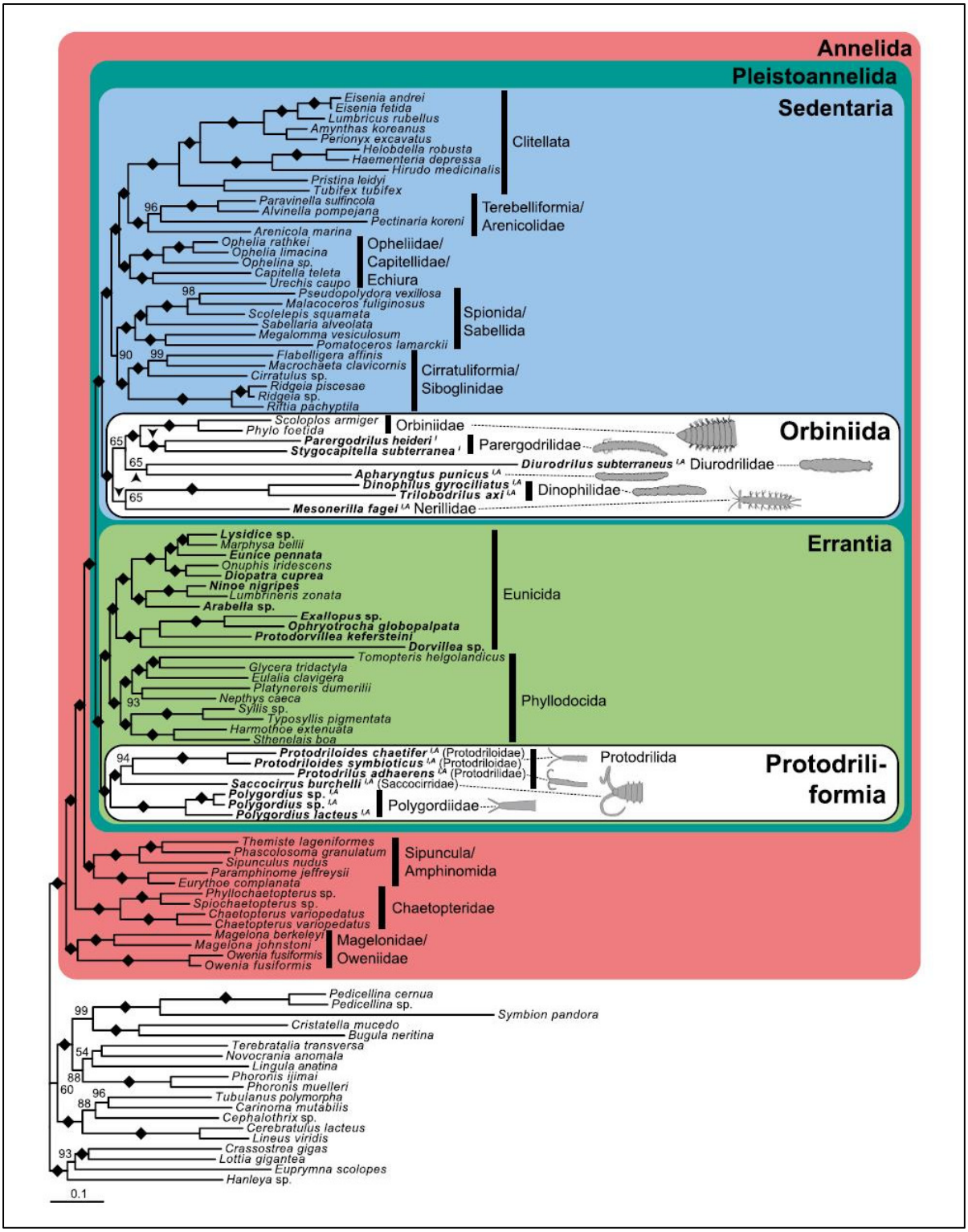


Fig. 10: Tree of maximum likelihood analysis of the largest dataset (dataset 1) with 100 species, 189,193 amino acid positions, and 41.2% sequence coverage using RAxML (from Struck et al., 2015). Bootstrap values above 50 are shown at the branches. Values of 100% are depicted as diamonds. Drawings of relevant taxa are displayed but are not to scale. Superscript I indicates interstitial species, and superscript A indicates former archiannelids. Arrowheads indicate possible events of progenesis.

Polygordiidae and Protodrilida show no morphological similarities to developmental stages of related taxa, therefore, a progenetic origin seems to be unlikely. Instead, miniaturization may explain the evolution of these taxa. This is further corroborated that a stepwise decrease of body size and loss of annelid-specific characters can be observed. For instance, several taxa are still large enough to reproduce via larval stages and others still possess parapodia and chaetae. The ancestral state reconstructions of Polygordiidae and Protodrilida revealed that the last common ancestor of these taxa was most likely saccocirrid-like evolving from an infaunal ancestor which inhabited coarse sediments. Because of this, the clade of Polygordiidae and Protodrilida was named Protodriliformia (Fig. 10).

Furthermore, similar to the last common ancestor of Annelida, the last common ancestor of Sedentaria was reconstructed to be a larger epibenthic or infaunal annelid characterized by a prostomium bearing palps and eyes, homonymous segments with parapodia and simple chaetae, and a pygidium with cirri (supplementary Tab. S4).

To conclude, generally evolution of interstitial annelid taxa by progenesis has been favored over miniaturization (Struck, 2006; Westheide, 1987). But based on the results of these analyses, progenetic evolution can be found in Orbiniida, and stepwise miniaturization occurs in Protodriliformia. Thus, there are two different evolutionary routes to adapt to the interstitium from larger ancestors.

5. Published articles

The following articles were included in the cumulative thesis submitted for review and differ slightly from the published journal articles.

- Bernt, M., Bleidorn, C., Braband, A., Dambach, J., Donath, A., Fritzsch, G., Golombek, A., Hadrys, H., Jühling, F., Meusemann, K., Middendorf, M., Misof, B., Perseke, M., Podsiadlowski, L., von Reumont, B., Schierwater, B., Schlegel, M., Schrödl, M., Simon, S., Stadler, P.F., Stöger, I., Struck, T.H. (2013).
  A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Molecular Phylogenetics and Evolution,* 69(2):352 – 364.
  http://dx.doi.org/10.1016/j.ympev.2013.05.002

- Perseke, M., Golombek, A., Schlegel, M., Struck, T.H. (2013).
  The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. *Molecular Phylogenetics and Evolution,* 66(3):898 – 905.
  http://dx.doi.org/10.1016/j.ympev.2012.11.019

- Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kück, P., Herlyn, H., Hankeln, T. (2014).
  Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of Spiralia. *Molecular Biology and Evolution,* 31(7):1833 – 1849.
  http://dx.doi.org/10.1093/molbev/msu143

- Golombek, A., Tobergte, S., Struck, T.H. (2015).
  Elucidating the phylogenetic position of Gnathostomulida and first mitochondrial genomes of Gnathostomulida, Gastrotricha and Polycladida (Platyhelminthes). *Molecular Phylogenetics and Evolution,* 86:49 – 63.
  http://dx.doi.org/10.1016/j.ympev.2015.02.013

- Zhong, M., Hansen, B., Nesnidal, M., Golombek, A., Halanych, K.M., Struck, T.H. (2011).
  Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. *BMC Evolutionary Biology,* 2011 11:369.
  http://dx.doi.org/10.1186/1471-2148-11-369

- Golombek, A., Tobergte, S., Nesnidal, M.P., Purschke, G., Struck, T.H. (2013). **Mitochondrial genomes to the rescue – Diurodrilidae in the myzostomid trap.** *Molecular Phylogenetics and Evolution,* 68(2):312 – 326. http://dx.doi.org/10.1016/j.ympev.2013.03.026

- Weigert, A., Golombek, A., Gerth, M., Schwarz, F., Struck, T.H., Bleidorn, C. (2016). **Evolution of mitochondrial gene order in Annelida.** *Molecular Phylogenetics and Evolution,* 94(A): 196 – 206. http://dx.doi.org/10.1016/j.ympev.2015.08.008

- Aguado, M.T., Richter, S., Sontowski, R., Golombek, A., Struck, T.H., Bleidorn, C. (2016). **Syllidae mitochondrial gene order is unusually variable for Annelida.** *Gene,* 594(1): 89 – 96. http://dx.doi.org/10.1016/j.gene.2016.08.050

- Struck, T.H., Golombek, A., Weigert, A., Franke, F.A., Westheide, W., Purschke, G., Bleidorn, C., Halanych, K.M. (2015). **The evolution of annelids reveals two adaptive routest o the interstitial realm.** *Current Biology,* 25(15):1993 – 1999. http://dx.doi.org/10.1016/j.cub.2015.06.007

5.1. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny

Matthias Bernt[a]*, Christoph Bleidorn[b], Anke Braband[c], Johannes Dambach[d], Alexander Donath[d], Guido Fritzsch[e], Anja Golombek[d], Heike Hadrys[f], Frank Jühling[e,g], Karen Meusemann[d], Martin Middendorf[a], Bernhard Misof[d], Marleen Perseke[h], Lars Podsiadlowski[i]*, Björn von Reumont[j], Bernd Schierwater[f], Martin Schlegel[b], Michael Schrödl[k], Sabrina Simon[l], Peter F. Stadler[e,m,n,o,p,q], Isabella Stöger[k], Torsten H. Struck[d]
(authors in alphabetical order)

Author affiliations
[a]*Parallel Computing and Complex Systems Group, Department of Computer Science, University of Leipzig, Augustusplatz 10, D-04109 Leipzig, Germany*
[b]*Molecular Evolution and Systematics of Animals, Institute of Biology, University of Leipzig, Talstraße 33, D-04103 Leipzig, Germany*
[c]*LGC Genomics GmbH, Ostendstr. 25, 12459 Berlin*
[d]*Centre for Molecular Biodiversity Research (ZMB), Zoologisches Forschungsmuseum Alexander Koenig (ZFMK), Adenauerallee 160, D-53113 Bonn, Germany*
[e]*Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*
[f]*ITZ, Ecology & Evolution, TiHo Hannover, Buenteweg 17d, 30559 Hannover, Germany*
[g]*Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC; Strasbourg, France*
[h]*Laboratory of Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Science, 164 West Xingang Road, 510301 Guangzhou, PR China*
[i]*Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, D-53113 Bonn, Germany*
[j]*Department of Life Sciences, The Natural History Museum, Cromwell Road SW7 5BD London, United Kingdom*
[k]*Zoologische Staatssammlung München, Münchhausenstraße 21, 81247 München, Germany*
[l]*Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA*
[m]*Max-Planck-Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany*
[n]*Fraunhofer Institut fuer Zelltherapie und Immunologie, Perlickstrasse 1, D-04103 Leipzig, Germany*
[o]*Department of Theoretical Chemistry, University of Vienna, Waehringerstrasse 17, A-1090 Wien, Austria*
[p]*Center for non-coding RNA in Technology and Health, University of Copenhagen, Grønnegardsvej 3, DK-1870 Frederiksberg, Denmark*
[q]*Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA*

*corresponding authors
Lars Podsiadlowski, Inst. Evolutionary Biology & Ecology, University of Bonn, An der Immenburg 1, D-53121 Bonn, Germany; fax +49228735129; email: lars@cgae.de
Matthias Bernt, Parallel Computing and Complex Systems Group, Department of Computer Science, University of Leipzig, Augustusplatz 10, D-04109 Leipzig, Germany; email: bernt@informatik.uni-leipzig.de

Abstract

About 2800 mitochondrial genomes of Metazoa are present in NCBI RefSeq today, two thirds belonging to vertebrates. Metazoan phylogeny was recently challenged by large scale EST approaches (phylogenomics), stabilizing classical nodes while simultaneously supporting new sistergroup hypotheses. The use of mitochondrial data in deep phylogeny analyses was often criticized because of high substitution rates on nucleotides, large differences in amino acid substitution rate between taxa, and biases in nucleotide frequencies. Nevertheless, mitochondrial genome data might still be promising as it allows for a larger taxon sampling, while presenting a smaller amount of sequence information. We present the most comprehensive analysis of bilaterian relationships based on mitochondrial genome data. The analyzed data set comprises more than 650 mitochondrial genomes that have been chosen to

represent a profound sample of the phylogenetic as well as sequence diversity. The results are based on high quality amino acid alignments obtained from a complete reannotation of the mitogenomic sequences from NCBI RefSeq database. However, the results failed to give support for many otherwise undisputed high-ranking taxa, like Mollusca, Hexapoda, Arthropoda, and suffer from extreme long branches of Nematoda, Platyhelminthes, and some other taxa. In order to identify the sources of misleading phylogenetic signals, we discuss several problems associated with mitochondrial genome data sets, e.g. the nucleotide and amino acid landscapes and a strong correlation of gene rearrangements with long branches.

Key words: Mitochondrial genomes, animal phylogeny

1. Introduction

The suitability of molecular markers for phylogenetic analysis can be evaluated according to a set of criteria (Cruickshank, 2002). (1) The orthology criterion should be fulfilled, meaning that the changes between gene sequences are results of underlying speciation events and not of gene duplication events (as is the case when comparing paralogous genes). Although orthology is often quickly assumed in phylogenetic datasets, its prediction is a non-trivial task and became an important part of phylogenomic approaches (Altenhoff and Dessimoz, 2012). (2) Marker genes should be present in all taxa under study. Thus, "housekeeping genes", responsible for basal cell functions and thus common to a wide array of organisms, were widely used in phylogenetics. Nevertheless, current phylogenomic studies often work with rather patchy data matrices, e.g. sets of genes derived from EST approaches which have a varying degree of incompleteness with respect to the whole matrix (Dunn et al., 2008; Pick et al., 2010). (3) Selection should only act as a stabilizing factor on marker genes. Otherwise phylogenetic signal may be blurred by positive or negative selection, e.g. by homoplasious changes in different taxa with similar selection pressure and by a strong difference of substitution rates depending on the strength of selective force. Again "housekeeping genes" seem to be a good choice, having the same functional role in basal cellular mechanisms of many organisms and being optimized for their functions long before the basal splits of the group under study (Butte et al., 2001). A recent study demonstrates that slowly evolving genes involved in the translation process provide best results in resolving basal metazoan relationships (Nosenko et al., 2013). (4) Ideal genetic markers exhibit constant character state frequencies (nucleotides or amino acids) and substitution rates in all studied lineages over time. However, these features are rarely met by real data sets. (5) Finally, a good mixture of conserved and variable parts must be present in the alignment. While conserved segments allow the construction of PCR primer sets suitable for many species and are important to obtain reliable sequence alignments, variable sites or segments provide a sufficient amount of phylogenetic signal.

At first glance animal mitochondrial genomes seem to fulfill most of these criteria. Gene duplications in mitochondrial genomes occur rarely. Therefore orthology prediction is apparently an easy task, especially for complete mitochondrial genomes. But the frequent detection of non-functional nuclear copies of mitochondrial sequences (numts) weakened this view. Identifying numts is an important problem, especially when only fragments of single mitochondrial genes are used as phylogenetic markers (Bensasson et al., 2001). Mitochondrial genomes are present in all Metazoa (with the single known exception of the Loricifera

(Danovaro et al., 2010)) and contain an almost perfectly conserved complement of "housekeeping genes". Their comparatively high mutation rate and a good mixture of conserved and variable sites facilitate the use of universal primer sets, enable unambiguous alignments, and provide sufficient phylogenetic signal (Moritz et al., 1987). In addition the lack of recombination and the strictly maternal mode of inheritance (for exceptions see Bernt et al., 2013a) make mitochondrial markers as well suitable to infer population structure (Avise, 2000). Currently (October 1st, 2012) 2765 mitochondrial genomes of Metazoa were present in NCBI RefSeq database, covering 1829 (66%) vertebrate species. About one half (479) of the remaining 936 entries are from arthropod species. However, complete mitochondrial genomes are available for most animal phyla. In comparison to phylogenomic datasets mitochondrial genome data still allow a larger taxon sampling for most of the animal phyla. But they include a much smaller amount of sequence information. Moreover, working with complete mitochondrial genomes enables the additional analysis of features like gene content and gene order.

Thus, animal mitochondrial genome data have been widely used addressing phylogenetic questions ranging from population to phylum level (Avise, 2000). With an increasing number of studies the limits and problems of mitochondrial data became more evident and its value for phylogenetic analyses was criticized for specific points or even in general (Ballard and Whitlock, 2004; Ballard and Rand, 2005; Galtier et al., 2009). Notable points are large divergence of substitution rates and base composition between taxa, the already mentioned presence of "numts", and frequent occurrence of mitochondrial introgression. Nevertheless, mitochondrial genome data often proved its value in phylogenetic studies (Rubinoff and Holland, 2005).

State-of-the-art in animal phylogenetics is the analysis of large multilocus datasets, derived from whole genomes or large scale EST approaches ("phylogenomics") (Hausdorf et al., 2007; Dunn et al., 2008; Philippe et al., 2009; Hejnol et al., 2009; Pick et al., 2010). These analyses largely confirmed the "new animal phylogeny" (Halanych, 1995; Aguinaldo et al., 1997; Adoutte et al., 2000; Halanych, 2004), with Bilateria subdivided into the major subtaxa Lophotrochozoa, Ecdysozoa, and Deuterostomia. However, a number of small phyla failed to be placed within this framework. The internal phylogeny of Lophotrochozoa and Ecdysozoa and the basal relationships between non-bilaterian taxa and Bilateria are far from being consistent between different published studies (e.g. Srivastava et al., 2008; Hejnol et al., 2009; Philippe et al., 2009; Pick et al., 2010; Philippe et al., 2011; Nosenko et al., 2013). To complement these approaches with a comparatively small set of genes, but larger taxon sampling, we exploit a comprehensive mitogenomic dataset for an analysis of metazoan phylogeny.

Former phylogenetic analyses of metazoan mitochondrial genomes with small taxon samplings frequently resulted in trees with problematic long branches (e.g. Nematoda, Platyhelminthes) and supported some barely reliable sistergroup relations (e.g. Hassanin et al., 2005; Steinauer et al., 2005; Yokobori et al., 2008; Jang and Hwang, 2009; Mwinyi et al., 2010). However, a broad comprehensive analysis was missing, which will clearly illustrate the prospects and limits of mitochondrial genome data in metazoan phylogenetics.

Here we present the most comprehensive analysis of metazoan phylogeny based on mitochondrial genome data, involving most invertebrate species with a RefSeq entry for a complete mitochondrial genome and a selection of vertebrate species. Together with outgroup taxa from fungi and protists we analyzed a dataset comprising more than 800 mitochondrial genomes. A new optimized automated annotation pipeline was set up to overcome annotation errors known to be widespread in NCBI RefSeq entries of mitochondrial genomes (Bernt et al., 2013b). Alignments of protein-coding genes were subject to carefully modeled ML analyses. Inconsistencies between phylogenetic analyses of nuclear genes and our results, as well as an overview concerning mitochondrial gene orders, will be discussed in more detail in the taxon-specific reviews (other articles in this special issue). Here we focus on the general landscape of mitochondrial genome variation in Metazoa and the problems resulting from departures of the above mentioned criteria of ideal phylogenetic markers.

2. Materials and methods

2.1. Data

The analyses are based on all metazoan mitogenome sequences in RefSeq (Pruitt et al., 2007) release 41, excluding the sequence of *Anopheles funestus* (NC_008070), which consists of 27.5% non-standard bases. In addition, the mitochondrial genome sequences of four metazoan species which have been added to RefSeq recently plus a few new and so far unpublished mitochondrial sequences of metazoan species (see supplementary material) were added. We used the mitochondrial genome sequences of 20 fungi species from RefSeq release 41 and eight contributed other non-metazoan eukaryote species as outgroup representatives (see supplementary material).

The phylogenetic reconstruction is solely based on protein coding genes. In order to avoid potential inconsistencies or errors in the published annotations (e.g. Boore, 2006) we re-annotated all sequences using the protein prediction pipeline of MITOS (Bernt et al., 2013b). For each protein coding gene the MITOS prediction with the best quality value was used to extract the corresponding amino acid sequence. For the two species with a mitogenome consisting of two sequences, i.e. *Hydra magnipapillata* (NC 011221, NC 011220) and *Brachionus plicatilis* (NC 010472, NC 010484), the best prediction from the two sequences was taken. This affects only the three genes *cox1*, *nad4*, and *nad6* where a prediction was made by MITOS for both mitochondrial genome sequences. In each case the values for the quality scores of the best predictions for the two sequences differ by more than a magnitude. For each protein coding gene an alignment of the determined amino acid sequences has been created (see Section 2.2). The concatenated alignments for the different protein coding genes for a group of species have then been used for phylogenetic reconstruction (see Section 2.3). In addition to the complete dataset (denoted as METAZOA) subsets, partly complemented with additional data, were used in analyses presented in other articles of this special issue: ARTHROPODA - without neopteran insects - (Podsiadlowski et al., 2013), DEUTEROSTOMIA (Perseke et al., 2013); DIPLOBLASTS (Osigus et al., 2013); HEXAPODA (Simon and Hadrys, 2013), and MOLLUSCA (Stoeger and Schroedl, 2013).

2.2. Creation and processing of alignments

Amino acid sequences were aligned separately for each protein coding gene with MAFFT version 6.716 (Katoh et al., 2002) using the default parameter values. The frayed ends of the aligned sequences were trimmed by employing a simple rule: Starting separately from both ends of an alignment, columns are removed until a column with less than 20% gaps is found or the total number of removed columns reaches 100. Homoplastic or random-like characters are removed by masking the trimmed alignments with the software noisy, rel. 1.5.9 (Dress et al., 2008), using a cutoff value of 0.8. The single protein alignments were concatenated in lexicographic order with respect to their names. In the few instances were an organism lacks a protein-coding gene the concatenated alignment is filled with gaps at the corresponding positions.

The phylogenetic analysis (Section 2.3) of the complete METAZOA data set is computationally extremely demanding. Therefore only a subset of species has been considered. The selection of such a subset has to regard the biases due to an over-representation of certain taxonomic groups. The reduction of the data set is carried out in such a way that the phylogenetic and sequence diversity within the data set is maintained. This is done with an automated approach as described in the following. A neighbor-joining tree of the concatenated alignments for the protein coding sequences has been calculated with QuickTree (Howe et al., 2002). Groups of very closely related sequences are identified as connected smallest subtrees with the property that the longest patristic distance between two leafs in the subtree is smaller than a cutoff value given as parameter. From such a group of sequences only two species having the sequence with the shortest and longest distances to the root node of the respective subtree are included in the data set. In order to prevent the exclusion of sequences belonging to species of high phylogenetic interest, all species from an expert curated list of 156 species (see supplement) are guaranteed to be included in METAZOA. The cutoff value is chosen to produce a data set of appropriate size (i.e. 684 species) such that a phylogenetic analysis is feasible in reasonable time.

In order to assess taxon sampling issues two smaller data sets have been analyzed. A data set containing 325 species (denoted as METAZOA-300) has been created by using a more restrictive threshold. Furthermore, a manually curated data set containing 114 species (denoted as METAZOA-100) has been analyzed (a detailed list of all taxa is provided in the supplement).

2.3. Phylogenetic reconstruction process

The Maximum Likelihood analysis was performed with RAxML version 7.2.8 (Stamatakis, 2006) by employing a protein mixed model, i.e. CAT+MTZOA+F (CAT+MTART+F for ARTHROPODA, and HEXAPODA, respectively) with GAMMA correction of the final tree. At least three batches of 100 rapid bootstrap trees were generated until all four convergence criteria provided by RAxML were met (Stamatakis et al., 2008; Pattengale et al., 2009). Additional batches of 100 rapid bootstraps were necessary for the data sets DIPLOBLASTS (400 in total), and HEXAPODA (400 in total). A best tree search for the best scoring ML tree was conducted. Except for the two large species sets METAZOA and DEUTEROSTOMIA 200 distinct starting trees were used. The run time requirements for the two larger datasets necessitated to select the best tree from separate runs with fewer starting trees, i.e. 10 times 10 and 50 times one starting tree for METAZOA and DEUTEROSTOMIA, respectively.

Bayesian Analysis was performed with PhyloBayes-MPI version 1.3b (Lartillot et al., 2009; Lartillot et al., 2013) on the smaller dataset (METAZOA-100) using the model CAT, MTZOA+Gamma. Six chains were run in parallel for at least 5500 iterations. The first 3000 samples were discarded as burn-in. From the remaining samples every tenth tree was used to compute a majority rule consensus tree and node support in form of Bayesian posterior probabilities. A PhyloBayes analysis was also started with the complete dataset (684 species), but the chains did not come to reasonable convergence and resolution of the consensus tree after comparatively long running time (four weeks).

2.4 Modeling amino acid substitution models

Two independend MAFFT (version 6.716, Katoh et al., 2002) alignments of amino acid sequences were obtained for light- and heavy strand encoded *nad5* genes from the METAZOA dataset. Best trees were calculated with RAxML version 7.2.8 (Stamatakis, 2006), model settings MTZOA+CAT+F for rapid bootstrapping and with MTZOA+GAMMA+F for the final tree. The resulting best trees were used for optimization of the model parameters under GTR+F model for amino acids. Substitution rates were obtained from the model parameters, amino acid frequencies were calculated directly from the alignments.

2.5 Nucleotide and amino acid statistics

AT and GC skew were determined for complete genomes (plus strand) according to the formula defined by Perna and Kocher (1995), AT skew = (A-T)/(A+T) and GC skew = (G-C)/(G+C), where the letters stand for the absolute number of the corresponding nucleotides in the sequences. We also analyzed the effect of AT content, GT and AC rich strands (measured by AT and GC skew) on the amino acid composition of mitochondrial protein coding genes. Considering the first two codon positions, which are crucial for coding, amino acids were grouped as follows: F, I, K, M, N, Y (encoded by AT-rich codons AAN, ATN, TAN and TTN) versus A, G, P, R (encoded by GC-rich codons GCN, CGN, CCN, GGN) and H, K, N, P, Q, T (encoded by CA-rich codons) versus C, F, G, V, W (encoded by GT-rich codons). For a species in an alignment the fraction of a set of amino acids denotes the fraction of these amino acids with respect to the total number of amino acids of the corresponding sequence (disregarding gaps). Leucine and serine are ignored since these amino acids are encoded by more than four codons (what means that the first two codon positions must not be the same). If not stated otherwise, statistics of the complete genome are determined for the plus strand, i.e. the strand given in RefSeq; statistics for single genes always refer to the coding strand.

2.6 Gene order divergence

Gene orders were compared using the breakpoint distance (Blanchette et al., 1999). An adjacency of a gene order G is a pair of genes that are adjacent in G. A conserved adjacency of two gene orders G and F is an adjacency in both gene orders where the corresponding genes are either in the same or opposite order and orientation. A breakpoint in a gene order with respect to another gene order is a pair of adjacent genes that is not conserved, i.e. not adjacent in the other genome. The breakpoint distance is the average number of breakpoints for two gene orders with respect to each other.

We tested the correlation between gene order rearrangements and branch length of the corresponding taxon. For gene order we excluded the highly variable positions of tRNAs, thus

in most cases 15 genes were considered. The branch length of each taxon from the base of Bilateria was determined as well as the minimal number of breakpoints needed to get from the taxons` gene order to one of three proposed ground patterns (corresponding to ground pattern hypotheses of Deuterostomia, Lophotrochozoa, and Ecdysozoa). As it is currently not possible to define a single most reasonable hypothesis for ground patterns of gene order for Metazoa, we used three different gene orders, defined as putative ground patterns for Ecdysozoa, Deuterostomia, and Lophotrochozoa, for an assessment of the derived nature of a given gene order. The deuterostome pattern is still realized in most of the deuterostome mitogenomes. The ecdysozoan pattern is the same as seen in most arthropods, an onychophoran species, and a tardigrade. The priapulid pattern is different from the ecdysozoan ground pattern by an inversion of half of the genome. Nematodes have a large variety of gene order patterns, not much resembling any of the presented three ground patterns. The lophotrochozoan pattern is one which is still realized in a brachiopod, some nemertean species, and in some molluscs. It is the only pattern realized in more than one phylum of Lophotrochozoa, and it is the lophotrochozoan pattern most similar to the ecdysozoan and deuterostome patterns.

2.7 Statistical analyses

Statistical analyses have been conducted with the R package (R Development Core Team, 2011). Pearson correlation coefficients have been computed with the function lm. The Wilcoxon signed-rank test (R function wilcox.test) was used to test statistical significance, using a p-value threshold of 0.01.

3. Results & Discussion

3.1 Phylogenetic trees obtained with mitochondrial genome data

Our most comprehensive dataset (METAZOA) includes almost all mitochondrial genome entries from invertebrate metazoans and a selection of vertebrate entries. A maximum likelihood analysis of this dataset using RAxML reveals an unbalanced tree with large differences in branch lengths and a lack of supported resolution for most basal nodes, clearly indicating some major problems in deep phylogeny reconstruction of Metazoa with mitochondrial genomes using up-to-date methods (Figure 1). This figure also displays that nucleotide frequencies and strand skews strongly vary among metazoan mt genomes.

At the base of this tree Cnidaria and Porifera appear polyphyletic, with Hydrozoa forming the sister group to Bilateria + Hexactinellida. The limited taxon sampling of mitochondrial genomes for several of these groups (e.g. Hydrozoa, Scyphozoa, Hexactinellida) clearly biases the analysis in this part of the tree. Only a few basal branches are well supported by bootstrap percentages (e.g. Bilateria, Bilateria+Hexactinellida). Mitochondrial genomics and the relationships of the basal metazoan splits are in focus of another article in this special issue (Osigus et al., 2013) and thus will not be discussed in detail here.

One remarkable feature of the tree presented in Figure 1 is the increase of branch lengths among bilaterian taxa in comparison to non-bilaterian taxa and outgroup members. Some unusual sistergroup relations found in the Bilateria part of the tree may be due to long-branch artifacts - most strikingly the assemblage of Nematoda, Platyhelminthes, Syndermata, and some long branching arthropod taxa like Acari and Phthiraptera. This group is nested within a likewise artificially assembled arthropod clade. Here only a small amount of “high-ranking” sister group

relations show support values above 80% (Onychophora + Priapulida, Branchiura + Pentastomida), whereas several well-established monophyla fail to be supported by bootstrapping and even by the best tree topology, e.g. Hexapoda, Chelicerata, and Malacostraca.

The lophotrochozoan part of the tree shows bootstrap support for some of the traditional phyla, e.g. Brachiopoda, Nemertea, Annelida sensu lato (with Sipuncula and Echiura), Entoprocta, and Bryozoa. Mollusca are not supported as a monophylum, but instead are scattered between the other lophotrochozoan taxa. As well interrelationships between the lophotrochozoan phyla are not resolved by this dataset and are essentially disturbed by the scattered distribution of molluscan subtaxa between the other lophotrochozoan taxa.

The only part of the tree which is largely congruent with phylogenetic analyses obtained with nuclear genome datasets is the Deuterostomia clade, except for the position of tunicates. The basal splits of deuterostomes are reasonable and well supported by bootstrap values. Tunicates have much longer branches and do not end up with the other deuterostomes, but instead are found as sister to the Acoela.

For an evaluation of the effects of large versus small taxon sets we conducted further analyses with smaller taxon samplings (METAZOA-300 and METAZOA-100 containing approximately 300 and 100 taxa, respectively). Results from maximum likelihood analysis for the METAZOA-300 dataset are largely similar to results of the 684 taxon dataset (see supplementary material). In the METAZOA-100 taxa dataset we omitted the long-branching Nematoda and Platyhelminthes, as well as most of the molluscan taxa, to see if these had a shifting effect on the other long branches (Figure 2). Even in this strongly reduced taxon set the topology and bootstrap support of the RAxML analysis not much differed in quality from the trees obtained from the two larger taxon sets. Again the arthropod assemblage seems arbitrarily arranged and includes the long branching Syndermata. In the lophotrochozoan part of the tree there is some resolution with moderate bootstrap support, probably due to the absence of many molluscan taxa. Brachiopods are sister to Annelida *sensu lato* and the remaining molluscs are combined in a clade with Nemertea and Phoronida. Thus a smaller taxon set results only in a slight improvement of phylogenetic support, especially when extreme long branching taxa are omitted.

Bayesian analysis of the METAZOA-100 dataset with (PhyloBayes-MPI) resulted in a different picture. Compared to the RAxML analysis long-branch phenomena did affect the outcome to a lesser extent, e.g in contrast to the RAxML tree Syndermata is found within a lophotrochozoan clade and some long-branching arthropods like Protura, Copepoda and Branchiura are now found among Pancrustacea. Ecdysozoa and Lophotrochozoa found maximum support by Bayesian posterior probabilities and the Mollusca are found to be monophyletic. Nevertheless, some other well-defined taxa are still not supported, e.g. Chordata. Furthermore branching patterns among arthropods and lophotrochozoans are not resolved. Thus, the PhyloBayes approach seems to be promising in phylogenetic analysis of mitochondrial genome data with strong differences in branch lengths, but is far more expensive in computational time and not yet feasible for our biggest dataset.

The unsatisfying outcome of a phylogenetic analysis using mt genome data shows that a large taxon sampling cannot solve the problems that have been shown in many former analyses using more limited taxon samplings (Steinauer et al., 2005; Jang and Hwang, 2009; Mwinyi et al., 2010). Other studies omitted long-branching taxa from the analysis to circumvent these

problems (Helfenbein et al., 2004; Yokobori et al., 2008). However, often the omitted taxa are of special interest concerning their phylogenetic position.

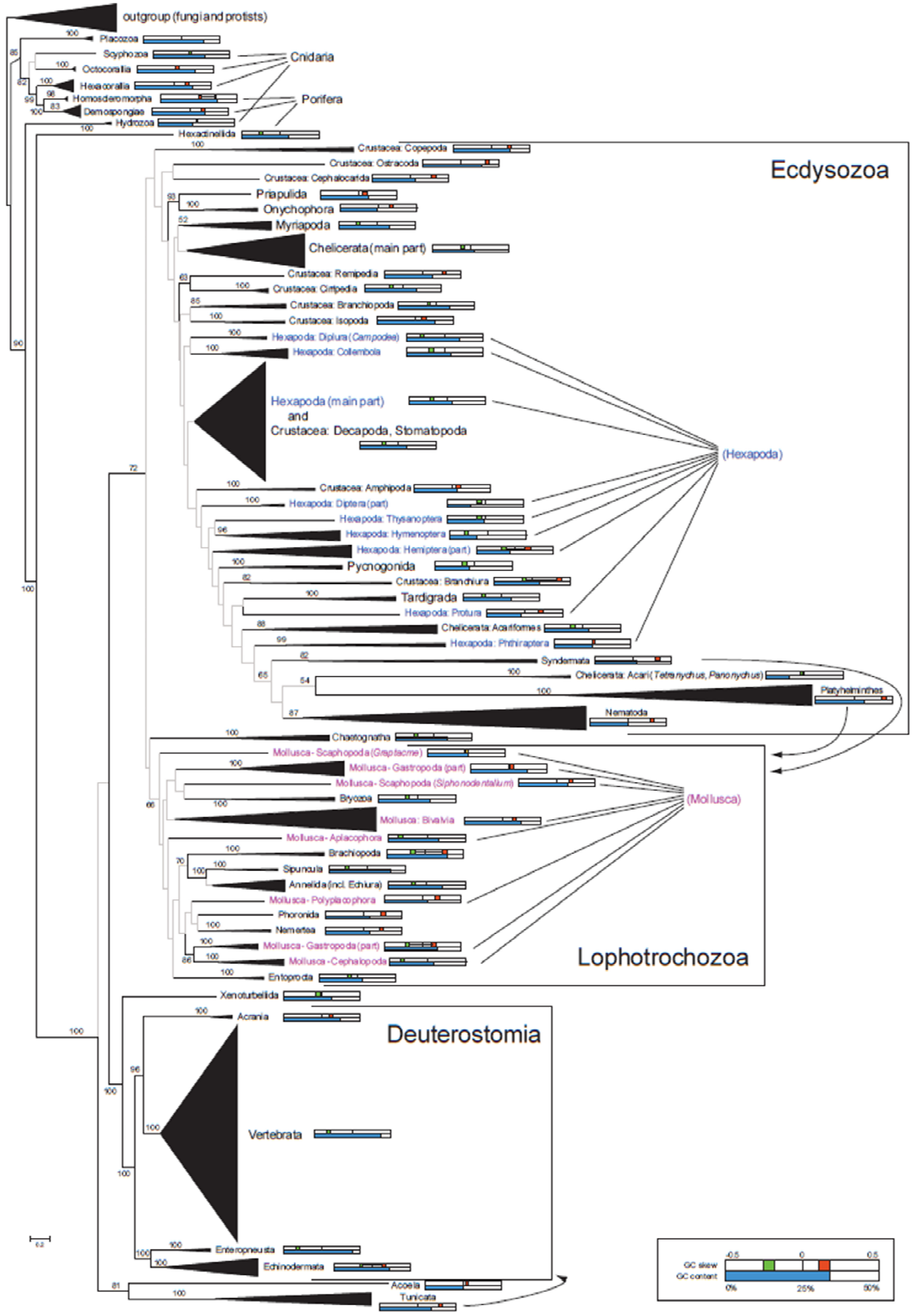
3.2 Nucleotide and amino acid frequencies of animal mitochondrial genomes

It is known that animal mitochondrial genomes vary significantly in nucleotide composition and almost all show a bias between the two strands of the genome (Perna and Kocher, 1995; Hassanin et al., 2005). This begs the question if the shifts in nucleotide composition affect amino acid alignments and subsequent phylogenetic analyses? The abundance of nucleotides demonstrates the AT-richness in animal mt genomes: A: 15.6%-48.7%, C: 4.4%-34.7%, G: 4.8%-31.3%, T: 21.0-54.9% (values from plus strand, due to NCBI RefSeq annotation). Almost balanced nucleotide frequencies (all four nucleotides around 25%) are found only in a few species, e.g. the placozoan *Trichoplax* species, the snail *Myosotella myositis,* and the anthozoan cnidarian *Savalia savaglia*. The lowest AT content is seen in *Balanoglossus* species (51.4% and 52.8%), as well as again *Savalia savaglia* (51.7%), and *Trichoplax adhaerens* (53%). Highest AT contents are found in insects, with an extreme value of 87.4% in the parasitic wasp *Diadegma semiclausum*. Several other species from Hymenoptera, Diptera, and Lepidoptera reach values higher than 80%, as well do some mites and nematodes.

Besides AT content variation, the strand bias is another factor yielding unbalanced nucleotide frequencies. Probably due to an asymmetry in the replication process of mitochondrial genomes, GC and AT skews characterize differences between the two strands of a mitochondria genome, with one strand favoring G/T over C/A (Perna and Kocher, 1995; Hassanin et al., 2005). Since G is by far the heaviest of the four nucleotides, the GT-rich strand corresponds to the "heavy strand". It is important to note that this is completely different from the major/minor coding strand or plus/minus strand terminology. When most genes are coded on the same strand it is easy to define this one as the major coding strand, but not in the case where both strands show a similar amount of coding genes. The plus strand is mostly defined according to the orientation of the *cox1* gene, an arbitrary convention given that gene order (and relative orientation) is variable and replication and transcription origins are difficult to detect automatically from sequence information.

Additionally, the asymmetric replication process creates nucleotide skews differing along the mitogenome (Reyes et al., 1998), depending on the position and orientation of the replication origins. Comparing GC and AT skews in arthropods gave evidence for a number of independent reversals of nucleotide skews, some of them with little or no changes in gene order, e.g. in most spiders, the varroa mite, scorpions (Hassanin, 2006), some pycnogonids (Arabi et al., 2010), and isopods (Kilpert et al., 2012). Inversion of the replication origin was discussed as a putative mechanism for a reversal of the strand bias (Hassanin et al., 2005; Wei et al., 2010).

Fig. 1: Phylogenetic tree obtained from Maximum Likelihood analysis with amino acid alignments from mitochondrial protein coding genes. Best tree from RAxML analysis with bootstrap support from 100 pseudoreplicates. Branches with bootstrap support below 85% are shown in gray. Some major derivations of otherwise well supported phylogenetic hypothesis are highlighted (arrows show expected placement of Platyhelminthes, Syndermata, and Tunicata; polyphyletic hexapods are blue, molluscs pink). Next to the taxon names information about GC content (blue) and GC skew (green/red) of mt genomes is given (according to plus-strand equence). Mean value of GC skew is shown in green if negative, and red if positive (if there is large variation, a span is given, e.g. Mollusca – Gastropoda). GC content is depicted as a left bound column with 50% at the right margin.


outgroup (fungi and protists)
Placozoa
Scyphozoa
Octocorallia
Hexacorallia
Homoscleromorpha
Demospongiae
Hydrozoa
Cnidaria
Porifera
Hexactinellida
Ecdysozoa
Crustacea: Copepoda
Crustacea: Ostracoda
Crustacea: Cephalocarida
Priapulida
Onychophora
Myriapoda
Chelicerata (main part)
Crustacea: Remipedia
Crustacea: Cirripedia
Crustacea: Branchiopoda
Crustacea: Isopoda
Hexapoda: Diplura (Campodea)
Hexapoda: Collembola
Hexapoda (main part)
and
Crustacea: Decapoda, Stomatopoda
(Hexapoda)
Crustacea: Amphipoda
Hexapoda: Diptera (part)
Hexapoda: Thysanoptera
Hexapoda: Hymenoptera
Hexapoda: Hemiptera (part)
Pycnogonida
Crustacea: Branchiura
Tardigrada
Hexapoda: Protura
Chelicerata: Acariformes
Hexapoda: Phthiraptera
Syndermata
Chelicerata: Acari (Tetranychus, Panonychus)
Platyhelminthes
Nematoda
Chaetognatha
Mollusca - Scaphopoda (Graptacme)
Mollusca - Gastropoda (part)
Mollusca - Scaphopoda (Siphonodentalium)
Bryozoa
Mollusca: Bivalvia
(Mollusca)
Mollusca - Aplacophora
Brachiopoda
Sipuncula
Annelida (incl. Echiura)
Mollusca - Polyplacophora
Phoronida
Nemertea
Mollusca - Gastropoda (part)
Mollusca - Cephalopoda
Entoprocta
Lophotrochozoa
Xenoturbellida
Acrania
Deuterostomia
Vertebrata
Enteropneusta
Echinodermata
Acoela
Tunicata
GC skew
GC content
-0.5
0
0.5
0%
25%
50%

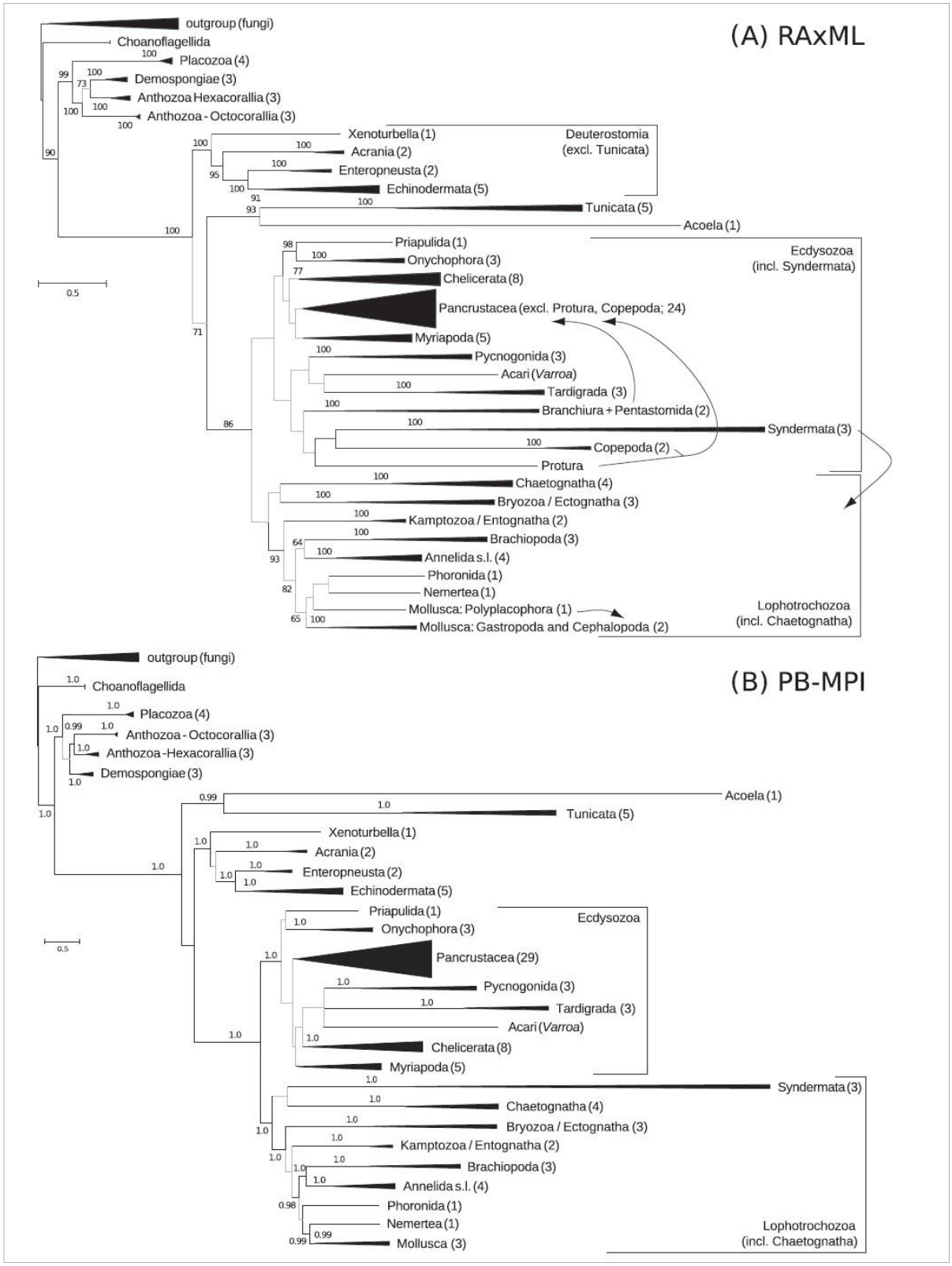


Fig. 2: Phylogenetic tree obtained with an alignment of amino acid sequences from the dataset reduced to 100 taxa (methods for taxon selection see text). (A) Best tree from RAxML analysis with bootstrap percentages (>50%) beneath the branches. Differences to the tree shown in subfigure B are highlighted by arrows. (B) Consensus tree from six independent chains of PhyloBayes-MPI. Bayesian posterior probabilities are given when >0.95. In both trees the numbers in brackets after taxon names refer to the number of species representing this taxon in this reduced data set. Insufficiently supported parts of the tree are light Grey.

Phylogenetic analyses in the above mentioned studies yielded longer branches (=more substitutions) for clades with a reversal of strand bias than for other clades.

There is a strong negative correlation between AT and GC skew, when all mt genomes (plus-strand) from our most comprehensive alignment (METAZOA) are compared (Figure 3). There are two clusters: one with positive GC skew and clearly negative AT skew, the other with predominantly negative GC skew and positive or moderately negative AT skew. Note that using the minus strand for one of the clusters would superimpose the clusters. It is unclear if the inversion of the skews is due to an inversion of the replication origin, which is not easily determinable. The long-branching taxa (red in Figure 3) have significantly larger GC and smaller AT skews. For instance, Nematoda, Platyhelminthes, and Tunicata have combinations of highly positive GC skew and a highly negative AT skew. The following alternative hypotheses are significantly supported (Wilcoxon test, $p < 10^{-16}$): a) AT skew for problematic taxa is less than the one for non-problematic taxa b) GC skew is larger than the one for non-problematic taxa.

Phylogenetic analyses on a high taxonomic level, like the study presented here, predominantly use amino acid sequences to overcome problems with aberrant nucleotide frequencies and nucleotide skews, which were assumed to have the strongest effect on synonymous substitutions. However, the variation in AT content and GC and AT skews obviously must lead to changes on the amino acid level, too (e.g. Foster et al., 1997; Min and Hickey, 2007). Figure 4 shows an example of amino acid frequency correlations for a single gene across Metazoa (*nad5* is shown here, all other genes are presented in the supplement). Because the distribution of genes on the two strands differs among metazoan mitochondrial genomes, correlations of strand bias and amino acid composition can only be analyzed separately per gene. We chose *nad5* for two reasons: (1) it is the largest and among the least conserved protein coding genes in metazoan mitochondrial genomes, thereby providing most information and (2) in metazoan species *nad5* is well distributed on the plus and minus strand (approximately 2:1). Analysis of *nad5* shows a clear negative correlation of the fraction of amino acids coded by AT rich codons and the fraction of amino acids coded by GC rich codons (Figure 4A). The slope of the linear regression for the fractions of AT-rich and GC-rich codons is approximately -0.5, i.e. as for AT and GC content of the genome. Problematic taxa with long branches in Figure 1, as depicted by red dots, are slightly shifted from the main regression line to lower proportions for both AT and GC rich codons. The effects are less prominent in more conserved genes like *cox1-3* and *cytb*, but nevertheless visible, as well as in the complete alignment (see supplementary material). Thus, a strong dependence of AT / GC content and amino acid composition can be attested. This suggests homoplasious effects, at least when extreme AT / GC contents are reached, e.g. in the case of some hexapods.

The effects of strand bias (heavy strand is GT rich; light strand is CA rich) are shown in Figure 4B. The usage of amino acids encoded by GT rich and CA rich codons has a clear negative correlation. Here the formation of two clusters is noticeable, corresponding almost perfectly to heavy and light strand encoded *nad5* genes. Problematic long branched taxa (according to Figure 1) tend to accumulate high fractions of GT-rich codons and low fractions of CA codons, corresponding to genes located on the GT rich, i.e. heavy strand (red dots in Figure 4b).

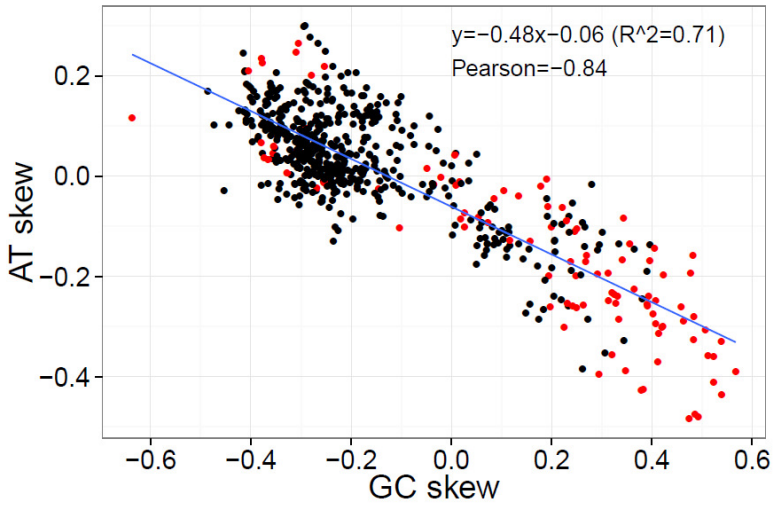


Fig. 3: GC-skew versus AT-skew in complete mitochondrial genomes (plus-strand) from Metazoa. Red: 153 species with long branches and unusual phylogenetic position in the tree shown in Figure 1 (Nematodes, Platyhelminthes, Syndermata, Acari, Tunicata, some hexapods, and molluscs), black: remaining taxa with reasonable phylogenetic position.

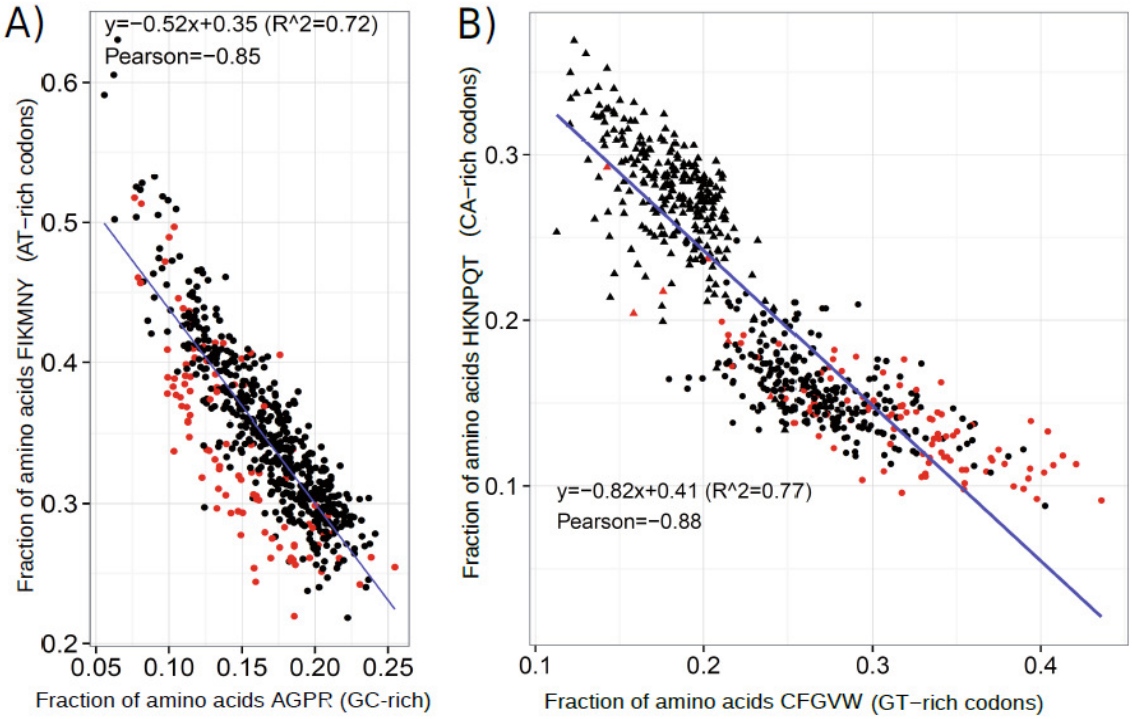


Fig. 4: Amino acid usage in the mitochondrial *nad5* gene. (A) The abundance of amino acids with AT-rich codons plotted against abundance of GC-rich codons. (B) The abundance of amino acids with CA-rich codons plotted against GT-rich codons. Data points corresponding to *nad5* genes located on the CA-rich strand are shown as triangles, those of nad5 genes on GT-rich strand as squares. Red data points correspond to long branched taxa as in Figure 1 (Nematodes, Platyhelminthes, Syndermata, Acari, Tunicata, some hexapods, and molluscs).

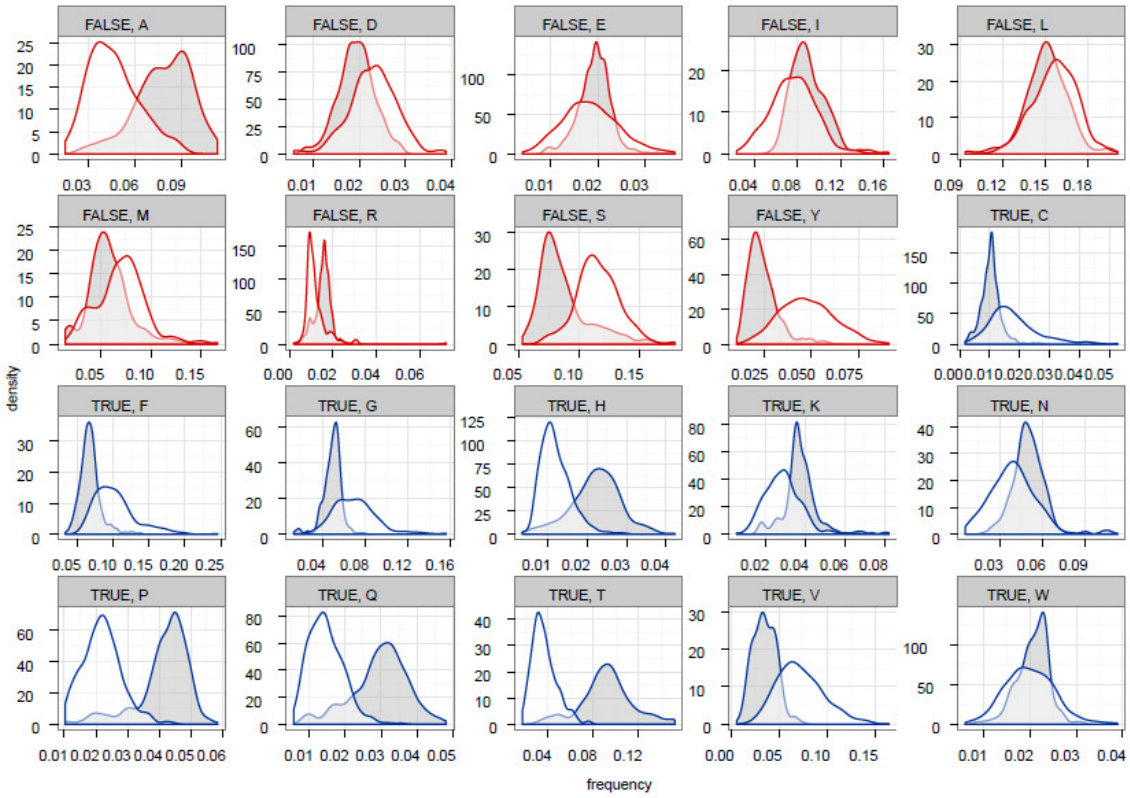


Fig. 5: Density of amino acid frequencies in the mitochondrial *nad5* gene. Density of each amino acid is plotted against its frequency in the *nad5* gene. For each amino acid two density plots were computed independently for *nad5* genes from CA rich (Grey area) and GT rich strands (white area). Blue curves are from amino acids which should be affected by strand bias (GT and CA rich codons), also indicated by the term "true".

The correlation of nucleotide composition and amino acid usage is as well reflected by the strong correlation of GC (resp. AT, GT, CA) content and the fraction of GC (resp. AT, GT, CA) rich codons (see figures for each protein coding gene in supplementary material). Thus, the amino acid composition of a gene strongly depends on whether it is located on the heavy or the light strand. This effect is visible also in the bimodal frequency distribution of several amino acids, e.g. those encoded by CA rich codons (Thr ACN; Gln CAA/C; His CAT/C), corresponding to the strand bias (Figure 5).

Optimized substitution model parameters for the two subsets of heavy strand and light strand encoded *nad5* genes have been determinate (see Section 2.4). In accordance with the previous results, the two optimized models differ strongly (Figure 6). Thus, the usage of a unified substitution model (as generally applied in most analyses) barely fits to a dataset where model parameters strongly depend on the orientation of the gene, or in other words, the model switches for a gene when a gene inversion occurs, while the model stays the same for other genes not involved in this event. Hence, instead of a "one fits all" model "heavy" and "light" strand models should be used in turn depending on which strand the gene is encoded in the corresponding part of the tree.

Altogether our results suggest a strong relation of the strand bias and amino acid sequences and thus the danger of homoplasious substitutions in taxa that achieved a similar genome organization independently (at least when inversions are involved). In addition an accelerated

substitution rate may occur each time a gene switches strands, hinting to a correlation of a high frequency of gene order changes with long branches in a phylogenetic tree (see next section).

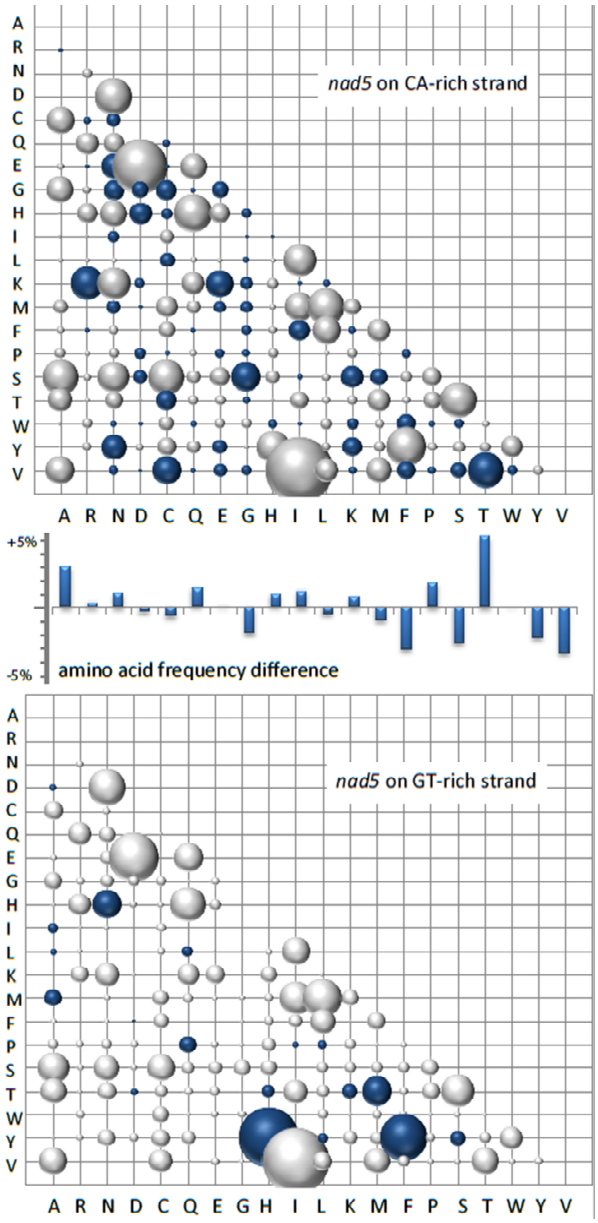


Fig. 6: Amino acid substitution models of *nad5* encoded on CA rich and GT rich strand. Between the two models the differences in amino acid frequencies of the two sets are shown (percent point difference of absolute proportion). Blue spheres indicate more than doubled substitution rate compared to the other model.

3.3 Correlations of gene rearrangements and substitution rates

The structural genome variation of mitochondrial genomes is another source of phylogenetic information. Boore et al. (1995) were the first to demonstrate phylogenetic signal in mitochondrial gene order diversity. Mitochondrial gene order stayed relatively stable in vertebrates and insects, while highly variable patterns are found in e.g. Mollusca (Boore et al., 2004), Bryozoa (Waeschenbach et al., 2006; Jang and Hwang, 2009; Nesnidal et al., 2011), Tunicata (Gissi et al., 2010; Stach et al., 2010), and Acari (Shao et al., 2006). It was mentioned

several times that a higher variation in gene order may correspond with higher substitution rates and therefore promotes long branches and problems in sequence-based analysis. Studies with arthropod examples show strong correlations between gene order and sequence distances (Shao et al., 2003; Xu et al., 2006). In the case of gene rearrangements involving strand switch of genes this could be explained with the strand bias of nucleotide frequencies (Hassanin et al., 2005), which also affects amino-acid frequencies in protein coding genes (Podsiadlowski and Braband, 2006; Min and Hickey, 2007).

In the absence of a coherent model for a ground pattern of mt gene order for all Metazoa or even Bilateria, we used three putative basal gene order patterns of protein coding and ribosomal RNA genes for Deuterostomia (Bourlat et al., 2009), Ecdysozoa (Webster et al., 2006), and Lophotrochozoa (Podsiadlowski et al., 2009) (Figure 7). Using data from our most comprehensive analysis, we determined for each taxon the branch length from the root and the breakpoint distance (Blanchette et al., 1999) of its gene order compared to the three basal patterns (Figure 8). We found a correlation of gene order change (quantified as the minimal number of breakpoints between the gene order under view and one of the three basal patterns) and amino acid substitution rate (here determined as root to leaf distance, i.e. the sum of the branch lengths in the phylogenetic tree from the root to the leaf). For up to seven breakpoints highly variable branch lengths were detected, but with more than seven breakpoints the number of taxa with short branches is in a minority. This is supported by the fact that the largest breakpoint distance where the null hypothesis (that the branch lengths are less or equal than those for equal gene orders) cannot be rejected is six. Thus, seven breakpoints lead to a significant increase in branch lengths. Complete shuffling of the mitochondrial genome is clearly correlated with long branches (=high substitution rate), while a moderate gene rearrangement (2-6 breakpoints) has virtually no effect. On the other hand extremely long branches (>5) are only found in genomes which are highly rearranged (eight or more breakpoints). Extreme values for both, branch lengths and breakpoint distance are found in Nematoda, Platyhelminthes, Tunicata, some Mollusca, and some Arthropoda (Acari, Copepoda). Nevertheless it should be mentioned that even taxa with the same gene order may have substantial variation in branch lengths, reaching mean substitution rates similar to those of taxa with highly rearranged genomes. Consequently, the gene order is only one of several factors related to substitution rate differences. An alternative explanation for this correlation would be that in some taxa unknown underlying features similarly affect both, substitution rates and rearrangement rates. These putative mechanisms may be relaxed repair mechanisms, high mutational stress in combination with lower importance of mitochondrial efficiency.

3.4 A more detailed discussion of problematic taxa

Several taxa were found in unexpected position within our tree (Figure 1), pointing to problems in constructing a reliable phylogenetic tree from a mitochondrial amino acid alignment. For an in depth discussion of selected phyla see the accompanying articles in this special issue. Here we will shortly re-examine some of the unexpected results from our phylogenetic analyses in the light of our results from nucleotide skew, amino acid frequencies, and gene order changes presented in Section 3.2 and 3.3.

Nematoda, Platyhelminthes, and Syndermata are most strongly affected by long-branch attraction, as seen in our phylogenetic tree (Figure 1). A clade composed of these groups was

never supported by datasets obtained from nuclear genome sequences, where Platyhelminthes and Syndermata are part of Lophotrochozoa and Nematoda are part of Ecdysozoa (Dunn et al., 2008; Hejnol et al., 2009; Pick et al., 2010). Amino acid substitution rate (estimated from branch lengths in the trees) in syndermatans and nematodes is more than doubled, while in platyhelminthes it appears to be more than four times higher than the average substitution rate in the remaining bilaterian taxa. It is apparent, that many of the long branches are comprised of species with endoparasitic lifestyle, but not all of them - rotifers and many non-parasitic nematodes are as well represented. In Syndermata the parasitic Acanthocephala have longer branches than free-living rotifers, but in nematodes no clear correlation between branch length and parasitic lifestyle is present. Anoxic conditions, a higher metabolic rate, a short generation time, and bottleneck effects, associated with low effective population size, were discussed to affect substitution rates of mitochondrial genomes (Martin, 1995; Min and Hickey, 2008). All of these effects are not restricted to an endoparasitic lifestyle, e.g. nematodes living in rotten plants, carcasses, or dung experience similar harsh conditions. Notable is that in phylogenomic datasets using nuclear genes Nematoda, Syndermata, and Platyhelminthes are among the longest branches as well (Dunn et al., 2008; Hejnol et al., 2009), suggesting an generally accelerated substitution rate in both, mitochondrial and nuclear genomes.

Tunicata show by far the longest branches among deuterostomes and never end up with Vertebrata and *Branchiostoma*, as clearly supported by nuclear genes and morphological data (Delsuc et al., 2006). Their gene order is completely different from all other Deuterostomia and highly variable, i.e. even congeneric species differ in gene order (Iannelli et al., 2007; Gissi et al., 2010; Stach et al., 2010). Tunicates are usually not confronted with anoxic conditions, thus their high substitution rate and gene order variation may have other (unknown) reasons. Phylogenomic analysis of nuclear genes show extremely derived sequences for *Oikopleura dioca*, but average branch lengths for *Ciona* species (Denoeud et al., 2010).

Mollusca remains one of the most problematic taxa in mitochondrial genome based analyses. Interestingly, the taxa showing the least derived gene orders (the polyplacophoran *Katharina tunicate*, some gastropods, e.g. *Haliotis, Ilyanassa,* and the cephalopodes) have also the shortest branches in the tree. This indicates that the most probable reason for problems in phylogenetic analysis of Mollusca relates to their frequent gene order shuffling, promoting differences in strand bias - remark the extremely different GC skews between mollusk taxa in Figure 1 - which in turn affects the amino acid usage. As for tunicates, a reason for the comparatively unstable gene order in mollusks is unknown. More details for molluscan mitochondrial genomes are found in an accompanying article of this special issue (Stoeger and Schroedl, 2013).

The placement of several hexapod taxa in our tree seems to be influenced by accelerated substitution rates, most prominently in Thysanoptera, Phthiraptera, some Hymenoptera, Diptera, and Hemiptera. Besides long-branch attraction there must be some other reason for the lack of support for Hexapoda. For instance the lack of inclusion of Collembola and Diplura into hexapods is a long known problem in phylogenetic analyses with mitochondrial datasets (Nardi et al., 2003; Pisani et al., 2004). In the case of Thysanoptera, Hemiptera, and Phthiraptera mitochondrial gene orders notably differ from the hexapod ground pattern (for more details see Simon and Hadrys, 2013). This is not the case for the dipteran and hymenopteran species placed outside of the main hexapod clade in Figure 1. Here the extreme values of AT content seem to contribute to the improper result of our phylogenetic analysis.

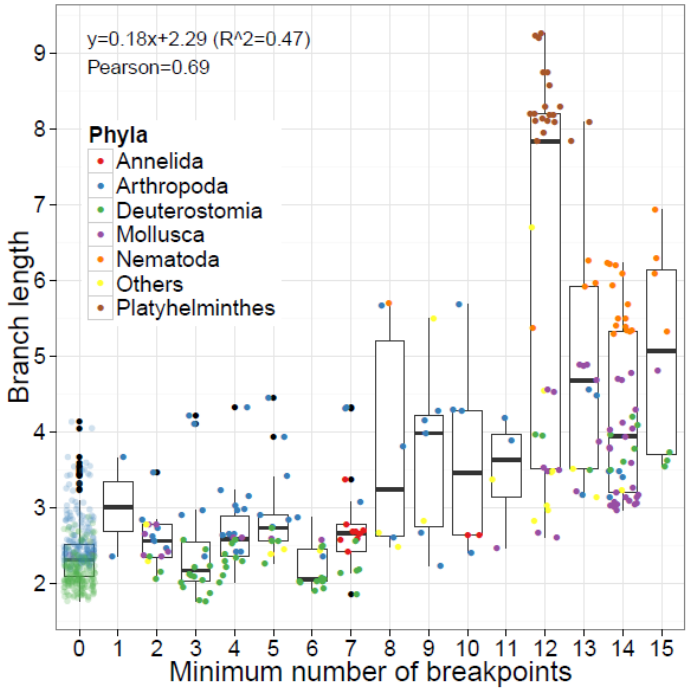


Fig. 8: Correlation of gene order and branch length in phylogenetic analyses. Gene order changes are recorded as minimum breakpoint distance to one of three alternative ground pattern of Bilateria. Only protein coding and ribosomal RNA genes are used in this comparison. Breakpoint number is integer, data points are slightly scattered around values on the x-axis for better display of their quantity.

Conclusions

Nucleotide frequencies vary broadly among Metazoa. While slight differences may be overcome by the usage of amino acid alignments, stronger deviations are also reflected in shifts in amino acid frequencies. Amino acids can be grouped according to shared physical/chemical properties and are often interchangeable without changing the functional efficiency of the corresponding protein. Thus considerable differences in AT content, changes in strand bias or replication origins, and inversion of genes strongly affect amino acid substitution rates and the outcome of phylogenetic analysis based on amino acid alignments. The correlation between gene order distances and substitution rate fits well into this picture. However, it explains the exceptional high substitution rates in, e.g. platyhelminthes, only to a certain degree. In our phylogenetic analysis of mitochondrial genome data from a broad taxon sampling of Bilateria sufficient resolution is lacking at the base of the tree as well as for ecdysozoan and lophotrochozoan interrelations. Several presuppositions for good phylogenetic markers seem to be violated (frequency stationarity, even substitution rates, directed substitutions via strand bias or selection) in this dataset. On the other hand some parts of the tree (e.g. deuterostomes) show reasonable branching pattern and good bootstrap support even for deep splits. This suggests that mitochondrial genomes may still have value in phylogenetic analyses, at least when gene order, nucleotide frequency, and strand bias does not vary extremely among the studied taxa. To understand the dramatic differences of nucleotide abundance, strand bias, and mitochondrial substitution rates between metazoan taxa we are in need of a more thorough comparative

analysis of mitochondrial functionality in cellular metabolism and the mitochondrial genetic machinery. Recent medical research revealed an unexpected complexity of the roles that mitochondria play for the maintenance of cellular functions (Zamzami et al., 1996; Szabadkai and Duchen, 2008; Dromparis and Michelakis, 2013), e.g., integrating energy metabolism, signaling pathways, and apoptotic processes - these functional roles may have varying degrees of importance among the different metazoan taxa.

Acknowledgements

We thank Omar Rota-Stabelli and Franz Lang for contributing amino acid sequences of additional taxa. This study was supported by the following grants from German Science foundation (DFG) priority programme "Deep Metazoan Phylogeny", SPP 1174: BL 787/2-1,2 (CB); BA 1520/10-1,2 (LP); SCHO 442/8-1,2 (AB); PO 765/4-3 (LP); STA 850/3-2; STA 850/2 (MB,AD,GF,FJ,MM,PFS); STR-683/5-2, 6-1 (THS); SCHL 229/14-1 (MP,M.Schlegel); HA 1947/5 (HH); SCHR 667/9-1 (IS,M.Schroedl); WA530/34 (BvR); MI 649/7 (BM,KM); Schi-227/20, Schi-227/26-1 (BS). BvR received additional support from DFG grant RE-3454/1-1. FJ was funded by French-German PROCOPE program [DAAD D/0628236, EGIDE PHC 14770PJ]; French-German University (DFH-UFA, Cotutelle de thèse CT-08–10); a doctoral fellowship of the German Academic Exchange Service [DAAD D/10/43622]; and a bridge scholarship of the Collège Doctoral Européen (CDE), Université de Strasbourg.

References

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. USA 97, 4453-4456.

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

Altenhoff, A.M., Dessimoz, C., 2012. Inferring orthology and paralogy. Methods Mol. Biol. 855, 259-279.

Arabi, J., Cruaud, C., Couloux, A., Hassanin, A., 2010. Studying sources of incongruence in arthropod molecular phylogenies: sea spiders (Pycnogonida) as a case study. C. R. Biol. 333, 438-453.

Avise, J.C., 2000. Phylogeography - the history and formation of species. Harvard University Press, Cambridge, Mass.

Ballard, J.W.O., Rand, D.M., 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Ann. Rev. Ecol. Evol. 36, 621-642.

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729-744.

Bensasson, D., Zhang, D., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314-321.

Bernt, M., Braband, A., Schierwater, B., Stadler, P.F., 2013a. Genetic aspects of mitochondrial genome evolution. Mol. Phylogenet. Evol. in press.

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., Putz, J., Middendorf, M., Stadler, P.F., 2013b. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. in press.

Blanchette, M., Kunisawa, T., Sankoff, D., 1999. Gene order breakpoint evidence in animal mitochondrial phylogeny. J. Mol. Evol. 49, 193-203.

Boore, J.L., 2006. Requirements and standards for organelle genome databases. OMICS. 10, 119-126.

Boore, J.L., Medina, M., Rosenberg, L.A., 2004. Complete Sequences of Two Highly Rearranged Molluscan Mitochondrial Genomes, Those of the Scaphopod *Graptacme eborea* and of the Bivalve *Mytilus edulis*. Mol. Biol. Evol. 21, 1492-1503.

Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L., Brown, W.M., 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature 376, 163-165.

Bourlat, S.J., Rota-Stabelli, O., Lanfear, R., Telford, M.J., 2009. The mitochondrial genome structure of *Xenoturbella bocki* (phylum Xenoturbellida) is ancestral within the deuterostomes. BMC Evol. Biol. 9, 107.

Butte, A.J., Dzau, V.J., Glueck, S.B., 2001. Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues". Physiol. Genomics 7, 95-96.

Cruickshank, R.H., 2002. Molecular markers for the phylogenetics of mites and ticks. Syst. Appl. Acarol. 7, 3-14.

Danovaro, R., Dell'anno, A., Pusceddu, A., Gambi, C., Heiner, I., Kristensen, R.M., 2010. The first Metazoa living in permanently anoxic conditions. BMC Biol. 8, 30.

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

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.M., Da, S.C., Brinkmann, H., Mikhaleva, J., Olsen, L.C. and others, 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330, 1381-1385.

Dress, A.W., Flamm, C., Fritzsch, G., Grunewald, S., Kruspe, M., Prohaska, S.J., Stadler, P.F., 2008. Noisy: identification of problematic columns in multiple sequence alignments. Algorithms Mol. Biol. 3, 7.

Dromparis, P., Michelakis, E.D., 2013. Mitochondria in vascular health and disease. Ann. Rev. Physiol. 75, 21.1 21.32.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W. and others, 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745-749.

Foster, P.G., Jermiin, L.S., Hickey, D.A., 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J. Mol. Evol. 44, 282-288.

Galtier, N., Nabholz, B., Glemin, S., Hurst, G.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol. Ecol. 18, 4541-4550.

Gissi, C., Pesole, G., Mastrototaro, F., Iannelli, F., Guida, V., Griggio, F., 2010. Hypervariability of Ascidian Mitochondrial Gene Order: Exposing the Myth of Deuterostome Organelle Genome Stability. Mol. Biol. Evol. 27, 211-215.

Halanych, K.M., 1995. The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. Mol. Phylogenet. Evol. 4, 72-76.

Halanych, K.M., 2004. The new view of animal phylogeny. Ann. Rev. Ecol. Evol. 35, 229-256.

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 -116.

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

Hausdorf, B., Helmkampf, M., Meyer, A., Witek, A., Herlyn, H., Bruchhaus, I., Hankeln, T., Struck, T.H., Lieb, B., 2007. Spiralian phylogenomics supports the resurrection of Bryozoa comprising Ectoprocta and Entoprocta. Mol. Biol. Evol. 24, 2723-2729.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguna, J. and others, 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. Roy. Soc. B 276, 4261-4270.

Helfenbein, K.G., Fourcade, H.M., Vanjani, R.G., Boore, J.L., 2004. The mitochondrial genome of *Paraspadella gotoi i*s highly reduced and reveals that chaetognaths are a sister group to protostomes. Proc. Natl. Acad. Sci. USA 101, 10639-10643.

Howe, K., Bateman, A., Durbin, R., 2002. QuickTree: building huge Neighbour-Joining trees of protein sequences. Bioinformatics 18, 1546-1547.

Hurst, G.D., Jiggins, F.M., 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc. Biol Sci. 272, 1525-1534.

Iannelli, F., Griggio, F., Pesole, G., Gissi, C., 2007. The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. BMC Evol. Biol. 7.

Jang, K.H., Hwang, U.W., 2009. Complete mitochondrial genome of *Bugula neritina* (Bryozoa, Gymnolaemata, Cheilostomata): phylogenetic position of Bryozoa and phylogeny of lophophorates within the Lophotrochozoa. BMC Genomics 10, 167.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066.

Kilpert, F., Held, C., Podsiadlowski, L., 2012. Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. Mol. Phylogenet. Evol. 64, 106-117.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286-2288.

Lartillot, N., Rodrigue, N., Stubbs, D., Richer, J., 2013. Phylobayes-MPI. A Bayesian software for phylogenetic reconstruction using mixed models, MPI version. Online ressource.

Martin, A.P., 1995. Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. Mol. Biol. Evol. 12, 1124-1131.

Min, X.J., Hickey, D.A., 2007. DNA asymmetric strand bias affects the amino acid composition of mitochondrial proteins. DNA Res. 14, 201-206.
Min, X.J., Hickey, D.A., 2008. An evolutionary footprint of age-related natural selection in mitochondrial DNA. J. Mol. Evol. 67, 412-417.
Moritz, C., Dowling, T.E., Brown, W.M., 1987. Evolution of Animal Mitochondrial-DNA - Relevance for Population Biology and Systematics. Ann. Rev. Ecol. Syst. 18, 269-292.
Mwinyi, A., Bailly, X., Bourlat, S.J., Jondelius, U., Littlewood, D.T.J., Podsiadlowski, L., 2010. The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of *Symsagittifera roscoffensis*. BMC Evol. Biol. 10.
Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., Frati, F., 2003. Hexapod origins: Monophyletic or paraphyletic? Science 299, 1887-1889.
Nesnidal, M.P., Helmkampf, M., Bruchhaus, I., Hausdorf, B., 2011. The complete mitochondrial genome of *Flustra foliacea* (Ectoprocta, Cheilostomata) – compositional bias affects phylogenetic analyses of lophotrochozoan relationships. BMC Genomics 12, 572.
Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Muller, W.E. and others, 2013. Deep metazoan phylogeny: When different genes tell different stories. Mol. Phylogenet. Evol. 67, 223-233.
Osigus, H.-J., Eitel, M., Schierwater, B., 2013. Mitogenomics at the base of Metazoa. Mol. Phylogenet. Evol.
Pattengale, N., Alipour,M., Bininda-Emonds,O., Moret,B., Stamatakis,A., 2009. How many bootstrap replicates are necessary? In: Batzoglou,S. (Ed.), Research in Computational Molecular Biology (Lecture notes in Computat. Mol. Biol. 5541). Springer Berlin, pp. 184-200.
Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353-358.
Perseke, M., Golombek, A., Schlegel, M., Struck, T.H., 2013. The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. Mol. Phylogenet. Evol. 66, 898-905.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T., Manuel, M., Worheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol 9, e1000602.
Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E. and others, 2009. Phylogenomics revives traditional views on deep animal relationships. Curr. Biol. 19, 706-712.
Pick, K.S., Philippe, H., Schreiber, F., Erpenbeck, D., Jackson, D.J., Wrede, P., Wiens, M., Alie, A. and others, 2010. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. Mol. Biol. Evol. 27, 1983-1987.
Pisani, D., Poling, L.L., Lyons-Weiler, M., Hedges, S.B., 2004. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. BMC. Biol. 2, 1.
Podsiadlowski, L., Braband, A., 2006. The complete mitochondrial genome of the sea spider *Nymphon gracile* (Arthropoda: Pycnogonida). BMC Genomics 7, 284.
Podsiadlowski, L., Braband, A., Struck, T.H., von Doehren J., Bartolomaeus, T., 2009. Phylogeny and mitochondrial gene order variation in Lophotrochozoa in the light of new mitogenomic data from Nemertea. BMC Genomics 10, 364.
Podsiadlowski, L., Meusemann, K., Reumont, B.v., Fahrein, K., DambachJ., Braband, A., Misof, B., 2013. Mitochondrial genome diversity in Ecdysozoa. Mol. Phylogenet. Evol.
Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucl. Acids Res. 35, D61-D65.
R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna , Austria www. r-project. org.
Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15, 957-966.
Rubinoff, D., Holland, B.S., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54, 952-961.
Shao, R., Barker, S.C., Mitani, H., Takahashi, M., Fukunaga, M., 2006. Molecula mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus *Leptotrombidium* (Acari: Acariformes). J. Mol. Evol. 63, 251-261.
Shao, R., Dowton, M., Murrell, A., Barker, S.C., 2003. Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol. Biol. Evol. 20, 1612-1619.
Simon, S., Hadrys, H., 2013. A comparative analysis of complete mitochondrial genomes among Hexapoda . Mol. Phylogenet. Evol. in press.
Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T. and others, 2008. The Trichoplax genome and the nature of placozoans. Nature 454, 955-960.
Stach, T., Braband, A., Podsiadlowski, L., 2010. Erosion of phylogenetic signal in tunicate mitochondrial genomes on different levels of analysis. Mol. Phylogenet. Evol. 55, 860-870.

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

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758-771.

Steinauer, M.L., Nickol, B.B., Broughton, R., Orti, G., 2005. First sequenced mitochondrial genome from the phylum Acanthocephala (*Leptorhynchoides thecatus*) and its phylogenetic position within Metazoa. J. Mol. Evol. 60, 706-715.

Stoeger, I., Schroedl, M., 2013. Mitogenomics does not resolve deep molluscan relationships (yet?). Mol. Phylogenet. Evol. in press.

Szabadkai, G., Duchen, M.R., 2008. Mitochondria: the hub of cellular Ca2+ signaling. Physiology 23, 84-94.

Waeschenbach, A., Telford, M.J., Porter, J.S., Littlewood, D.T.J., 2006. The complete mitochondrial genome of *Flustrellidra hispida* and the phylogenetic position of Bryozoa among the Metazoa. Mol. Phylogenet. Evol. 40, 195-207.

Webster, B.L., Copley, R.R., Jenner, R.A., kenzie-Dodds, J.A., Bourlat, S.J., Rota-Stabelli, O., Littlewood, D.T.J., Telford, M.J., 2006. Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral Ecdysozoan. Evol. Dev. 8, 502-510.

Wei, S.J., Shi, M., Chen, X.X., Sharkey, M.J., van, A.C., Ye, G.Y., He, J.H., 2010. New views on strand asymmetry in insect mitochondrial genomes. PLoS One. 5, e12708.

Xu, W., Jameson, D., Tang, B., Higgs, P.G., 2006. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. J. Mol. Evol. 63, 375-392.

Yokobori, S., Iseto, T., Asakawa, S., Sasaki, T., Shimizu, N., Yamagishi, A., Oshima, T., Hirose, E., 2008. Complete nucleotide sequences of mitochondrial genomes of two solitary entoprocts, *Loxocorone allax* and *Loxosomella aloxiata*: Implications for lophotrochozoan phylogeny. Mol. Phylogenet. Evol. 47, 612 628.

Zamzami, N., Susin, S.A., Marchetti, P., Hirsch, T., Gomez-Monterrey, I., Castedo, M., Kroemer, G., 1996. Mitochondrial control of nuclear apoptosis. J Exp. Med. 183, 1533-1544.

5.2. The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny

Marleen Perseke [a,1,†], Anja Golombek [b,2,†], Martin Schlegel [a], Torsten H. Struck [b,2]

[a] Institute of Biology: AG Molekulare Evolution und Systematik der Tiere, University of Leipzig, Talstrasse 33, D-04103 Leipzig, Germany; perseke@rz.uni-leipzig.de & schlegel@rz.uni-leipzig.de
[b] University of Osnabrück; FB05 Biology/Chemistry; AG Zoology; Barbarastr. 11; D- 49076 Osnabrück; Germany; A.Golombek@gmx.de
[1] Laboratory of Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Science, 164 West Xingang Road, 510301 Guangzhou, PR China
[2] Zoological Research Museum Alexander Koenig; Adenauerallee 160; D-53113 Bonn; Germany; torsten.struck.zfmk@uni-bonn.de
† Shared first authorship, both authors contributed equally to this work

Corresponding author:
Torsten H. Struck; Zoological Research Museum Alexander Koenig; Adenauerallee 160; D-53113 Bonn; Germany
Phone: +49-225-9122-401; Fax: +49-225-9122-212; E-Mail: torsten.struck.zfmk@uni-bonn.de

Abstract

Deuterostomia, one of the three major lineages of Bilateria, comprises many wellknown animals such as vertebrates, sea squirts, sea stars and sea urchins. Whereas monophyly of Deuterostomia and several subtaxa is well supported, the relationships of these to each other and, hence, deuterostome relationships are still uncertain. To address these issues in deuterostome phylogeny we analyzed datasets comprising more than 300 complete deuterostome mitochondrial genomes. Based on sequence information, the results revealed support for several relationships such as a basal position of *Xenoturbella* within Deuterostomia or for taxa like Craniota or Ambulacraria, but yielded also problems in some taxa, e.g. Tunicata, Pterobranchia and Ophiuroidea, due to long-branch artifacts. However, within tunicates the relationships are well supported. Variation in the genetic code was also informative and, e.g., supported the taxon Ambulacraria including Pterobranchia.

Key words: Tunicata, Ambulacraria, *Xenoturbella*, Echinodermata, Chordata, Mitochondrial Genomes

1. Introduction

The diverse morphological forms of Deuterostomia range from solitary to colonial bodyplans accompanied by sessile or free-living life styles. Traditionally, the group is recognized based on shared embryonic developmental patterns such as radial cleavage, gastrulation occuring at the vegetal pole, blastopore becoming the anus and the mouth being a secondary formation (deuterostomy), as well as coelom formation by enterocoely (summarized in Swalla, 2006). The close relationship of the three major lineages (i.e., Chordata comprising Craniota (Vertebrata), Cephalochordata (Acrania) and Tunicata (Urochordata); Hemichordata consisting of Enteropneusta and Pterobranchia; and Echinodermata) has been revealed in many molecular analyses, albeit with weak support only (see Hejnol et al., 2009; Philippe et al., 2011; Swalla and Smith, 2008 and references therein). The phylogenetic relationships within and between these three lineages are a longstanding issue with different implications for bodyplan and life style evolution within Deuterostomia and Bilateria (see Cameron et al., 2000 and references therein). Closer relationships of major lineages to other groups with partly less complex

morphological features such as *Xenoturbella* and Acoelomorpha have been hypothesized and used to assess the evolution of simple and complex morphological features of Bilateria (e.g., Adoutte et al., 2000; Hejnol et al., 2009; Philippe et al., 2011; Swalla and Smith, 2008).
Several embryological and larval features support monophyly of Chordata, e.g., a hollow nerve cord dorsal to a notochord, a postanal tail, an endostyle, and pharyngeal gill slits (see Stach, 2008; Swalla and Smith, 2008 and references therein). Most chordate taxa belong to the subgroup Craniota, whose monophyly is supported by autapomorphies like an endoskeleton, multilayered epidermis, endothelially lined blood vessels, and neural crest. The fish-shaped Cephalochordata is characterized by, among others, cyrtopodocytes. Tunicata exhibit a remarkable diversity of developmental and life history traits and their monophyly is supported, e.g., by a tunic or a pylorus gland (reviewed in Stach, 2008; Swalla et al., 2000).
The traditional view of Chordata phylogeny suggests Cephalochordata as sister to Craniota supported by the possession of a notochord and several other characters (Notochordata hypothesis) (Garstang, 1928; Stach, 2008). Analyses of molecular data, however, are uncertain in this respect. Whereas some analyses based on nuclear rDNA or mitochondrial data or on gene synteny substantiated the Notochordata hypothesis (e.g., Bourlat et al., 2009; Cameron et al., 2000; Putnam et al., 2008; Tsagkogeorga et al., 2009), others based on mitochondrial or phylogenomic data favored a closer relationship of Tunicata and Craniota (Olfactores hypothesis) (e.g., Delsuc et al., 2006; Singh et al., 2009; Stach et al., 2010). On the other hand, revision of morphological data found support for the Olfactores hypothesis by the discovery of a migratory cell population identified in Tunicata and Craniota, which is missing in Cephalochordata (Jeffrey, 2004; reviewed in Stach, 2008).
Echinodermata show a primary sessile life style and pentamerism of the adult body organization clearly distinguishing them from other deuterostomes (Brusca and Brusca, 1990). Hemichordata unite the sessile, colonial-living Pterobranchia and the solitary, worm-like Enteropneusta (e.g., Benito and Pardos, 1997). The phylogenetic positions of the hemichordate groups and Echinodermata with respect to each other and Chordata are uncertain based on morphological data (see Winchell et al., 2002 and references therein). Especially, a close relationship of Enteropneusta to Chordata (Cyrtotreta hypothesis) (see Schaeffer, 1987) or a monophyletic Hemichordata with a close relationship to Echinodermata (Ambulacraria hypothesis) (Metschnikoff, 1881) has been suggested. Analyses of rDNA, mitochondrial (mt) genome and phylogenomic data unequivocally support the latter hypothesis (Bourlat et al., 2006; Cameron et al., 2000; Halanych, 1995; Perseke et al., 2011; Swalla and Smith, 2008; Winchell et al., 2002), which is congruent to their similar larval morphologies, but in contrast to similarities in chorda-like structures and gill slits in adult forms of enteropneusts and chordates.
Traditionally, the lophophorate taxa (Branchiopoda, Phoronida, and Bryozoa) and Chaetognatha had been placed within or close to Deuterostomia as well (e.g., Barnes, 1968; Zimmer, 1973), but molecular data including mt genomes placed them within Protostomia (see Halanych, 2004 and references therein; Helfenbein et al., 2004; Marletaz et al., 2006 and references therein; Matus et al., 2006; Stechmann and Schlegel, 1999). In contrast, molecular analyses yielded a close relationship of *Xenoturbella* to Deuterostomia (Bourlat et al., 2006; Bourlat et al., 2003; Bourlat et al., 2009; Perseke et al., 2007; Philippe et al., 2011). *Xenoturbella* is a small, ciliated marine worm with an unusual and also simple morphology (Westblad, 1949). Based on morphological data several positions of *Xenoturbella* have been discussed: a close relationship to Acoelomorpha due to similarities in the nervous system and epithelial cells (see Nielsen, 2010 and references therein), to Enteropneusta and Holothuroidea (Echinodermata) due to similar epithelial structures (Reisinger, 1960), to Mollusca due to similar developmental features (Israelsson, 1997) or a basal bilaterian position (Ehlers and Sopott-Ehlers, 1997). Recently, phylogenomic analyses with increased data of Acoelomorpha yielded a sistergroup relationship of *Xenoturbella* and Acoelomorpha and placed them as sister

to all other Bilateria (Hejnol et al., 2009) or to Ambulacraria within Deuterostomia (Philippe et al., 2011), though with low bootstrap support for both placements as well as their sistergroup relationship.
Mitochondrial genomes of all deuterostome lineages and their subgroups have been determined. Most complete mt genomes belong to Craniota (1,391 genomes as of April 20th 2012 in the NCBI genome database), and their transcription and replication mechanisms are best understood (Fernandez-Silva et al., 2003), especially regarding the knowledge of dysfunctions causing specific diseases in humans (Lane, 2006). Several mitochondrial genomes are also known from Cephalochordata (8), Tunicata (12), Echinodermata (29), and Enteropneusta (3) as well as one for Pterobranchia (Bernt et al., 2012; Kon et al., 2007; Nohara et al., 2005; Perseke et al., 2011; Stach et al., 2010).
The sea urchin *Strongylocentrotus purpuratus* represents a model organism for developmental biology (e.g., Gilbert, 2000; Sodergren et al., 2006) and analyses of its mt genome yielded few differences to the craniote mt genome gene order, replication mechanisms and gene transcription (e.g., Cantatore et al., 1990; Mayhook et al., 1992). Moreover, the mtDNA of Cephalochordata is also similar to the one of Craniota with only a few gene order differences, but conserved sequence elements for replication initiation are missing in Cephalochordata (Nohara et al., 2005). Such conserved elements were also only rarely recovered in other mitochondrial genomes, e.g. in *Balanoglossus* (Castresana et al., 1998b) and *Xenoturbella* (Bourlat et al., 2009). Finally, Tunicata show a quite unique genetic code and unusual genome architectures of their mt genomes (see Gissi et al., 2010; Stach et al., 2010).
On the other hand, mt genomes were successfully applied at different levels of phylogenetic reconstruction using protein-sequence analyses, gene order arrangements, but also features of the genetic code and tRNA secondary structure (e.g., Perseke et al., 2010; Perseke et al., 2007; Perseke et al., 2011; Stach et al., 2010). For example, mitochondrial genome data support the deuterostome affiliation of *Xenoturbella* (Bourlat et al., 2009; Perseke et al., 2007), but not of Acoelomorpha (Mwinyi et al., 2010; Ruiz-Trillo et al., 2004).
As part of the "Deep Metazoan Phylogeny" project assessing the evolution and phylogeny of Metazoa using all available complete mt genomes of metazoan groups (except for Arthropoda and Craniota due to the shear amount of genomes available for these taxa and computational time) we addressed the phylogenetic relationships of major deuterostome lineages. Our analyses recovered with strong support the three chordata subgroups, Echinodermata and their five subgroups, Enteropneusta and Ambulacraria. *Xenoturbella* branched basal within Deuterostomia but distinct to the acoel *Symsagittifera*. The position of Tunicata was highly unstable due to long-branch artifacts resulting in the non-monophyly of both Chordata and Deuterostomia in the largest dataset.

2. Material and Methods

A detailed description of the pipeline to generate and analyze the datasets is given in Bernt et al. (2012) in this special issue. In brief, all complete mt genomes as of February 22nd 2010 were retrieved from the NCBI genome database and annotation errors in the NCBI entry were fixed using an automated pipeline. Following the extraction of the protein-coding genes from the genomes and their alignment using MAFFT (Katoh et al., 2005) combined with position masking by noisy (Drees et al., 2008), the number of arthropod and craniote taxa was reduced to balance between coverage of biodiversity and computation time. Unfortunately, the pterobranch genome had also to be excluded at this step due to the strong bias of the proteincoding genes that highly influenced the protein composition and disrupted the total alignment. The individual gene datasets were concatenated into a single supermatrix with more than 800 metazoan taxa (M800). To assess the influence of unstable taxa on the reconstruction,

reduced datasets were generated from the large dataset. First, taxa were reduced over all metazoan groups resulting in two datasets comprising only either 300 or 100 metazoan taxa (M300 or M100, respectively). Second, to especially investigate the placement of the long-branched tunicates, two datasets were generated from the large one consisting only of the deuterostome taxa plus three short-branched protostome outgroup taxa and additionally the five chaetognath taxa in one of the two datasets (D or DC, respectively). All datasets were analyzed employing the Maximum Likelihood method implemented in RAxML (Stamatakis et al., 2008) and robustness of nodes was assessed using the bootstrap (BS) approach.
To study genome rearrangements we used breakpoint analyses implemented in the web-based CREx software (Bernt et al., 2007). The gene orders were compared between the consensus order of Craniota (Boore, 1999) as well as of Cephalochordata (Nohara et al., 2005), the hypothetical basal order of Echinodermata (Perseke et al., 2010), and all known gene orders of Tunicata (Stach et al., 2010), of the hemichordates *Balanoglossus* (Castresana et al., 1998b) and *Saccoglossus* (NC_007438), of *Xenoturbella bocki* (Perseke et al., 2007) and the acoel *Symsagittifera roscoffensis* (Mwinyi et al., 2010). The gene orders were compared with two different gene sets: „all genes" included all 37 mitochondrial genes and „non-tRNA genes" included only the two ribosomal genes and the 13 protein-coding genes.

3. Results and Discussion

3.1 Chordate phylogeny and the position of Tunicata
Our analyses of mitochondrial (mt) sequences recovered the chordate groups Craniota, Cephalochordata and Tunicata as maximally supported monophyla (Fig. 1). Unique features in the mt genetic codes also supported monophyly of Tunicata and Craniota, respectively. Whereas the codons "AGA" and "AGG" were assigned to the amino acid glycine in Tunicata (Fig1. C-1) (see Kondow et al., 1999 and references therein), they were used as termination codons in Craniota (Fig. 1 C-2) (Ivanov et al., 2001; Osawa et al., 1992). The mt genetic code of Cephalochordata was identical to the invertebrate one (CodeTab 9, NCBI), though the codon "AGG" was either absent or occurred only in extremely low frequencies (Nohara et al., 2005). In contrast to the variable genetic code, the basal mt gene orders of Craniota and Cephalochordata were highly conserved differing in only four rearranged tRNA genes in Cephalochordata (Fig. 1 A-3) (Nohara et al., 2005). Moreover, the craniote arrangement (Fig. 1 A-2) may represent the basal arrangement of Deuterostomia (Boore, 1999; Bourlat et al., 2009; Castresana et al., 1998b) showing high similarities to the arrangements in the hemichordate *Balanoglossus* and in *Xenoturbella bocki* (3 and 14 breakpoints between the arrangements in the "non-tRNA gene" and "all genes" datasets, respectively; Tab. 1) and has been hypothesized to be similar to the basal arrangement of Metazoa (Lavrov and Lang, 2005). In contrast, the mt genome architectures of Tunicata were unique in several features with respect to the other deuterostomes (Stach et al., 2010). First, the gene content was variable with usually two additional tRNA genes (Fig. 1 A-1) and *atp8* as well as the two mt rRNA genes were considerably shorter than usually. Furthermore, *atp8* was absent (or too derived to get annotated) in the tunicate genome of *Clavelina lepadiformis* (Stach et al., 2010). Second, the AT content was higher. Third, all genes were located only on one strand. Fourth, extensive gene order rearrangements occurred nonetheless and precluded the reconstruction of a basal tunicate gene order or conclusions about phylogenetic relationships to other deuterstomes based on mt gene order (more than 11 and 27 breakpoints to other deuterostome gene orders in the "non-tRNA gene" and "all gene" datasets, respectively; Tab. 1). There was only one conserved gene block (*coxII-cytb*), which was observed in some Ascidiacea, that was, in all Phlebobranchia and some Stolidobranchia, as well as in the single Thaliacea, *Doliolum nationalis* (Gissi et al., 2010).

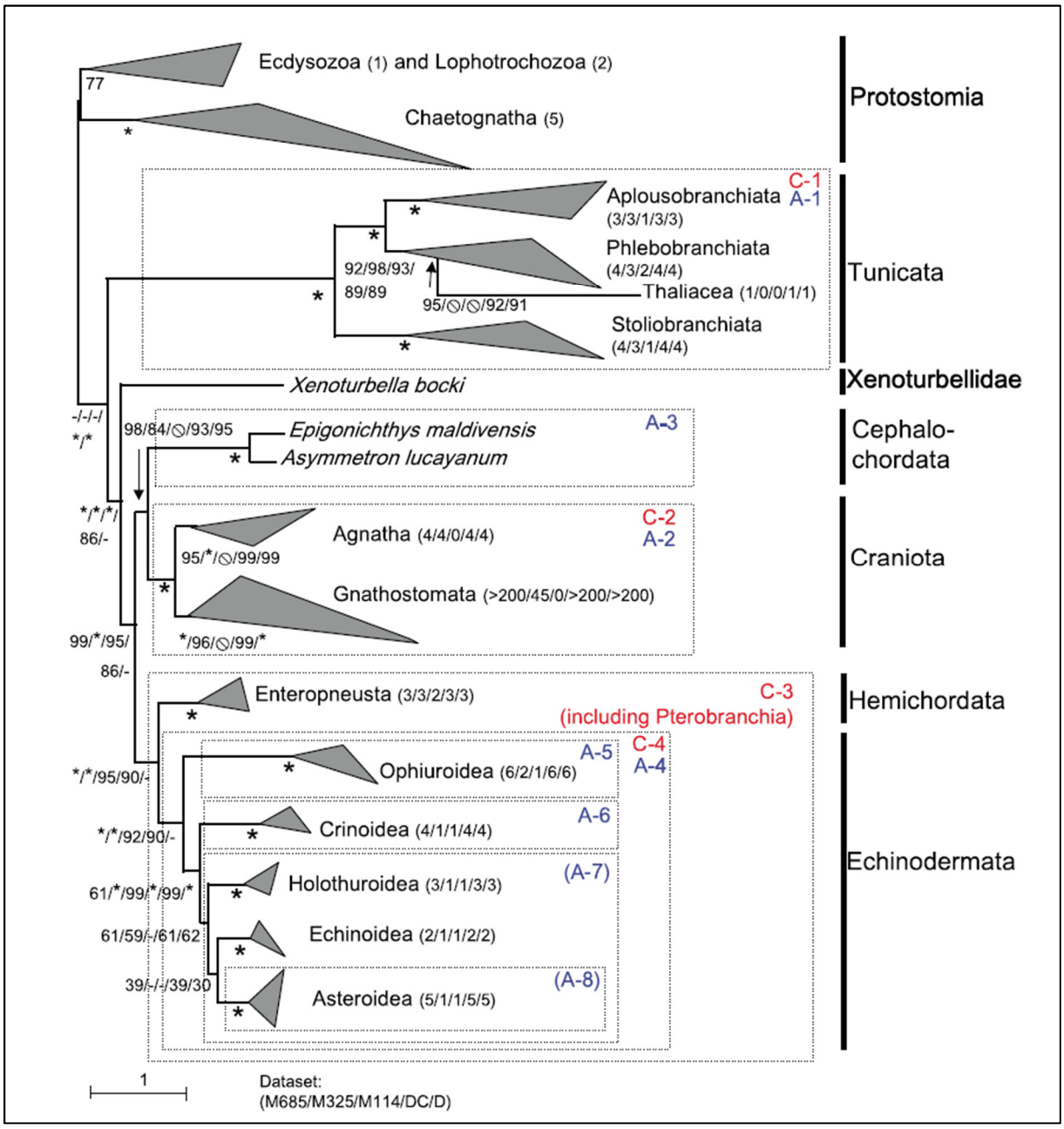


Fig. 1: ML tree of the dataset DC. The numbers on the nodes show support values from all analyses in the order of dataset M800, M300, M100, DC and D. Asterisksindicate highest support value obtained in all datasets (bold) or in an individual analysis. Species are collapsed to groups and the variability of the group is indicated by a triangle for the shortest and longest branch length. The numbers in parenthesis behind the groups show the number of analyzed species in the same order as shown for the support values. In addition, mt features of gene arrangements ("A") and genetic codes ("C") supporting a group are shown. Discontinuous lines frame the groups and the features are serially numbered in accordance to the text; a detailed description of these features is provided in the text. □ = not applicable due to limited taxon sampling.

Tab. 1: Pairwise gene order comparisons using breakpoint analyses (Watterson et al., 1982) implemented in CREx (Bernt et al., 2007). The number of breakpoints between gene order arrangements using "all genes" (lower triangle) or "non-tRNA genes" (upper triangle) is shown. For the 12 available tunicate genomes, only the highest and lowest breakpoint number is given.

	Craniota	Cephalochordata	*Xenoturbella bocki*	*Balanoglossus* spp.	*Saccoglossus kowalevskii*	Hypothetical Echinodermata	*Rhabdopleura compacta*	Tunicata	Acoela
Craniota		0	3	3	7	6	9	12-15	13
Cephalo-chordata	12		3	3	7	6	9	12-15	13
Xenoturbella bocki	14	19		3	6	8	9	13-15	13
Balanoglossus spp.	14	15	14		4	9	9	12-15	13
Saccoglossus kowalevskii	25	29	23	21		10	10	13-15	14
Hypothetical Echinodermata	27	29	27	27	29		11	13-15	14
Rhabdopleura compacta	35	35	35	33	35	34		13-15	14
Tunicata	29-36	30-36	28-36	30-36	32-36	30-36	32-36		10-14
Acoela	34	34	34	34	34	33	34	29-34	

Cephalochordata was placed as sister to Craniota with strong support in all our analyses (Fig. 1, BS ≥ 84) substantiating the Notochordata hypothesis. However, the placement of long-branched Tunicata was highly unstable. In the analyses using the large metazoan datasets M800, M300 or M100, respectively, Tunicata was the sistergroup of long-branched Acoela (Figs. 1-2 in Bernt et al., 2013 in this special issue). Based on the Deuterostomia dataset with Chaetognatha (dataset DC), Tunicata was sister to all other deuterostomes and in close proximity to the long-branched Chaetognatha (Fig. 1). In the deuterostome dataset without Chaetognatha (dataset D), Tunicata was the sistergroup of Ophiuroidea (data not shown), which is an echinoderm subgroup with obviously accelerated evolutionary rates (see Scouras et al., 2004). Hence, the placement of long-branched Tunicata strongly depended on the taxon sampling and was most likely affected by long-branch artifacts as our analyses showed that Tunicata was always drawn to other long branches in the dataset. Although this long-branch attraction of Tunicata precluded the resolution of the Chordata phylogeny, the relationships of the other deuterostome lineages to each other (i.e., not considering Tunicata) were identical in all analyses and well supported with BS values of 95 and higher (Fig. 1).

The difficult placement of Tunicata and failure to recover monophyly of the morphologically well-supported Chordata in sequence-based analyses of mt data had been reported before, albeit analyses by Singh et al. (2009) and Stach et al. (2010) were able to recover a monophyletic clade Chordata using mitochondrial sequences. In their analyses a sistergroup relationship of Tunicata and Craniota was supported, congruent with the Olfactores hypothesis, but with low bootstrap support. In contrast to our analyses, the taxon sampling was much smaller in these two studies (with only 31 and 28 deuterostome taxa, respectively), containing selected species (only four representatives of Craniota, none or one of Ophiuroidea). In the study of Stach et al.

(2010) no non-deuterostome outgroup taxa were present. In addition, Singh et al. (2009) used a Bayesian approach implementing a time-heterogenous model, which is known to alleviate the problems of long-branched taxa to a certain degree. Unfortunately, in our analyses with 100 taxa and more this approach is computationally not feasible in the moment. Moreover, as discussed above, the mitochondrial gene order of Tunicata also did not provide evidence for their phylogenetic placement as it was, e.g., the case for the long-branched Myzostomida (Bleidorn et al., 2007). Therefore, based on mitochondrial data the phylogenetic position of Tunicata remained uncertain due to their high variability at all levels.

3.1.1 Tunicate phylogeny

In all analyses the tunicate taxon Stolidobranchiata was sister to all other tunicates covered in our datasets with highest support values (Fig. 1). Moreover, in all analyses Aplousobranchiata has been recovered as a monophylum, whereas Thaliacea was placed within Phlebobranchiata and all placements were highly supported by bootstrap values (Fig. 1). Traditionally, Tunicata were classified into Ascidiacea (sea squirts) comprising Stolidobranchiata, Aplousobranchiata and Phlebobranchiata, Thaliacea (salps) and Appendicularia, for which neither a partial nor a complete mt genome is available yet (Stach and Turbeville, 2002). The same phylogeny as in Fig. 1 has also been recovered in other mt analyses (Singh et al., 2009; Stach et al., 2010). Moreover, an analysis based on 35 house-keeping genes also found a sistergroup relationship of Aplousobranchiata and Phlebobranchiata with Stolidobranchiata as sister to this clade and obtained highest support values at all nodes (Tsagkogeorga et al., 2010). Thaliacea were not covered by this study. No morphological data is known to date to corroborate the close relationship of Thaliacea and Phlebobranchiata, but 18S rDNA also strongly supported this result (e.g., Stach, 2009; Stach and Kirbach, 2009; Stach and Turbeville, 2002; Swalla et al., 2000; Zeng et al., 2006). Moreover, Pherophoridae (Phlebobranchiata), Aplousobranchiata and Appendicularia have in common the horizontal orientation of the tail. Hence, depending on a possible close relationship of Appendicularia and Aplousobranchiata as found in some other molecular studies and supported by morphological data, this feature might further substantiate a close relationship of Appendicularia, Aplousobranchiata and Phlebobranchiata (Stach and Kirbach, 2009). However, the position of Appendicularia was highly unstable within Tunicata and alternative positions have also been found (see Stach and Kirbach, 2009 and references therein). Moreover, several characters such as internal longitudinal blood vessels and a complex opening of the ciliated funnel support a clade of Stolidobranchiata, Phlebobranchiata and Thaliacea (e.g., Stach, 2009). Molecular studies of 18S rDNA also seem support this clade (e.g., Stach and Turbeville, 2002). However, in these analyses Aplousobranchiata exhibited extremely long branches in comparison to the other tunicates and more sophisticated analyses of the 18S rDNA data also found a close relationship of Aplousobranchiata, Phlebobranchiata and Thaliacea (Tsagkogeorga et al., 2009). In the mt data Aplousobranchiata did not show any increased substitution rates in comparison to the other tunicates (see Fig. 1). Finally, a morphological cladistic analysis of tunicates resulted in a large basal polytomy providing no resolution at all for Tunicata (Stach and Turbeville, 2002).

3.2 Phylogeny of Ambulacraria

Almost all of our analyses supported the Ambulacraria hypothesis, i.e. the sistergroup relationship of Hemichordata (represented by Enteropneusta) and Echinodermata with high support values (Fig. 1, BS $\geq$ 95). Strong support for the Ambulacraria hypothesis was first gained from a study based on 18S rDNA data (Cameron et al., 2000). Tunicata was placed within Ambulacraria only in the deuterostome dataset without Chaetognatha (D) rendering Ambulacraria paraphyletic (data not shown). However and as discussed above, this is most likely due to the misplacement of long-branched Tunicata as sister to Ophiuroidea. As Perseke et al. (2011) could show, the placement of Pterobranchia within both Hemichordata and

Ambulacraria was problematic to identify with mt sequence analyses and gene order comparisons due to the unusual genome of the pterobranch *Rhabdopleura compacta*. It is for this reason that this genome had to be excluded from this study. However, analyses based on nuclear 18S rDNA and mitochondrial 16S rDNA and *cytochrome b* genes strongly supported a placement of Pterobranchia within Hemichordata (Cannon et al., 2009). In addition, the mt genetic code of Ambulacraria (including the pterobranch *Rhabdopleura*) was characterized by the assignment of "ATA" to isoleucine in contrast to Chordata or *Xenoturbella* (Fig. 1, C-3) (see Castresana et al., 1998a; Perseke et al., 2011) supporting the inclusion of Pterobranchia in Ambulacraria. Moreover, a unique anticodon sequence CUU for the $tRNA_{Lys}$ was further suggested as an autapomorphy for Ambulacraria (Perseke et al., 2011) and that could have caused the assignment of "AAA" to asparagine in mt genomes of Echinodermata (Asakawa et al., 1995), of "AGG" to lysine in the one of the pterobranch *Rhabdopleura* (Perseke et al., 2011) and lack of "AGG" in the enteropneust genome of *Balanoglossus* (Castresana et al., 1998a). Thus, mt data supported Ambulacraria as proposed on larval features (Metschnikoff, 1881) and recovered in many molecular analyses (e.g., Cameron et al., 2000; Halanych, 1995; Hejnol et al., 2009; Perseke et al., 2011; Winchell et al., 2002).

The monophyly of Echinodermata was supported by different levels of mt genome information: (i) all sequence analyses recovered the monophyly (Fig. 1 BS≥ 95, with exception of dataset D), (ii) the genetic code "AGA" was assigned to asparagine (Fig.1, C-4) and (iii) the gene order showed the deuterostome arrangement of proteincoding genes except for the location of *nd4L* between *coxI* and *coxII* (Fig. 1 A-4). Assuming the hypothetical basal echinoderm arrangement proposed by Perseke et al. (2010), two further gene order rearrangements occured in the lineage to Echinodermata: an inversion of the fragment containing *nd1* and *nd2* and the inversion of 16S rRNA leading to a medium amount of breakpoints compared to other deuterostome arrangements (Tab. 1). Thus, monophyly of Ambulacraria and Echinodermata was well supported suggesting a solitary lifestyle of the last common ancestor of Deuterostomia and convergent origins of colonial lifestyles within Deuterostomia (i.e., Pterobranchia and within Tunicata). However, further sequence data including mt genomes are still necessary for the hemichordate subgroups Pterobranchia and Enteropneusta for resolving their relationship to each other as well as within Ambulacraria and, hence, the evolution of coloniality within deuterostomes (Swalla and Smith, 2008).

3.2.1 Echinoderm phylogeny

The five extant echinoderm subgroups Crinoidea, Ophiuroidea, Asteroidea, Echinoidea and Holothuroidea were recovered as monophyla in all our analyses with highest support values (Fig.1), congruent with other molecular markers and their obviously different morphology. Crinoids (sea lilies) filter the water with many branched arms connected to a central cup-shaped body and are sessile at least in juvenile stages (Brusca and Brusca, 1990). Whereas asteroids (sea stars) possess five or more thick arms extending from a central disk, ophiuroids (brittle stars) have five thin arms containing vertebrae-like elements for highly flexible movements (Brusca and Brusca, 1990). Echinoids (sea urchins) are rotund with a strong calcareous teeth apparatus (Aristotle's lantern) and holothuroids (sea cucumbers) are bilateral symmetric, but both show the pentamery still in different organs (Brusca and Brusca, 1990). The gene orders of mt genomes further substantiate the monophyly of Ophiuroidea (by one inversion of a fragment containing the genes *cytb*, 12S rRNA and several tRNA genes; Fig.1 A-5), Crinoidea (by an inversion of a small fragment containing 12S rRNA, the flanked tRNA genes and the non-coding region with the elements for the replication; Fig. 1 A-6) and Asteroidea (by an inversion of a fragment containing *nd1*, *nd2*, 16S rRNA and several tRNA genes; Fig. 1 A-8) (see Perseke et al., 2010 and references therein). Ophiuroidea was a well-supported sister to all other echinoderms in all our analyses (Fig. 1, not considering Tunicata in dataset D). Within the clade of the remaining four taxa, most analyses supported a sistergroup relationship of

Crinoidea to a clade of Asteroidea, Echinoidea and Holothuroidea (Fig. 1). Only the analyses with one representative species per taxon (M300, M100) resulted in a basal branching of Holothuroidea (BS = 73) and a sistergroup relationship of Crinoidea to Asteroidea (BS = 44) (Figs. 2 & Suppl. Fig. in Bernt et al., 2013 in this special issue). The resolution in the clade of Holothuroidea, Asteroidea and Echinoidea was always low and different relationships were proposed by different datasets (Fig.1).

These relationships of the echinoderm subgroups were similar to previous analyses of echinoderm mt sequences, but in strong contrast to the Eleutherozoa hypothesis placing Crinoidea as sister group to all other echinoderms which is supported by morphological characters as well as nuclear rRNA analyses (e.g., Janies, 2001; Littlewood et al., 1997; Mallatt and Winchell, 2007). mtDNA of Ophiuroidea showed highest substitution rates, which may lead to the basal branching in phylogenetic analyses even with highest support (see Scouras et al., 2004). The placement of Crinoidea should also be taken with care because all crinoid mtDNAs showed opposed directed strand-specific nucleotide pressure that was most likely caused by the gene inversion including the origin of replication in mt genomes of Crinoidea (Scouras and Smith, 2006). mt genomes supported also the monophyly of Holothuroidea, Asteroidea, and Echinoidea by highly similar gene orders (Fig.1 A-7). Although these were also similar to the craniote arrangement (see Boore, 1999), recent analyses suggest that this similarity arose by convergence due to an inversion from a fragment containing the genes *nd1*, *nd2*, 16S rRNA and few tRNA genes in comparison to the hypothetical basal echinoderm ground pattern (Perseke et al., 2010). Whereas a sistergroup relationship of Echinoidea and Asteroidea was recovered by most of our analyses (Fig. 1), others found sistergroup relationships of either Holothuroidea and Asteroidea or Holothuroidea and Echinoidea. However, nodal support was low for all three groupings in our analyses. The sistergroup relationship of Holothuroidea and Echinoidea has been recovered in our largest dataset (M800) and was congruent with morphological and nuclear rRNA results (e.g., Littlewood et al., 1997; Mallatt and Winchell, 2007) forming the taxon Echinozoa.

3.3 Xenoturbella and Acoelomorpha affiliation to Deuterostomia

Our analyses of mt sequences placed *Xenoturbella* always as a well-supported sister to all other deuterostomes except for tunicates (Fig. 1), similar to most previous analyses of mt sequence data. In contrast, a close relationship to the acoel *Symsagitiffera* was never recovered in any of our metazoan datasets. Instead longbranched Acoela was always sister to long-branched Tunicata (Figs. 1-2 & Suppl. Fig. in Bernt et al., 2013 in this special issue). Thus, our analyses revealed neither a sistergroup relationship of *Xenoturbella* to Acoela nor to Ambulacraria. The close relationship of *Xenoturbella* to Deuterostomia was also substantiated by the mt gene order, which was similar to those of chordates and the hemichordate *Balanoglossus* (3 and 14 breakpoints between the arrangements in the "non-tRNA gene" and "all gene" datasets, respectively; Tab. 1) and by mt sequence elements for replication initiation comparable to craniote genomes (Bourlat et al., 2009). However, the mt genetic code of *Xenoturbella* was identical to the invertebrate genetic code (CodeTab 5, NCBI) and the codons "AGA" and "AGG", which are often reassigned in deuterostomes, were lacking in *Xenoturbella* mt genomes. Therefore, a closer relationship to Ambulacraria was not supported by any apomorphic mt feature which was suggested by several EST and 18S rRNA analyses (Bourlat et al., 2006; Bourlat et al., 2003), some analyses of mt sequences (Bourlat et al., 2006; Bourlat et al., 2003) and immunoprecipitation analyses (Stach et al., 2005).

The basal position of *Xenoturbella* in Bilateria and a close relationship to Acoelomorpha was suggested by *hox* gene analyses (Fritzsch et al., 2008) as well as from a phylogenomic study (Hejnol et al., 2009) assuming similarities in the mt genome of *Xenoturbella* to Deuterstomia as plesiomorphic features. However, as in the case of mt sequence analyses (Figs. 1-2 & Suppl. Fig. in Bernt et al., 2013 in this special issue), Acoelomorpha exhibited also in these analyses

extremely long branches rendering their phylogenetic placement within Bilateria generally problematic. Another phylogenomic study (Philippe et al., 2011) also recovered a sistergroup relationship of *Xenoturbella* and Acoelomorpha, however, with a sistergroup relationship to Ambulacraria. In contrast, the available partial and complete mt genomes of Acoelomorpha (Mwinyi et al., 2010; Ruiz-Trillo et al., 2004) are very unique in Metazoa and showed no similarity in mt genome architecture to other deuterostomes (high number of breakpoints to all other deuterostome arrangements; Tab. 1). The sparsity of *hox* genes in *Xenoturbella* could be either a plesiomorphic condition as indicated by our results or a reduction in number due to a secondary simplification of their body organization assuming a placement within Deuterostomia. mt data clearly favor as sistergroup relationship of *Xenoturbella* and the other deuterostomes. Hence, the simple bauplan of *Xenoturbella* and its basal phylogenetic position within Deuterostomia in combination with the basal placement of Acoelomorpha within Bilateria suggests that both the last common ancestor of Deuterostomia and Acoelomorpha did not exhibit a high degree of morphological complexity.

4. Conclusion

Our comprehensive analyses based on more than 300 complete deuterostome mitochondrial genomes revealed that mitochondrial (mt) data has the potential to solve several issues of deuterostome phylogeny and evolution. Besides mt sequencebased phylogenetic analyses other mt features such as the genetic code or gene order can be exploited. Furthermore, we showed that monophyly of each major deuterostome lineage and some subgroups (i.e., Craniota, Cephalochordata, Tunicata, Hemichordata, Echinodermata and their subgroups Crinoidea, Ophiuroidea, Asterioidea, Holothuroidea and Echinoidea) was well supported by mt data at different levels of information. Moreover, mt data also clearly supported monophyly of Ambulacraria, a clade comprising Hemichordata and Echinodermata, as well as a basal position of *Xenoturbella* within deuterostomes. However, the problematic placement of Tunicata and Ophiuroidea as well as the necessary exclusion of Pterobranchia early on in the analyses also point to eventual problems of mt data. In some taxa the variability of the mt data is very high precluding a definite conclusion regarding the placement of these taxa based on mt data alone (e.g., Tunicata, Pterobranchia, and Ophiuroidea). On the other hand, the tunicate phylogeny shows that mt data can still be useful within these taxa.

5. Acknowledgements

We thank Matthias Bernt, Alexander Donath as well as their colleagues in the bioinformatic group of Peter F. Stadler (University of Leipzig) for their help in analyzing the data. This work was support by the grants DFG-STR-683/5-2 & 6-1 to T.H.S and DFG- SCHL 229/14-1 to M.S. from the Deutsche Forschungsgemeinschaft.

6. References

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, 4453-4456.

Asakawa, S., Himeno, H., Miura, K., Watanabe, K., 1995. Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics 140, 1047-1060.

Barnes, R.D., 1968. Invertebrate zoology. WB Saunders Co., Philadelphia, PA.

Benito, J., Pardos, F., 1997. Hemichordata. In: Harrison, M.A. (Ed.), Microscopic Anatomy of Invertebrates. Wiley Liss., New York, pp. 15-101.

Bernt, M., Bleidorn, C., Braband, A., Dambach, J., Donath, A., Fritzsch, G., Hadrys, H., Jörger, K., Jühling, F., Meusemann, K., Middendorf, M., Misof, B., Perseke, M., Podsiadlowski, L., von Reumont, B., Schierwater, B., Schlegel, M., Schrödl, M., Simon, S., Stadler, P.F., Stöger, I., Struck, T.H.,

Podsiadlowski, L., 2012. A comprehensive analysis of metazoan mitochondrial genomes and animal phylogeny. (in this special issue). Mol. Phylogenet. Evol.

Bernt, M., Merkle, D., Ramsch, K., Fritzsch, G., Perserke, M., Bernhard, D., Schlegel, M., Stadler, P., Middendorf, M., 2007. CREx: Inferring Genomic Rearrangements Based on Common Intervals. Bioinformatics 23, 2957-2958.

Bleidorn, C., Eeckhaut, I., Podsiadlowski, L., Schult, N., McHugh, D., Halanych, K.M., Milinkovitch, M.C., Tiedemann, R., 2007. Mitochondrial genome and nuclear sequence data support Myzostomida as part of the annelid radiation. Mol. Biol. Evol. 24, 1690-1701.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767- 1780.

Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B., Heyland, A., Moroz, L.L., Copley, R.R., Telford, M.J., 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444, 85-88.

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

Bourlat, S.J., Rota-Stabelli, O., Lanfear, R., Telford, M.J., 2009. The mitochondrial genome structure of *Xenoturbella bocki* (phylum Xenoturbellida) is ancestral within the deuterostomes. BMC Evol. Biol. 9, 107.

Brusca, R.C., Brusca, G.J., 1990. Invertebrates, First edition. Sinauer Associates, Inc., Sunderland, Massachusetts.

Cameron, C.B., Garey, J.R., Swalla, B.J., 2000. Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. Proc. Natl. Acad. Sci. U. S. A. 97, 4469-4474.

Cannon, J.T., Rychel, A.L., Eccleston, H., Halanych, K.M., Swalla, B.J., 2009. Molecular phylogeny of hemichordata, with updated status of deep-sea enteropneusts. Mol. Phylogenet. Evol. 52, 17-24.

Cantatore, P., Roberti, M., Loguercio Polosa, P., Mustich, A., Gadalet, M.A., 1990. Mapping and characterization of *Paracentrotus lividus* mitochondrial transcripts: multiple and overlapping transcription units. Curr. Genet. 17, 235-245.

Castresana, J., Feldmaier-Fuchs, G., Paabo, S., 1998a. Codon reassignment and amino acid composition in hemichordate mitochondria. Proc. Natl. Acad. Sci. U. S. A. 95, 3703-3707.

Castresana, J., Feldmaier-Fuchs, G., Yokobori, S.-i., Satoh, N., Paabo, S., 1998b. The mitochondrial genome of the hemichordate *Balanoglossus carnosus* and the evolution of deuterostome mitochondria. Genetics 150, 1115-1123.

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

Drees, A.W.M., Flamm, C., Fritzsch, G., Grünewald, S., Kruspe, M., Prohaska, S.J., Stadler, P.F., 2008. Noisy: identification of problematic columns in multiple sequence alignments. Algorithms Mol. Biol. 3, 7.

Ehlers, U., Sopott-Ehlers, B., 1997. Ultrastructure of the subepidermal musculature of *Xenoturbella bocki*, the adelphotaxon of the Bilateria. Zoomorphology 117, 71-79.

Fernandez-Silva, P., Enriquez, J.A., Montoya, J., 2003. Replication and transcription of mammalian mitochondrial DNA. Exp. Physiol. 88, 41-56.

Fritzsch, G., Böhme, M.U., Thorndyke, M., Nakano, H., Israelsson, O., Stach, T., Schlegel, M., Hankeln, T., Stadler, P.F., 2008. PCR survey of *Xenoturbella bocki* Hox genes. J. Exp. Zool. (Mol Dev Evol) 310B, 278 284.

Garstang, W., 1928. The morphology of the Tunicata and its bearing on the phylogeny of the Chordata. Q. J. Microsc. Sci. 72, 51-187.

Gilbert, S.F., 2000. Developmental Biology. Sinauer Associates, Sunderland, MA.

Gissi, C., Pesole, G., Mastrototaro, F., Iannelli, F., Guida, V., Griggio, F., 2010. Hypervariability of Ascidian Mitochondrial Gene Order: Exposing the Myth of Deuterostome Organelle Genome Stability. Mol. Biol. Evol. 27, 211-215.

Halanych, K.M., 1995. The phylogenetic position of the pterobranch hemichordates based on 18S rDNAsequence data. Mol. Phylogenet. Evol. 4, 72-76.

Halanych, K.M., 2004. The new view of animal phylogeny. Annu. Rev. Ecol. Evol. Syst. 35, 229-256.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E.G., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W., 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. R. Soc. B: Biol. Sci. 276, 4261-4270.

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

Israelsson, O., 1997. …and molluscan embryogenesis. Nature 390, 32.

Ivanov, V., Beniaminov, A., Mikheyev, A., Minyat, E., 2001. A mechanism for stop codon recognition by the ribosome: A bioinformatic approach. RNA 7, 1683-1692.

Janies, D., 2001. Phylogenetic relationships of extant echinoderm classes. Can. J. Zool. 79, 1232-1250.

Jeffrey, W.R., 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature 431, 696-699.

Katoh, K., Kuma, K.-i., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518.

Kon, T., Nohara, M., Yamamoue, Y., Fujiwara, Y., Nishida, M., Nishikawa, T., 2007. Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences. BMC Evol. Biol. 7, 127.

Kondow, A., Suzuki, T., Yokobori, S.-i., Ueda, T., Watanabe, K., 1999. An extra tRNAGly(U*CU) found in ascidian mitochondria responsible for decoding nonuniversal codons AGA/AGG as glycine. Nucleic Acids Res. 27, 2554-2559.

Lane, N., 2006. Mitochondrial disease: powerhouse of disease. Nature 440, 600-602.

Lavrov, D.V., Lang, B.F., 2005. Poriferan mtDNA and Animal Phylogeny Based on Mitochondrial Gene Arrangements. Syst. Biol. 54, 651 - 659.

Littlewood, D.T.J., Smith, A.B., Clough, K.A., Emson, R.H., 1997. The interrelationships of the echinoderm classes: morphological and molecular evidence. Biol. J. Linn. Soc. Lond. 61, 409-438.

Mallatt, J., Winchell, C.J., 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. Mol. Phylogenet. Evol. 43, 1005-1022.

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

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

Mayhook, A.G., Rinaldi, A.M., Jacobs, H.T., 1992. Replication Origins and Pause Sites in Sea Urchin Mitochondrial DNA. Proc. R. Soc. London, B 248, 85-94.

Metschnikoff, V.E., 1881. Über die systematische Stellung von *Balanoglossus*. Zool. Anz. 4, 139-157.

Mwinyi, A., Bailly, X., Bourlat, S.J., Jondelius, U., Littlewood, D.T.J., Podsiadlowski, L., 2010. The phylogenetic position of Acoela as revealed by the complete mitochondrial geneome of *Symsagittifera roscoffensis*. BMC Evol. Biol. 10, 309.

Nielsen, C., 2010. After all: *Xenoturbella* is an acoelomorph! Evol. Devel. 12, 241-243.

Nohara, M., Nishida, M., Miya, M.U., Nishikawa, T., 2005. Evolution of the Mitochondrial Genome in Cephalochordata as Inferred from Complete Nucleotide Sequences from Two *Epigonichthys* Species. J. Mol. Evol. 60, 526-537.

Osawa, S., Jukes, S.T., Watanabe, K., Muto, A., 1992. Recent evidence for evolution of the genetic code. Microbiol. Rev. 56, 229-264.

Perseke, M., Bernhard, D., Fritzsch, G., Brümmer, F., Stadler, P.F., Schlegel, M., 2010. Mitochondrial genome evolution in Ophiuroidea, Echinoidea, and Holothuroidea: Insights in phylogenetic relationships of Echinodermata. Mol. Phylogenet. Evol. 56, 201-211.

Perseke, M., Hankeln, T., Weich, B., Fritzsch, G., Stadler, P., Israelsson, O., Bernhard, D., Schlegel, M., 2007. The mitochondrial DNA of *Xenoturbella bocki*: genomic architecture and phylogenetic analysis. Theory Biosci. 126, 35-42.

Perseke, M., Hetmank, J., Bernt, M., Stadler, P.F., Schlegel, M., Bernhard, D., 2011. The enigmatic mitochondrial genome of *Rhabdopleura compacta* (Pterobranchia) reveals insights into selection of an efficient tRNA system and supports monophyly of Ambulacraria. BMC Evol. Biol.

Philippe, H., Brinkmann, H., Copley, R.R., Moroz, L.L., Nakano, H., Poustka, A.J., Wallberg, A., Peterson, K.J., Telford, M.J., 2011. Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. Nature 470, 255 258.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutierrez, E., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W.H., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064-1071.

Reisinger, E., 1960. Was ist *Xenoturbella*? Z. Wiss. Zool. 164, 188-198.

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

Schaeffer, B., 1987. Deuterostome monophyly and phylogeny. Evolutionary Biology 21, 179-235.

Scouras, A., Beckenbach, K., Arndt, A., Smith, J.M., 2004. Complete mitochondrial genome DNA sequence for two ophiuroids and a holothuroid: the utility of protein gene sequence and gene maps in the analyses of deep deuterostome phylogeny. Mol. Phylogenet. Evol. 31, 50-65.

Scouras, A., Smith, J.M., 2006. The complete mitochondrial genomes of the sea lily *Gymnocrinus richeri* and the feather star *Phanogenia gracilis*: signature nucleotide bias and unique nad4L gene rearrangement within crinoids. Mol. Phylogenet. Evol. 39, 323-334.

Singh, T.R., Tsagkogeorga, G., Delsuc, F., Blanquart, S., Shenkar, N., Loya, Y., Douzery, E.J.P., Huchon, D., 2009. Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. BMC Genomics 10, 534.

Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., Coffman, J.A., Dean, M., Elphick, M.R., Ettensohn, C.A., Foltz, K.R., Hamdoun, A., Hynes, R.O., Klein, W.H., Marzluff, W., McClay, D.R., Morris, R.L., Mushegian, A., Rast, J.P., Smith, L.C., Thorndyke, M.C., Vacquier, V.D., Wessel, G.M., Wray, G., Zhang, L., Elsik, C.G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M.J., Mackey, A.J., Maglott, D., Panopoulou, G., Poustka, A.J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C.A., Worley, K., Durbin, K.J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M.L., Jackson, A.R., Milosavljevic, A., Tong, M., Killian, C.E., Livingston, B.T., Wilt, F.H., Adams, N., Bellé, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Genevière, A.- M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A.J., Goldstone, J.V., Cole, B., Epel, D., Gold, B., Hahn, M.E., Howard-Ashby, M., Scally, M., Stegeman, J.J., Allgood, E.L., Cool, J., Judkins, K.M., McCafferty, S.S., Musante, A.M., Obar, R.A., Rawson, A.P., Rossetti, B.J., Gibbons, I.R., Hoffman, M.P., Leone, A., Istrail, S., Materna, S.C., Samanta, M.P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K.-F., Brandhorst, B.P., Whittle, J., Berney, K., Bottjer, D.J., Calestani, C., Peterson, K., Chow, E., Yuan, Q.A., Elhaik, E., Graur, D., Reese, J.T., Bosdet, I., Heesun, S., Marra, M.A., Schein, J., Anderson, M.K., Brockton, V., Buckley, K.M., Cohen, A.H., Fugmann, S.D., Hibino, T., Loza-Coll, M., Majeske, A.J., Messier, C., Nair, S.V., Pancer, Z., Terwilliger, D.P., Agca, C., Arboleda, E., Chen, N., Churcher, A.M., Hallböök, F., Humphrey, G.W., Idris, M.M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R.P., Raible, F., Rowe, M., Taylor, J.S., Tessmar-Raible, K., Wang, D., Wilson, K.H., Yaguchi, S., Gaasterland,T., Galindo, B.E., Gunaratne, H.J., Juliano, C., Kinukawa, M., Moy, G.W., Neill, A.T., Nomura, M., Raisch, M., Reade, A., Roux, M.M., Song, J.L., Su, Y.-H., Townley, I.K., Voronina, E., Wong, J.L., Amore, G., Branno, M., Brown, E.R., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Röttinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S.-Y., Xu, R., Brown, C.T., Chen, L., Gray, R.F., Lee, P.Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L.V., Okwuonu, G., Parker, D., Pu, L.-L., Thorn, R., Wright, R., 2006. The Genome of the Sea Urchin *Strongylocentrotus purpuratus*. Science 314, 941-952.

Stach, T., 2008. Chordate phylogeny and evolution: a not so simple three-taxon problem. J. Zool. 276, 117-141.

Stach, T., 2009. Anatomy of the trunk mesoderm in tunicates: homology considerations and phylogenetic interpretation. Zoomorphology 128, 97-109.

Stach, T., Braband, A., Podsiadlowski, L., 2010. Erosion of phylogenetic signal in tunicate mitochondrial genomes on different levels of analysis. Mol. Phylogenet. Evol. 55, 860-870.

Stach, T., Dupont, S., Israelsson, O., Fauville, G., Nakano, H., Kånneby, T., Horndyke, M., 2005. Nerve cells of *Xenoturbella bocki* (phylum uncertain) and *Harrimania kupfferi* (Enteropneusta) are positively immunireactive to antibodies raised against echinoderm neuropeptids. J. Mar. Biol. Ass. UK 85, 1519 1524.

Stach, T., Kirbach, A., 2009. Larval convergence in a colonial tunicate: the organization of the sarcotubular complex in *Ecteinascidia turbinata* (Perophoridae, Phlebobranchiata, Tunicata, Chordata). Zoomorphology 128, 1-11.

Stach, T., Turbeville, J.M., 2002. Phylogeny of Tunicata inferred from molecular and morphological characters. Mol. Phylogenet. Evol. 25, 408-428.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web-servers. Syst. Biol. 75, 758-771.

Stechmann, A., Schlegel, M., 1999. Analysis of the complete mitochondrial DANN sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. Proc. R. Soc. B: Biol. Sci. 266, 1-10.

Swalla, B.J., 2006. Building divergent body plans with similar genetic pathways. Heredity 97, 235-243.

Swalla, B.J., Cameron, C.B., Corley, L.S., Garey, J.R., 2000. Urochordates Are Monophyletic Within the Deuterostomes. Syst. Biol. 49, 52-64.

Swalla, B.J., Smith, A.B., 2008. Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363, 1557-1568.

Tsagkogeorga, G., Turon, X., Galtier, N., Douzery, E., Delsuc, F., 2010. Accelerated Evolutionary Rate of Housekeeping Genes in Tunicates. J. Mol. Evol. 71, 153-167.

Tsagkogeorga, G., Turon, X., Hopcroft, R., Tilak, M.-K., Feldstein, T., Shenkar, N., Loya, Y., Huchon, D., Douzery, E., Delsuc, F., 2009. An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. BMC Evol. Biol. 9, 187.

Westblad, E., 1949. *Xenoturbella bocki* n.g, n.sp, a peculiar, primitive turbellarian type. Arkiv för Zoologi 1, 3 - 29.

Winchell, C.J., Sullivan, J., Cameron, C.B., Swalla, B.J., Mallatt, J., 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol. Biol. Evol. 19, 762 776.

Zeng, L., Jacobs, M.W., Swalla, B.J., 2006. Coloniality has evolved once in Stolidobranch Ascidians. Integr. Comp. Biol. 46, 255-268.

Zimmer, R.L., 1973. Morphological and developmental affinities of the lophophorates. In: Larwood, G.P. (Ed.), Living and Fossil Bryozoa. Academic Press, New York, pp. 593-599.

5.3. Platyzoan paraphyly based on phylogenomic data supports a nonacoelomate ancestry of Spiralia

Torsten H. Struck[a,b], Wey-Fabrizius, Alexandra R.[c], Golombek, Anja[a], Hering, Lars[d], Weigert, Anne[e], Bleidorn, Christoph[e], Klebow, Sabrina[c], Iakovenko, Nataliia[f,g], Hausdorf, Bernhard[h], Petersen, Malte[a], Kuck, Patrick[a], Herlyn, Holger[i], Hankeln, Thomas[c]

Author affiliation

[a] *Zoological Research Museum Alexander Koenig; Adenauerallee 160; 53113 Bonn; Germany*

[b] *University of Osnabruck; FB05 Biology/Chemistry; AG Zoology; Barbarastr. 11; 49069 Osnabruck; Germany*

[c] *Johannes Gutenberg University; Institute of Molecular Genetics, Biosafety Research and Consulting; Johann-Joachim-Becher-Weg 30A, 55099 Mainz; Germany*

[d] *University of Leipzig; Institute for of Biology II; Animal Evolution and Development; Talstr. 33; 04103 Leipzig; Germany*

[e] *University of Leipzig; Institute of Biology; Molecular Evolution and Systematics of Animals; Talstr. 33; 04103 Leipzig; Germany*

[f] *Department of Biology and Ecology; Ostravian University in Ostrava; Chittussiho 10; 71000, Ostrava; Czech Republic*

[g] *Department of Invertebrate Fauna and Systematics; Schmalhausen Institute of Zoology NAS of Ukraine; Bogdana Khmelnitskogo str. 15; 01601 Kyiv; Ukraine*

[h] *Zoological Museum, University of Hamburg, Martin-Luther-King-Platz 3, 20146 Hamburg, Germany*

[i] *Johannes Gutenberg University; Institute of Anthropology; Anselm-Franz-von-Bentzel-Weg 7; 55099 Mainz; Germany*

Corresponding author:

Torsten H. Struck; Zoological Research Museum Alexander Koenig; Adenauerallee 160; 53113 Bonn; Germany; torsten.struck.zfmk@uni-bonn.de; Phone: +49-228-9122401; Fax: +49-228-9122212

Abstract

Based on molecular data three major clades have been recognized within Bilateria: Deuterostomia, Ecdysozoa and Spiralia. Within Spiralia, small-sized and simply organized animals such as flatworms, gastrotrichs and gnathostomulids have recently been grouped together as Platyzoa. However, the representation of putative platyzoans was low in the respective molecular phylogenetic studies, in terms of both, taxon number and sequence data. Furthermore, increased substitution rates in platyzoan taxa raised the possibility that monophyletic Platyzoa represents an artefact due to long-branch attraction. In order to overcome such problems, we employed a phylogenomic approach, thereby substantially increasing i) the number of sampled species within Platyzoa and ii) species-specific sequence coverage in datasets of up to 82,162 amino acid positions. Using established and new measures (long-branch score) we disentangled phylogenetic signal from misleading effects such as long-branch attraction. In doing so, our phylogenomic analyses did not recover a monophyletic origin of platyzoan taxa that, instead, appeared paraphyletic with respect to the other spiralians. Platyhelminthes and Gastrotricha formed a monophylum, which we name Rouphozoa. To the exclusion of Gnathifera, Rouphozoa and all other spiralians represent a monophyletic group, which we name Platytrochozoa. Platyzoan paraphyly suggests that the last common ancestor of Spiralia was a simple-bodied organism lacking coelomic cavities, segmentation and complex brain structures, and that more complex animals such as annelids evolved from such a simply organized ancestor. This conclusion contradicts alternative evolutionary scenarios proposing an annelid-like ancestor of Bilateria and Spiralia and several independent events of secondary reduction.

Introduction

Molecular data have profoundly changed the view of the bilaterian tree of life by recognizing three major clades: Deuterostomia, Ecdysozoa and Spiralia (Halanych 2004; Edgecombe et al. 2011). The term Spiralia is occasionally used as a synyonym for Lophotrochozoa (Halanych 2004). However, the term Lophotrochozoa is actually reserved for all descendants of the last common ancestor of Annelida, Mollusca and the three lophophorate taxa (Halanych 2004), while the more comprehensive taxon Spiralia includes all animals with spiral cleaveage and, hence, also Platyhelminthes (Edgecombe et al. 2011). Herein we use Spiralia in the terms of the more inclusive definition.

Previous results of the molecular phylogenetic analyses initiated a still on-going debate about the evolution of complexity in Bilateria. It was proposed that the last common ancestor of Deuterostomia, Ecdysozoa and Spiralia had a segmented and coelomate body organization resembling that of an annelid, and that morphologically more simply organized taxa like nematodes or flatworms (Platyhelminthes) evolved by secondary reductions (Brinkman and Philippe 2008; De Robertis 2008; Couso 2009; Tomer et al. 2010; Chesebro et al. 2013). This is in stark contrast to the traditional "acoeloid-planuloid" hypothesis favouring evolution of Bilateria from a simple body organization towards more complex forms with a last common ancestor resembling a flatworm without segmentation and coelomic cavities (Hyman 1951; Halanych 2004; Hejnol et al. 2009). Unravelling the phylogenetic relationships within Bilateria is crucial to resolve this controversy (Halanych 2004; Edgecombe et al. 2011).

While recent phylogenomic studies recovered most of the relations of the major branches within Deuterostomia and Ecdysozoa, the internal phylogeny of Spiralia is still unclear (Edgecombe et al. 2011). Indeed, spiralian animals exhibit a wide variety and plasticity in development and morphology including body organization (Nielsen 2012) which gave rise to the distinction of two major taxa: Lophotrochozoa and Platyzoa (Halanych 2004; Edgecombe et al. 2011). As mentioned above, Lophotrochozoa comprises at least annelids (ringed worms), lophophorates and molluscs (Halanych 2004) and hence animals with a more complex morphology. In contrast, Platyzoa subsumes more simple appearing taxa such as flatworms, hairy backs (Gastrotricha), wheel animals (classical Rotifera), thorny-headed worms (Acanthocephala), and jaw worms (Gnathostomulida) (Cavalier-Smith 1998). While some authors regard Platyzoa as sister to Lophotrochozoa (Edgecombe et al. 2011), others place Platyzoa within Lophotrochozoa, thus rendering Spiralia synonymous with Lophotrochozoa (Halanych 2004). Importantly, unique morphological autapomorphies supporting the monophyly of Platyzoa are lacking (Giribet 2008) and phylogenetic analyses of nuclear and mitochondrial data failed to resolve the question as well (Paps et al. 2009a; Paps et al. 2009b; Bernt et al. 2013). Nevertheless, there seems to be a tendency for a weakly supported monophylum Platyzoa as long as larger datasets were analysed (Halanych 2004; Hausdorf et al. 2007; Struck and Fisse 2008; Hejnol et al. 2009; Paps et al. 2009a; Witek et al. 2009). However, across all these analyses placement of platyzoan taxa appeared unstable, probably due to low data and taxa coverage (Edgecombe et al. 2011). Moreover, parallel evolution of character states on long branches (also known as long branch attraction, LBA) might also have confounded these analyses (Edgecombe et al. 2011). In summary, monophyly of Platyzoa and the phylogenetic positions of the platyzoan taxa within Spiralia are still contentious although their positions have major implications for bilaterian evolution. In particular, monophyly of Platyzoa and a placement within Lophotrochozoa would be in line with the theory of a more complex ancestry (Brinkman and Philippe 2008), whereas paraphyletic Platyzoa with respect to Lophotrochozoa would support the "acoeloid-planuloid" hypothesis.

Results and Discussion

To address the major outstanding issues of bilaterian phylogeny with respect to spiralian and more specifically platyzoan relationships, we applied a phylogenomic approach, generating transcriptome sequence data for 10 putative platyzoan and two nemertean species using second-generation sequencing technology and a modified RNA amplification method, which allowed the generation of sequencing libraries from as few as 10 specimens of microscopic species of Gnathostomulida, Gastrotricha and classical Rotifera (Supplementary Table S1). These data were complemented with transcriptomic or genomic data of 53 other spiralian and ecdysozoan species, including additional representatives of Platyzoa (Supplementary Table S2). Hereby, the taxon coverage of Platyzoa increased 3.5-fold and for individual platyzoan taxa like Syndermata (wheel animals and thorny-headed worms) and Gastrotricha even 5-fold in comparison to previous large-scale analyses of spiralian relationships (Dunn et al. 2008; Hejnol et al. 2009). After orthology assignment (Ebersberger et al. 2009) the data were further screened for sequence redundancy (Kvist and Siddall 2013), potentially paralogous sequences (Struck 2013a) and contamination (Struck 2013a) resulting in a pruning of about 7% of sequence data (Supplementary Tables S3-S8).

Brute-force approach: more taxa and data.
Phylogenetic reconstructions based on the largest datasets d01 with 82,162 amino acid positions and 38.3% sequence coverage (Supplementary Table S9) recovered monophyly of both Platyzoa and Lophotrochozoa with strong bootstrap support (BS) of 99 for both (Fig. 1). Within Platyzoa, monophyly of Platyhelminthes, of Syndermata as well as of Gnathostomulida was maximally supported, whereas monophyly of Gastrotricha was not recovered. The chaetonotid gastrotrich *Lepidodermella squamata* appeared as sister to Platyhelminthes (BS 46), whereas the macrodasyidan gastrotrichs formed a monophylum (BS 74) as sister to all other platyzoan taxa (BS 68). Finally, Gnathostomulida was sister to Syndermata (BS 61) consistent with the Gnathifera hypothesis (Ahlrichs 1997; Herlyn and Ehlers 1997).
To study the influence of unstable taxa, leaf stability analyses were performed. With a leaf stability index of 0.876 the gastrotrich *L. squamata* was the most unstable species within the sampled platyzoans, followed by the two gnathostomulid species (0.941) and the macrodasyidan gastrotrich *Dactylopodola baltica* (0.969) (Fig. 2, Supplementary Table S10). Excluding these four platyzoan taxa from dataset d01 and conducting a new phylogenetic reconstruction did not influence the remaining topology, but led to an increased BS value of 99 for a clade uniting Platyhelminthes and Syndermata and decreased values for the monophyly of both Platyzoa and Lophotrochozoa (BS 82 & 86, Table 1). Thus, the four unstable taxa showed some influence on support for the phylogenetic placement of other platyzoan taxa. Therefore, we excluded these four taxa from the following analyses, which addressed the potential role of LBA on platyzoan phylogeny in more detail.

LBA accounts for monophyly of Platyzoa.
Monophyletic Platyzoa as sister to Lophotrochozoa gained strong support in the analyses described above. However, thorough inspection of the topology (Fig. 1) revealed considerable branch length heterogeneity, with long branches in the analysed platyzoan lineages and rather short branches in lophotrochozoan and ecdysozoan lineages. Hence, the observed strong support for monophyletic Platyzoa might originate from artificial rather than phylogenetic signal (Bergsten 2005; Edgecombe et al. 2011; Kuck et al. 2012). For the tree derived from dataset d01 and shown in Fig. 1 the LB scores showed a bimodal distribution with a minimum between the two highest optima at a LB score value of 0 (Fig. 2B). Putative platyzoan species had generally higher LB score values than lophotrochozoan and ecdysozoan species (Fig. 2, Supplementary Table S11). Only the LB scores inferred for *Stylochoplana* and *Paraplanocera*

within Platyhelminthes, the two *Brachionus* species and *Lecane* in Syndermata, and *Megadasys* and *Macrodasys* in Gastrotricha approximated those of most lophotrochozoans and ecdysozoans. On the other hand, *Symbion* (Cycliophora), *Alcyonidium* and *Tubulipora* (Ectoprocta) showed values > 0, resembling those of most of the platyzoan species sampled (Fig. 2).

To assess the effect of long branches on tree reconstruction, all species with LB scores above 0 were excluded from dataset d01 (82,162 positions). Interestingly, monophyly of Platyzoa was no longer recovered (Fig. 3). Gastrotricha now emerged as sister to Platyhelminthes (BS 96, Table 1), and this clade was sister to monophyletic Lophotrochozoa (BS 95, Table 1), while Syndermata was sister to all other spiralian taxa. Thus, exclusion of long-branched species had a tremendous effect on the analyses rendering a strongly supported monophyly of Platyzoa with BS values above 95 into a paraphyletic assemblage, in which the clade consisting of Gastrotricha+Platyhelminthes and Lophotrochozoa obtained strong support with a BS value of 95.

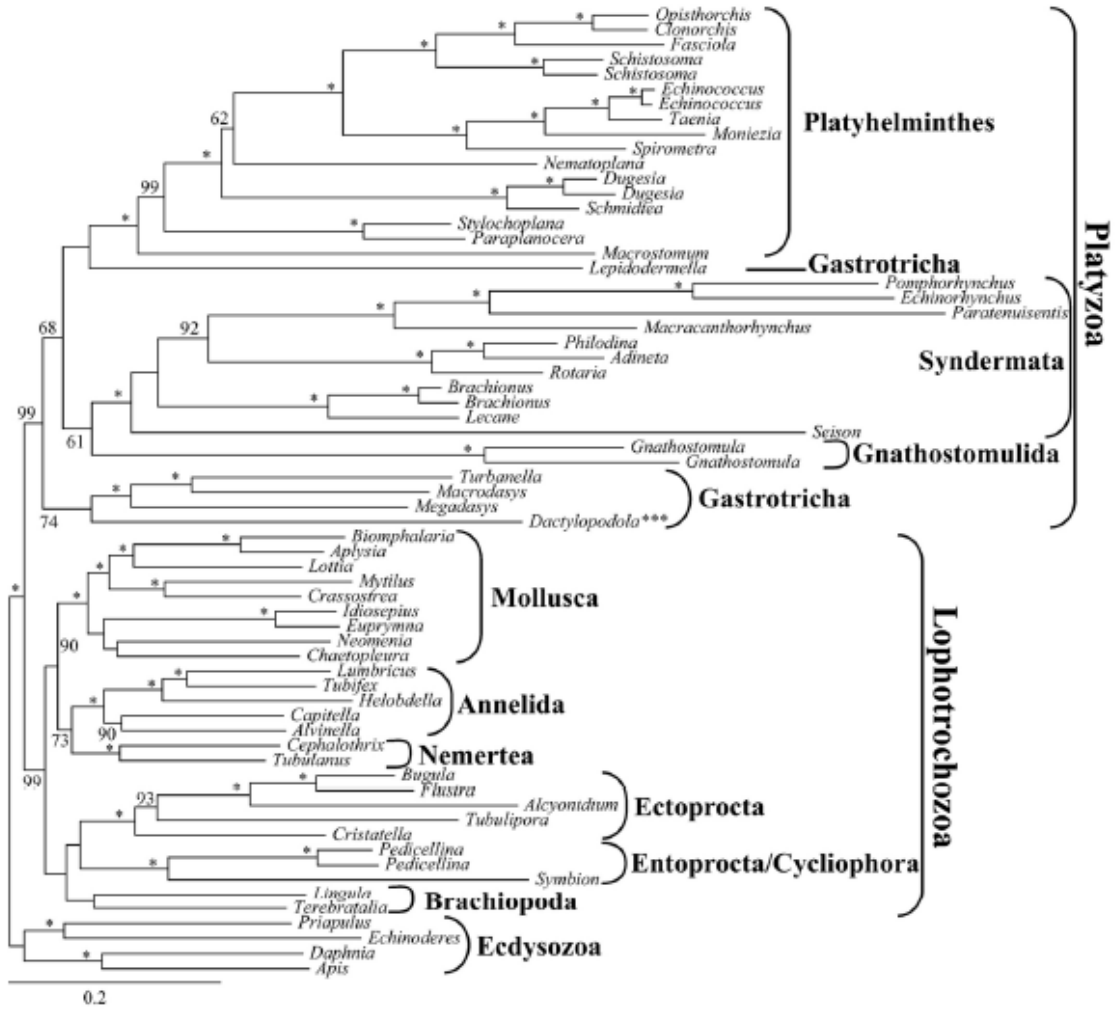


Fig. 1: Maximum Likelihood (ML) tree obtained by analysis of dataset d01 with 65 taxa and 82,162 amino acid positions. Only bootstrap support (BS) values ≥ 50 are shown at the branches, * indicates maximal support of 100. Higher taxonomic units are indicated.

Biases causing monophyletic Platyzoa.

To gain further insights into the issue of mono- vs. paraphyletic Platyzoa we analyzed the data with respect to the different properties of individual genes. In detail, we studied the effect of

gene-specific proportions of hydrophobic amino acids and missing data, base composition and branch length heterogeneity, and evolutionary rates on tree reconstruction. A common procedure is to choose one of these properties as the most influential one either based *a priori* on literature or *a posteriori* on the obtained results (e.g., Brinkman and Philippe 2008; Simmons 2012b; Nesnidal et al. 2013; Nosenko et al. 2013; Roure et al. 2013; Salichos and Rokas 2013). Herein we used another procedure based on the variability exhibited in the data itself prior to analyses of alternative datasets reflecting different degrees of data reduction. According to the principal component analysis (Alexe et al. 2008) the first principal component explained 31.0% of the variance between the different genes. It was mainly derived from the proportion of missing data and base composition heterogeneity with eigenvectors pointing into opposite directions (Supplementary Fig. S3 & Table S12). Branch length heterogeneity and evolutionary rate were the largest factors in the second component, which explained 26.8% of the variance. Correlation analyses showed that in our case evolutionary rate, which is often used as a proxy for branch length heterogeneity (Brinkman and Philippe 2008), did not correlate with actual measurements of branch length heterogeneity (R2 = 0.0324 and 0.0635, Supplementary Fig. S4).

As we wanted to test for LBA we used the direct measurement of branch length heterogeneity instead of evolutionary rate. Thus, we generated datasets with either different degrees of missing data (d02-d06), proportion of low base composition heterogeneity (d07) or low branch length heterogeneity (d08) as well as genes being part of the 70% or 95% confidence intervals of the first two principal components (d09 & d10) (Supplementary Tables S9 & S13). Based on the results of the principal component analysis we present in detail the results of three datasets d07 (low base composition heterogeneity), d08 (low branch length heterogeneity) and d02. The latter combines a low degree of missing data with a high number of positions.

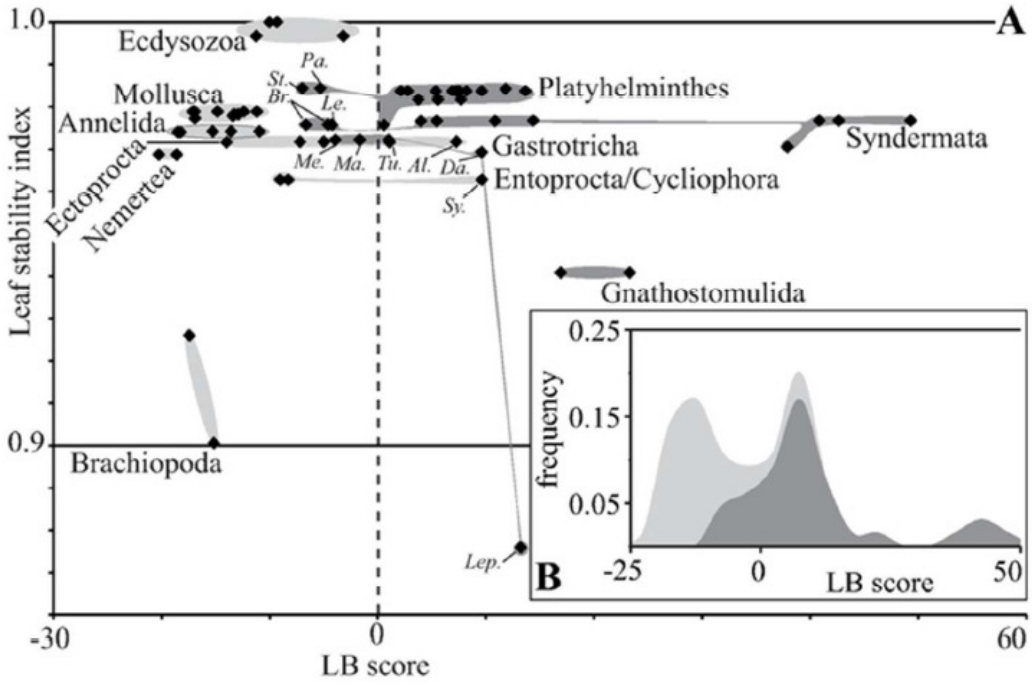


Fig. 2: Leaf stability indices and LB scores based on the ML analysis of dataset d01 with 65 taxa and 82,162 amino acid positions. (A) Plot of leaf stability indices against LB scores. (B) Distribution of LB scores. The dashed line in A indicates LB score = 0. Shades distinguish ecdysozoan and lophotrochozoan species (light grey) from putative platyzoan species (dark grey). Some species are also labelled: Pa= Paraplanocera, St= Stylochoplana, Br= Brachionus, Le= Lecane, Me= Megadasys, Ma= Macrodasys, Da= Dactylopodola, Lep= Lepidodermella, Sy= Symbion, Tu= Tubulipora, Al= Alcyonidium.

Analyses of these three datasets excluding the four above-mentioned unstable platyzoan taxa consistently resulted in paraphyletic Platyzoa (Fig. 4, Table 1) as observed before when excluding long branched taxa from the large dataset d01. Once more, Platyhelminthes was sister to Gastrotricha (BS 76, 84 & 71, Rouphozoa in Table 1) and Lophotrochozoa was recovered as a monophyletic group (BS 98, 47 & 95). The clade of Gastrotricha/Platyhelminthes was sister to Lophtrochozoa (BS 72, 86 & 75, Table 1) and Syndermata was sister to the all other spiralian taxa again. Thus, either by increasing the coverage (d02) or decreasing base composition or branch length heterogeneity (d07 & d08) paraphyletic Platyzoa was recovered (Table 1), as a clade comprising Gastrotricha, Platyhelminthes and Lophotrochozoa gained strong branch support exceeding values of 70.

Table 1: Bootstrap support for monophyly and paraphyly of Platyzoa as well as monophyly of Rouphozoa.

Dataset	Excl. taxa[a]	#pos.[b]	#taxa[c]	Platyzoa mono.	Platyzoa para.	Rouphozoa mono.
d01	None	82,162	65	99**	0	3
(all data)	unstable	82,162	61	82*	1	1
	LB[d]	82,162	34	3	95**	96**
d02	unstable	36,513	61	3	86*	84*
(high coverage)	LB[d]	36,513	34	0	100**	93*
d07	unstable	37,907	61	19	75*	71*
(low base frequency heterogenety)	LB[d]	37,907	34	0	100**	97**
d08	unstable	29,133	61	18	72*	76*
(low branch length heterogenety)	LB[d]	29,133	34	10	50	54

a Excl. = excluded
b # pos. = number of positions
c # taxa = number of taxa
d LB = long branched taxa
* Support values are part of the 70% confidence set
** Support values are part of the 95% confidence set

Additional exclusion of long-branched species (Figs. 2 & 3) reproduced paraphyly of Platyzoa in all analyses, even with maximum bootstrap support in some analyses. Again Platyhelminthes was sister to Gastrotricha (BS 93, 97 & 54, Fig. 5, Rouphozoa in Table 1) and both were more closely related to the lophotrochozoan taxa than to Syndermata (BS 100, 100 & 50, Fig. 5, Rouphozoa in Table 1). Moreover, comparing the trees without longbranched species (Figs. 3 & 5) to the one with all species (Fig. 1) shows that now similar branch lengths lead to the "platyzoan" and lophotrochozoan species (Figs. 3 & 5). Additionally, the standard deviation of the species-specific LB scores for the trees shown in Figs. 3 & 5 are 10.3 and 9.3, respectively, and, hence, lower than the standard deviation of 15.4 for the tree of Fig. 1. This means that the latter exhibits much stronger branch length heterogeneity across all taxa than the former two. Similarly, the standard deviations for the classical tip-to-root distances are lower for the trees of Figs. 3 & 5 with 0.054 and 0.143 than for the tree of Fig. 1 with 0.202.

We also used a Bayesian approach with the GTR+CAT model, as this is known to be more robust towards long-branch attraction than classical ML models such as LG (Lartillot et al. 2007). Due to computational time restrictions and high memory requirements we were not able to use the large dataset d01 (82,162 positions). Instead, we chose dataset d02 (low to medium-low degree of missing data; 36,513 positions; 46.1% coverage; Supplementary Table S9) as the principal component analysis indicated coverage as the most influential property in the first component. Importantly, the Bayesian approach did not recover monophyletic Platyzoa, but instead a clade including Gastrotricha+Platyhelminthes and monophyletic Lophotrochozoa

(posterior probability (PP) = 1.00, Fig. 6) and again Gnathostomulida+Syndermata was sister to this clade (PP=1.00, Fig.6).
Thus, combining Bayesian and Maximum Likelihood analyses with different data and taxa exclusion strategies could not recover monophyletic Platyzoa in contrast to analyses using only large numbers of data (Figs. 1, 3-6, Table 1). Considering all 10 datasets (i.e., d01 to d10) bootstrap support for monophyletic Platyzoa substantially increased with additional amino acid positions (dark grey line in Fig. 7A), while support for paraphyly decreased (black line in Fig. 7A). In contrast, support for monophyly of Lophotrochozoa was not strongly affected by the number of positions analysed (light grey line in Fig. 7A). It is a well-known phenomenon of LBA that it is positively misleading; that is, with increasing numbers of positions the artificial group is more robustly recovered (Felsenstein 1978; Huelsenbeck 1997; Bergsten 2005). On the other hand, excluding long-branched species from analyses did not lead to such correlations. In particular, support for the monophyly of Platyzoa remained low irrespective of the number of alignment positions (dark grey line in Fig. 7B).

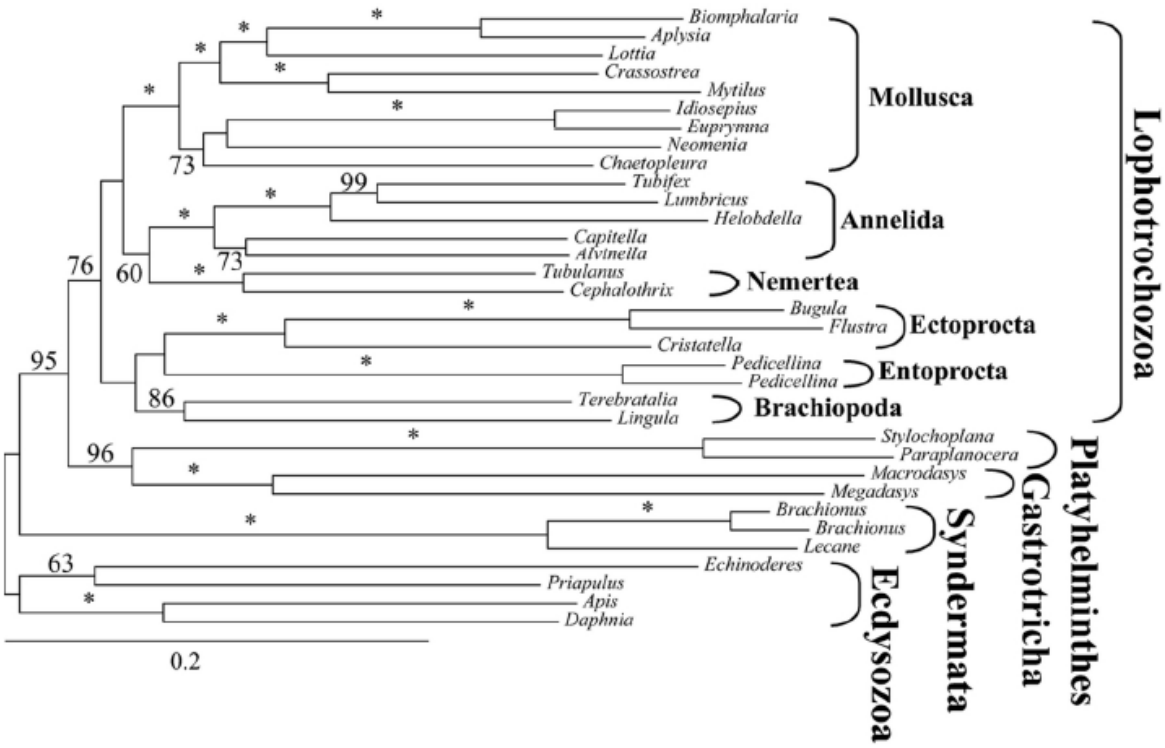


Fig. 3: ML tree obtained by analysis of dataset d01 with 34 taxa and 82,162 amino acid positions. All taxa exceeding LB scores > 0 in tree of Fig. 1 were excluded. Only BS ≥ 50 are shown at the branches, * indicates maximal support of 100. Higher taxonomic units are indicated.

Additionally, we determined for each dataset the number of single-gene trees supporting monophyly or paraphyly of Platyzoa. Across all datasets the percentage of single genes supporting platyzoan paraphyly ranged from 8.6% to 11.7% and, thus, was higher than the percentage supporting monophyletic Platyzoa, ranging from 0.5% to 3.2% (Table 2). Interestingly, decreasing the degree of missing data (i.e., d01-d06) and, hence, increasing the number of taxa per gene, the ratio of the percentage of trees supporting paraphyly relative to the percentage of trees supporting monophyly strongly increased (black line in Fig. 7C). Directly addressing biases in the data such as base or branch length heterogeneity did not have

such an effect on the ratio. In the case of LBA only strategies as used herein, which are able to attenuate its misleading effect by excluding either biased data or species or by increasing taxon coverage per gene, can reveal whether or not an assembly of long branched taxa is artificially grouped together (Bergsten 2005). In conclusion, our analyses support platyzoan paraphyly, while recovery of monophyletic "Platyzoa" is most probably due to LBA.

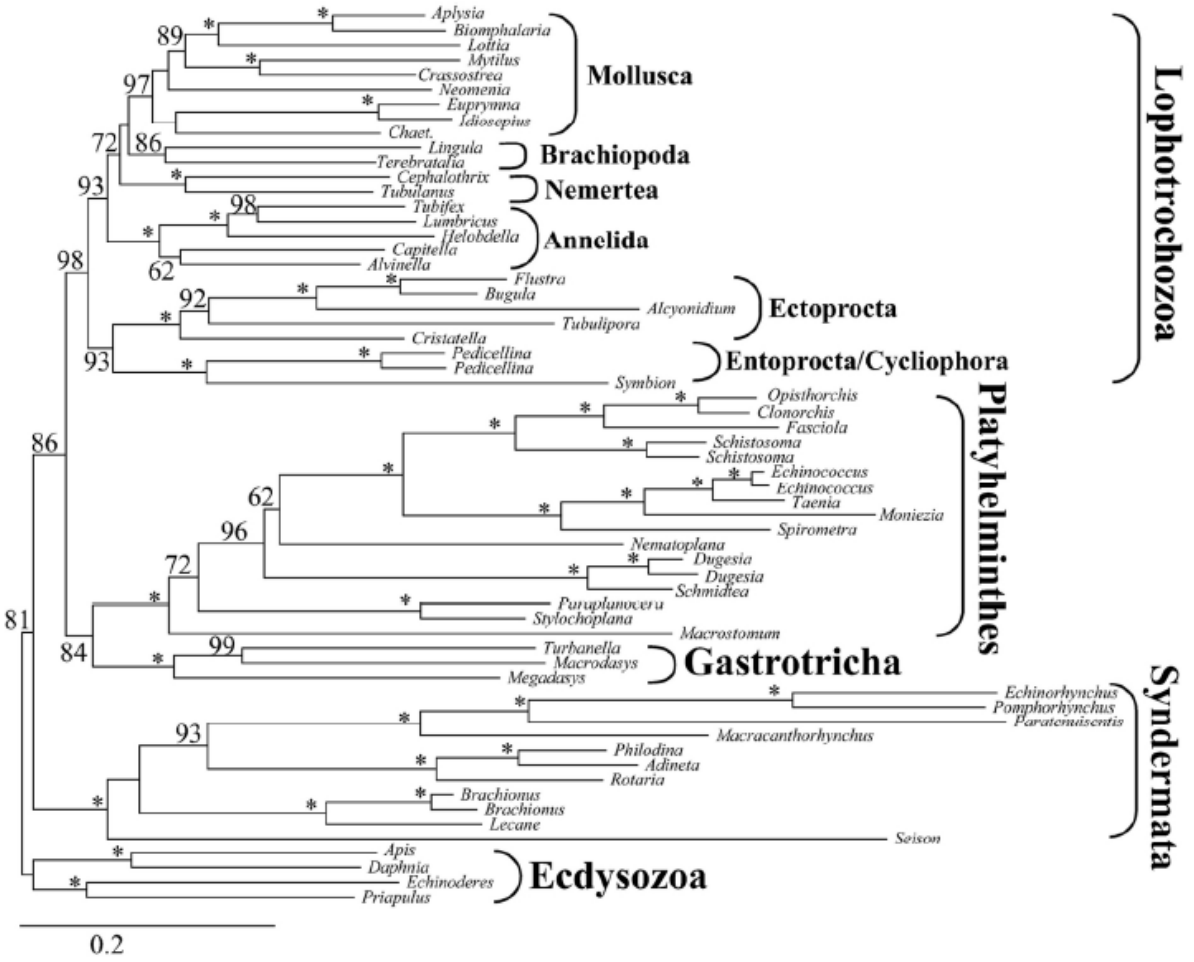


Fig. 4: ML tree obtained by analysis of dataset d02 with 61 taxa and 36,513 amino acid positions. Only partitions with low to medium up to low degrees of missing data were included and the four unstable taxa (L. squamata, D. baltica and the two Gnathostomulida species) were excluded. Only BS ≥ 50 are shown at the branches, * indicates maximal support of 100. Higher taxonomic units are indicated.

Position of Gnathostomulida.

In addition to LBA, the inference of a stable topology was hampered by the inclusion of Gnathostomulida and the two gastrotrichs *Lepidodermella* and *Dactylopodola*. In order to elucidate the phylogenetic position of Gnathostomulida within Spiralia we re-included the two formerly excluded gnathostomulid species into different datasets. Importantly, their inclusion did not alter the topology with respect to platyzoan paraphyly in any tree reconstruction (e.g., compare Figs. 4 & 8). Analysis of dataset d07 excluding unstable taxa except Gnathostomulida (i.e., *Lepidodermella* and *Dactylopodola*) and of dataset d02 excluding all long-branched taxa, recovered Gnathostomulida as part of a clade with Gastrotricha and Platyhelminthes (Table 3). However, all other analyses placed Gnathostomulida as sister to Syndermata with BS values of up to 91, even though overall bootstrap support remained low (Fig. 8, Table 3). Moreover, the Bayesian analysis also recovered a sistergroup-relationship of Gnathostomulida and

Syndermata with strong support (PP=0.98, Fig. 6). This position of Gnathostomulida as sister to Syndermata is consistent with the Gnathifera hypothesis (Ahlrichs 1997; Herlyn and Ehlers 1997). Monophyly of Gnathifera has also been found in previous studies based on ribosomal protein data (Witek et al. 2009; Hausdorf et al. 2010) and is also strongly supported by the likely homology of gnathostomulidan jaws and rotiferan trophi (Rieger and Tyler 1995; Haszprunar 1996; Ahlrichs 1997; Herlyn and Ehlers 1997; Jenner 2004a). For a thorough analysis of the phylogenetic relations within Syndermata and the implication for their evolution we refer to a recent transcriptome-based study (Wey-Fabrizius et al. 2014).

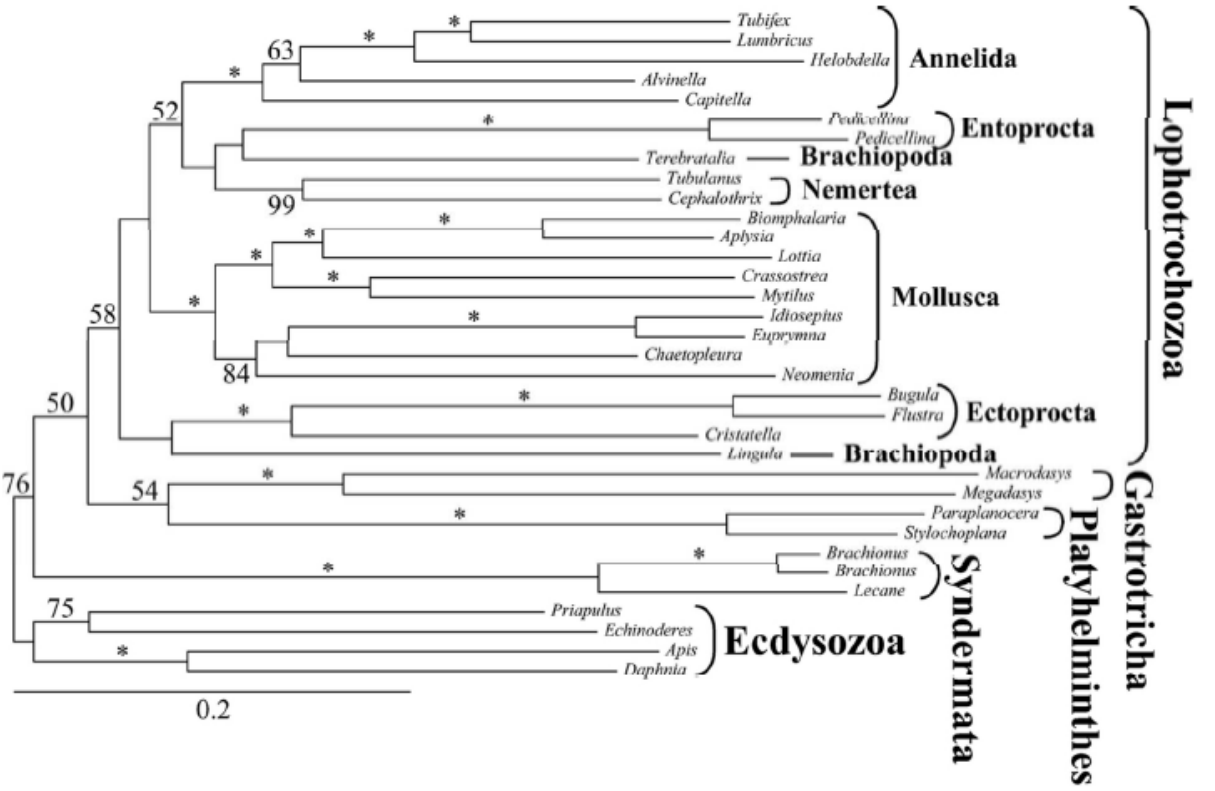


Fig. 5: ML tree obtained by analysis of dataset d08 with 34 taxa and 29,133 amino acid positions. Only partitions with low degrees of branch length heterogeneity were included and all taxa exceeding LB scores > 0 in tree of Fig. 1 were excluded. Only BS ≥ 50 are shown at the branches, * indicates maximal support of 100. Higher taxonomic units are indicated.

A novel view on spiralian phylogeny.

In summary, our analyses support the monophyly of Lophotrochozoa and of a clade combining Gastrotricha and Platyhelminthes. Gnathifera is sister to a clade comprising the aforementioned taxa (Fig. 9). No morphological apomorphy is known to date supporting either a monophyletic origin of Platyhelminthes and Gastrotricha or of Platyhelminthes, Gastrotricha and Lophotrochozoa (Jenner 2004a; Rothe and Schmidt-Rhaesa 2009) and, hence, could be used for naming these two clades. However, whereas most of the other spiralian taxa exhibit additional structures for food gathering in their ground pattern (e.g. palps in annelids, proboscis in nemerteans, filter feeding apparatuses in lophophorates, entoprocts, and cycliophorans, as well as jaw-like elements in rotifers, gnathostomulids and molluscs), gastrotrichs and most flatworm species ingest food without such extra-structures, just by dilating their rather simple pharynx. The respective pharynx simplex is part of the ground pattern of Platyhelminthes and enables the swallowing of prey by either sucking action or engulfment (Doe 1981). Gastrotricha

possess a Y-shaped or inverted Y-shaped sucking pharynx (Kieneke et al. 2008). Although gathering food by sucking is not necessarily an autapomorphy of these two taxa, this common characteristic can nonetheless be utilised for naming the clade. We therefore suggest the name Rouphozoa (derived from the Greek word *rouphao* for ingesting by sucking) to define the last common ancestor of Platyhelminthes and Gastrotricha and all its descendants. The clade of Rouphozoa + Lophotrochzoa can be named Platytrochozoa, reflecting that it comprises Platyhelminthes and taxa with a trochophore larva and all extant descendants of the last common ancestor of Platyhelminthes and Lophotrochozoa. Spiralia then comprises Gnathifera (Syndermata+Gnathostomulida) and Platytrochozoa.

Implications for bilaterian evolution.

The paraphyly of Platyzoa with respect to Lophotrochozoa is more in line with the traditional "acoeloid-planuloid" hypothesis than with the scenario of a last common ancestor of Deuterostomia, Ecdysozoa and Spiralia with a segmented and coelomate body organization resembling an annelid. Within Spiralia the non-coelomate, small-sized taxa successively branch off first (Fig. 9). Both Gnathostomulida and Gastrotricha comprise small interstitial organisms with an acoelomate body organization and less than 4 or 2 mm of length, respectively (Nielsen 2012). Within Syndermata only the highly modified, parasitic Acanthocephala are larger than a few millimeters and all exhibit a pseudocoelomate organization (Herlyn and Rohrig 2003; Nielsen 2012). Similarly, in Platyhelminthes the ancestral condition is also a small-sized, acoelomate organization as seen today in Catenulida and Macrostomorpha, which are less than 5 mm in length (Nielsen 2012). Within Spiralia, animals with a coelomate body organization are, according to our analyses, only found in Lophotrochozoa (Fig. 9). Thus, it is epistemologically more parsimonious to assume that the last common ancestor of Spiralia was an animal lacking a coelomic body cavity. Although many relationships within Lophotrochozoa are still unresolved in our study and warrant further investigations, our analyses suggest that within Spiralia coelomic cavities with a lining epithelium might have originated at the earliest in the stem lineage of Lophotrochozoa. Additionally, recent investigations of development and formation of coelomic cavities using a comparative anatomical approach revealed considerable differences between Annelida and Panarthropoda already in the earliest steps of coelomogenesis (for review see Koch et al. 2014). Hence, segmental coeloms in annelids and arthropods are not necessarily homologous structures (Koch et al. 2014). In addition, the developmental origins of coelomic cavities in deuterostomes differ from those in lophotrochozoans and panarthropods (Nielsen 2012). Considering these differences and our results, it is more probable that coelomic cavities evolved independently within the major bilaterian clades Deuterostomia, Ecdysozoa and Spiralia. Clearly, further analyses of the underlying genetic regulatory networks in coelom formation across a wide variety of coelomate and non-coelomate taxa are necessary to substantiate or reject this conclusion.

The position of coelomate Chaetognatha within Bilateria is also of interest in this aspect, but still enigmatic based on both molecular and morphological data. Deuterostome as well as protostome affinities including a sistergroup relationship to Spiralia have been proposed (Marletaz et al. 2006; Matus et al. 2006; Dunn et al. 2008; Perez et al. 2014). Moreover, Chaetognatha possess a unique type of coelom formation, heterocoely, which exhibits no strong similarities to the other types of coelom formation (Kapp 2000; Perez et al. 2014) and, hence, might be indicative of a convergent evolution of coelomic cavities in Chaetognatha. However, ultrastructural studies of coelom formation are lacking at the moment (Perez et al. 2014).

The alternative scenario whereupon evolution progressed from complex to simple in Bilateria is mainly based on similarities in segmentation in vertebrates, arthropods and annelids (De Robertis 2008; Couso 2009; Chesebro et al. 2013). However, in our analyses Annelida was always deeply nested within Lophotrochozoa. Thus, similar to the evolution of coelomic cavities, a segmented ancestry of Spiralia would imply several independent losses of this

organization, which we regard as less parsimonious. Moreover, Annelida and Arthropoda exhibit high plasticity in segmentation and, on the other hand, other spiralian and ecdysozoan taxa exhibit varying degrees of repetitive organization in organ systems. This includes Kinorhyncha, Monoplacophora and Polyplacophora, Eucestoda and other platyhelminths, some nematodes and nematomorphs, and a nemertean (Hannibal and Patel 2013; Struck 2013b). In addition, segmentation is mostly restricted to tissue derived from the ectoderm in arthropods, from the mesoderm in vertebrates, and from both germ layers in annelids (Nielsen 2012). A possible explanation for similarities in segment formation including developmental pathways like the *notch* oscillation could be that these gene regulatory networks have been co-opted from ancestral networks involved in the organization of repetitive organ systems (Davidson and Erwin 2006; Chipman 2010). However, this hypothesis cannot be conclusively proven due to a current lack of data on developmental gene pathways in taxa with such repetitive organ systems (Chesebro et al. 2013). Nonetheless, the spiralian phylogeny derived herein provides additional support for the hypothesis that segmentation evolved independently within Deuterostomia, Ecdysozoa and Spiralia.

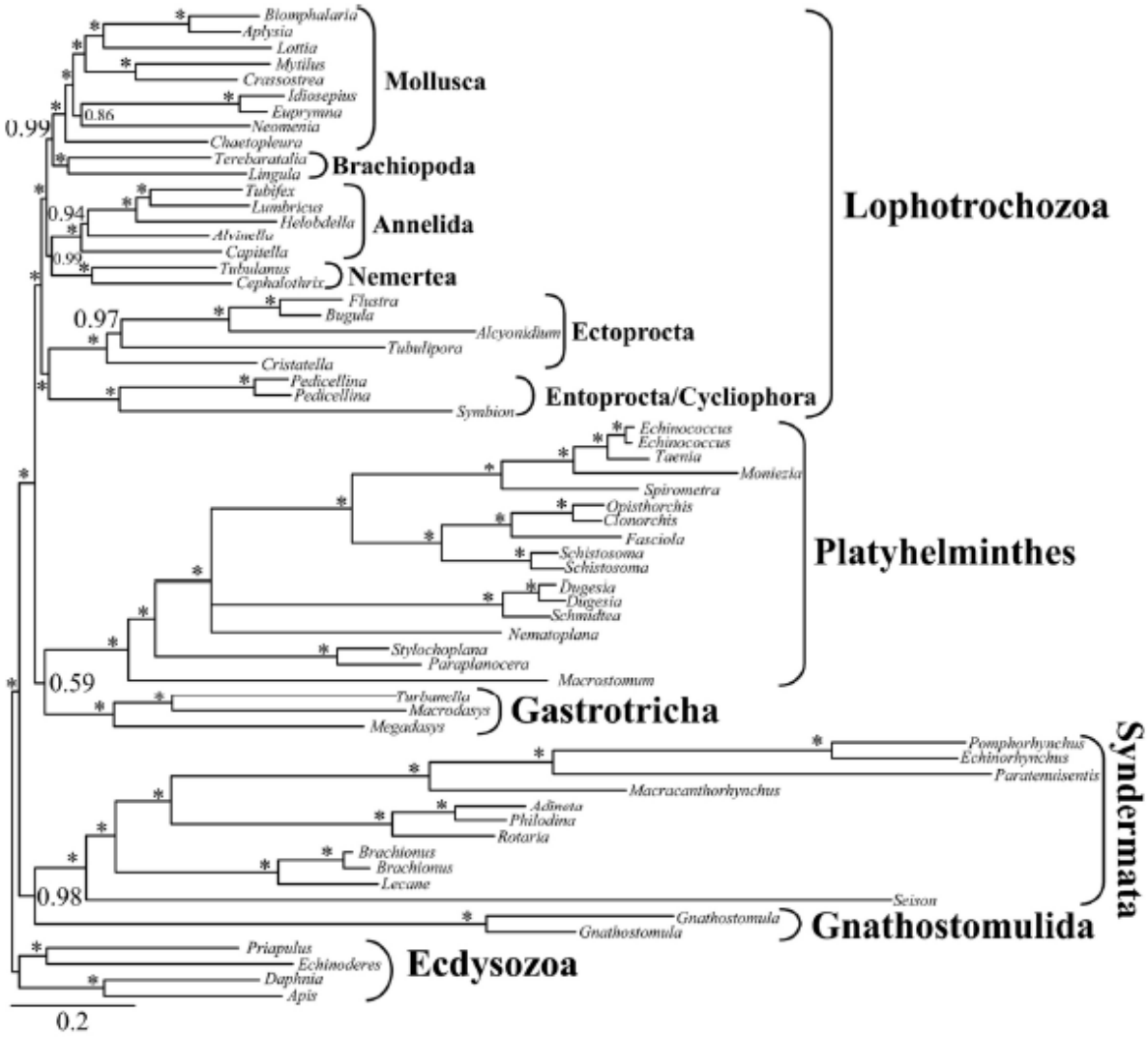


Fig. 6: Bayesian Inference (BI) tree obtained by analysis of dataset d02 with 63 taxa and 36,513 amino acid positions. Only partitions with low to medium up to low degrees of missing data were included and only the two unstable gastrotrich taxa L. squamata and D. baltica were excluded. Only posterior probabilities (PP) ≥ 0.50 are shown at the branches, * indicates maximal support of 1.00. Higher taxonomic units are indicated.

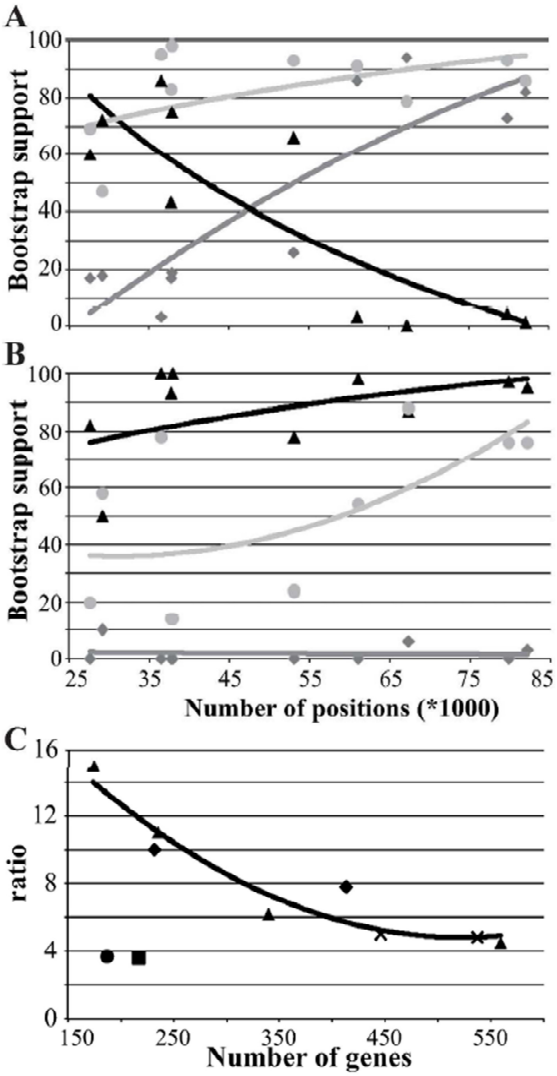


Fig. 7: Bootstrap support and the ratio of single-genes supporting paraphyly over monophyly relative to the number of alignment positions or genes. (A, B) Bootstrap support for monophyly and paraphyly of Platyzoa as well as monophyly of Lophotrochozoa relative to the number of positions. (A) Analyses based on 61 taxa, from which the four unstable taxa (L. squamata, D. baltica and the two Gnathostomulida species) were excluded. (B) Analyses based on 34 taxa, from which all taxa exceeding LB scores > 0 in tree of Fig. 1 were excluded. Light grey = monophyly of Lophotrochotrochzoa, dark grey = monophyly of Platyzoa, black = paraphyly of Platyzoa. Best-fitting trend lines generated by Excel are also shown in the same colours. (C) Ratio of the percentage of single-gene trees supporting paraphyly of Platyzoa to the percentage of singlegene trees supporting monophyly of Platyzoa relative to the number of genes. Diamonds = datasets d02 & d03 with reduced missing data; triangles = datasets d01, d04-d06 generated using MARE; circle = dataset d07 with reduced base heterogeneity; square = dataset d08 with reduced branch length heterogeneity; crosses = datasets d09 & d10 based on confidence intervals of PCA. The best-fitting trend line generated by Excel for the datasets d01-d06 with decreasing degrees of missing data is shown in black.

Support for a complex bilaterian ancestor also arose from the observation of neuronal structures called mushroom bodies that were consistently present in arthropods and some annelids, as well as similar gene expression patterns noted in these bodies and in the vertebrate pallium (Heuer et al. 2010; Tomer et al. 2010). However, within annelids, mushroom bodies occur exclusively in five families of the subgroup Errantia, which are all characterized by a high vagility (Heuer et al. 2010; Struck et al. 2011), while they are not known for any other annelid or spiralian taxa (Rothe and Schmidt-Rhaesa 2009; Heuer et al. 2010; Nielsen 2012; Loesel 2014). Thus, if such

distinct higher brain centres are taken as an ancestral condition of a complex last common spiralian ancestor (Heuer et al. 2010), several losses within Spiralia, including even several ones within Annelida, have to be assumed. On the other hand, the gastrotrich nervous system consists of a brain with a solid arch-like dorsal commissure with laterally positioned cell somata and a fine ventral commissure as well as a pair of longitudinal, lateroventral nerve cords joining posteriorly (Rothe and Schmidt-Rhaesa 2009). This organization is similar to the organization of the nervous system of Acoelomorpha. Hence, in comparison to the net-like plexus without a cerebral ganglion in non-bilaterian animals, both Gastrotricha and Acoelomorpha express a certain degree of condensation at the anterior end to form a more or less condensed commissural brain, but to a lesser degree than other bilaterian taxa (Rothe and Schmidt-Rhaesa 2009). Thus, Gastrotricha might still exhibit the ancestral bilaterian condition indicative that also the last common ancestor of Spiralia showed that characteristic. Moreover, also, for example, platyhelminths, syndermatans, gnathostomulids, or entoprocts show anterior condensations of the central nervous system, but not to the same degree as in elaborate brains, which can be found in some molluscs or annelids (Northcutt 2012; Loesel 2014). Such a condensation is in general agreement with a small-sized, non-coelomate ancestor for Spiralia showing no complex body organization. On the other hand, the observed similarities in the expression profiles of mushroom bodies of arthropods and annelids as well as the vertebrate pallium support the view that the evolution of more complex brain centres occurred early on in Bilateria (Heuer et al. 2010; Tomer et al. 2010). However, all three organs are part of the olfaction system. Analyses of these expression profiles in the brains of other bilaterian taxa are lacking in the moment. Hence, instead of being indicative of elaborative morphological structures the observed similar expression profiles could be part of ancestral gene regulatory networks involved in the integration of chemosensory input in clusters of cells of more simply organized brains. However, developmental biological studies of the olfaction system of other bilaterian taxa such as Gastrotricha or Platyhelminthes are required to substantiate either hypothesis.

Table 2: Percentage of single-genes supporting monophyly or paraphyly of Platyzoa.

	Degree of missing						data heterogenety		PCA[f]	
Dataset	d01	d02	d03	d04	d05	d06	d07	d08	d09	d10
#genes[a]	559	232	413	340	235	174	217	187	446	537
%mono.[b]	2.1	0.9	1.2	1.5	0.9	0.6	3.2	3.2	2.0	2.0
%para.[c]	9.7	8.6	9.4	9.1	9.4	8.6	11.5	11.8	10.1	9.9
%lack[d]	88.2	90.5	89.3	89.4	89.8	90.8	85.3	85.0	87.9	88.1
para./mono.[e]	4.5	10	7.8	6.2	11	15	3.6	3.7	5	4.8

[a] # genes = number of genes in dataset
[b] % mono. = percentage of single-gene trees supporting monophyly of Platyzoa
[c] % para. = percentage of single-gene trees supporting paraphyly of Platyzoa
[d] % lack = percentage of single-gene trees lacking resolution regarding this question
[e] para./mono. = ratio of the percentage of single-gene trees supporting paraphyly of Platyzoa to the percentage of single-gene trees supporting monophyly of Platyzoa
[f] PCA = principal component analysis

In conclusion, paraphyly of "Platyzoa" with respect to Lophotrochozoa and the spiralian phylogeny presented herein provide support for the view that the last common ancestor of Spiralia was an organism without coelomic cavity, segmentation and elaborate brain structures,

which probably inhabited the marine interstitial realm. This implies that evolution in Bilateria progressed most likely from a simple ancestor to more complex descendants independently within the three major bilaterian clades. However, we cannot rule out that miniaturization or a progenetic origin of the discussed taxa lead to loss of their morphological complexity. Several such examples are known from annelids and arthropods as in these cases it was more parsimonious to assume secondary simplification than convergent evolution (Jenner 2004b; Bleidorn 2007). However, the above discussion also shows that besides a robust phylogeny of Spiralia and Bilateria developmental biological studies of gene regulatory networks and expression profiles beyond the few standard model organisms are necessary to understand the evolution of Spiralia.

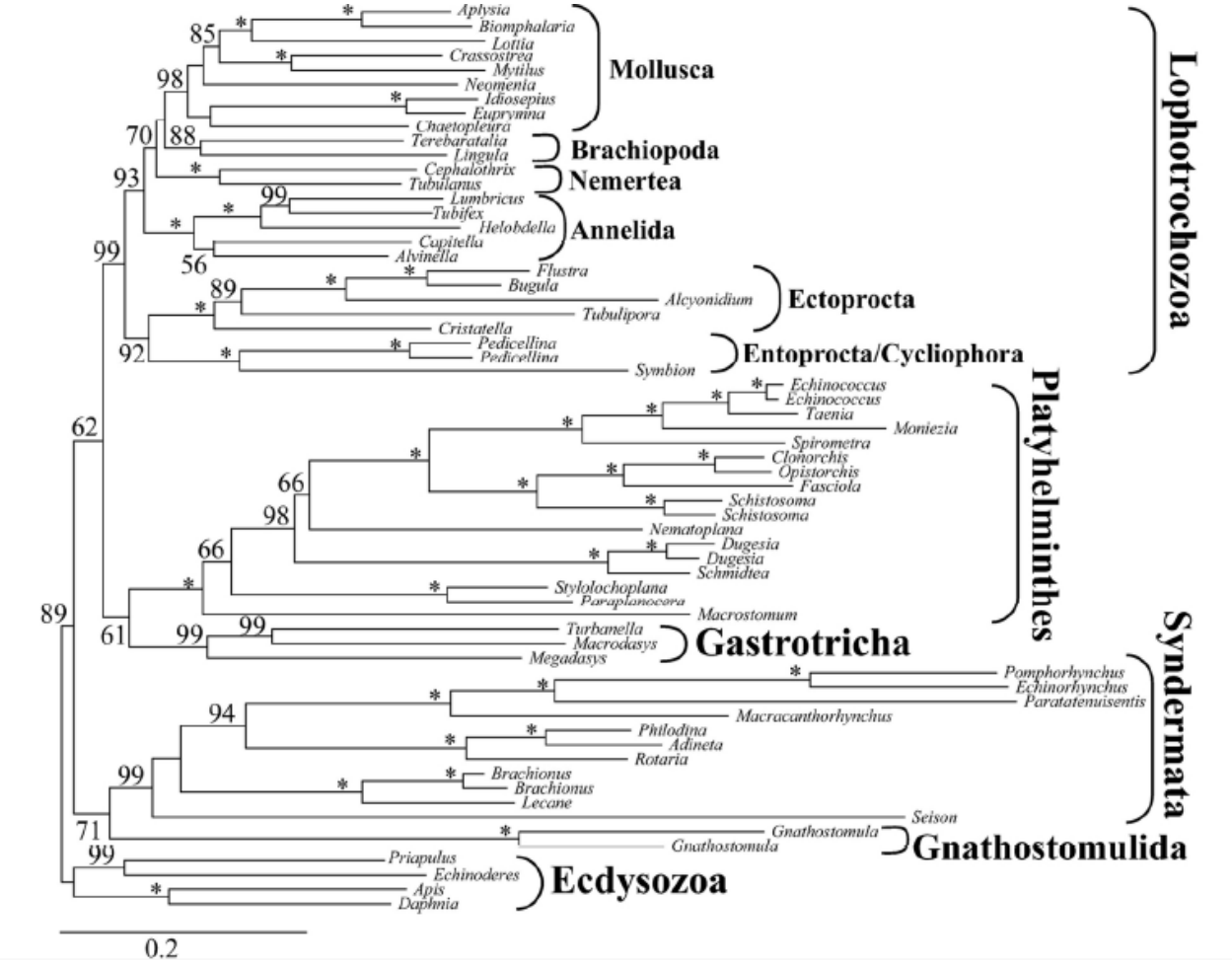


Fig. 8: ML tree obtained by analysis of dataset d02 with 63 taxa and 36,513 amino acid positions. Only partitions with low to medium up to low degrees of missing data were included and only the two unstable gastrotrich taxa L. squamata and D. baltica were excluded. Only BS ≥ 50 are shown at the branches, * indicates maximal support of 100. Higher taxonomic units are indicated.

Table 3: Bootstrap support for monophyly of Gnathifera.

Dataset	Excl. taxa[a]	#taxa[b]	Gnathifera
d01	None	65	61
(all data)	unstable	63	91*
	LB[c]	36	67
d02	unstable	63	71*
(high coverage)	LB[c]	36	10[d]
d07	unstable	63	48[d]
(low base frequency heterogeneity)	LB[c]	36	86[d]
d08	unstable	63	24
(low branch length heterogeneity)	LB[c]	36	12

[a] Excl. = excluded (same as in Table 1 except for Gnathostomulida)
[b] # taxa. = number of taxa
[c] LB = long branched taxa
[d] Gnathostomulida not placed as sister to Syndermata in the ML tree, but in a clade with Gastrotricha and Platyhelminthes
* Support values are part of the 70% confidence set

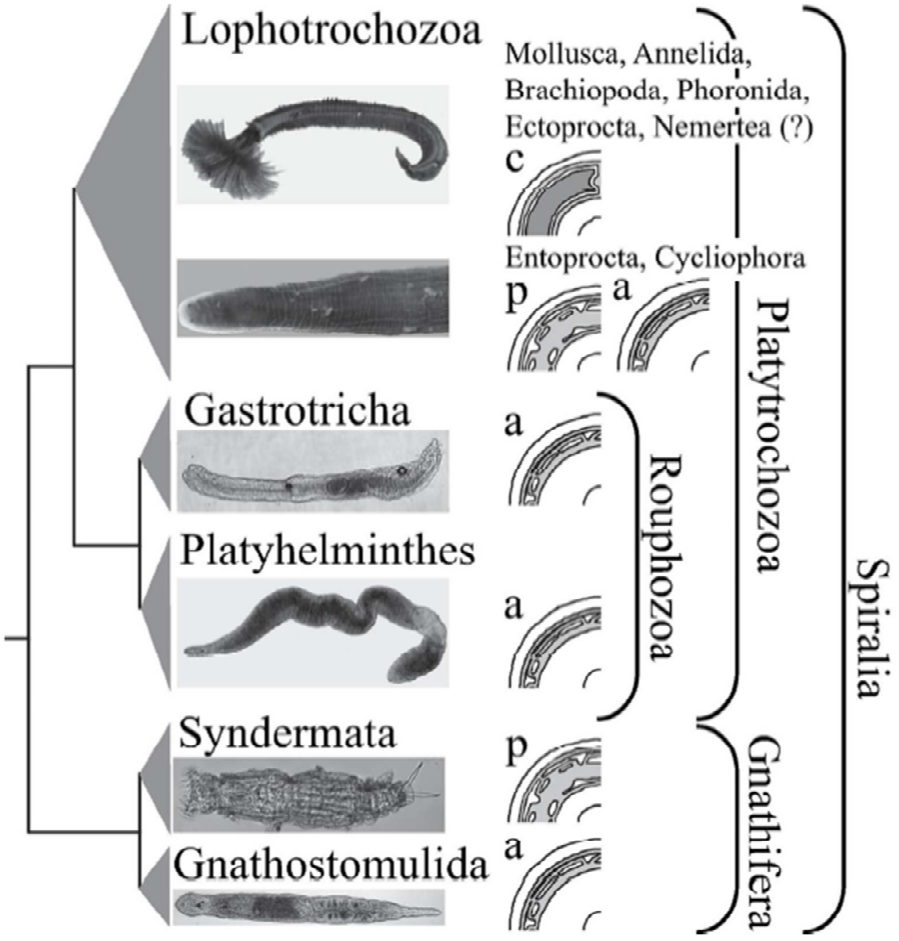


Fig. 9: Proposed phylogeny of Spiralia. Higher taxonomic units and names are given. Drawings depict the acoelomate (= a), pseudocoelomate (= p) and coelomate (= c) body organization. Picture of Rotaria neptunoida (Syndermata) was courtesy of Michael Plewka. (?) means that it is still discussed if the lateral vessels of the nemertean circulatory system are homologous to coelomic cavities of other lophotrochozoan taxa (Turbeville 1986).

Material and Methods

Data generation.
Supplementary Table S1 lists species (four gastrotrich, two flatworms, two wheel animals, one acanthocephalan, one gnathostomulid, as well as two nemertean species) collected for this study. As deeply sequenced transcriptome libraries were lacking for nemerteans we additionally constructed them for representatives of this taxon. Upon collection, samples were either snap-frozen at -80oC or stored in RNAlater. Total RNA was isolated using the NucleoSpin RNA XS Kit (Macherey-Nagel) for *Rotaria rotatoria* and *Lecane inermis* (both Syndermata, classical Rotifera), the peqGOLD MicroSpin Total RNA kit (peqlab) for *Gnathostomula paradoxa* (Gnathostomulida), *Megadasys* sp., *Macrodasys* sp., *Dactylopodola baltica* and *Lepidodermella squamata* (all Gastrotricha) or the peqGOLD Total RNA kit (peqlab) for *Tubulanus polymorphus*, *Cephalothrix linearis* (both Nemertea), *Nematoplana coelogynoporoides* and *Stylochoplana maculata* (both Platyhelminthes) and *Macracanthorhynchus hirudinaceus* (Syndermata, Acanthocephala).
For all species, except the nemerteans, total RNA was reverse-transcribed to doublestranded cDNA with the MINT UNIVERSAL cDNA synthesis kit (Evrogen) to produce amplified cDNA libraries. For *R. rotatoria*, Gnathostomulida and Gastrotricha a modified amplification protocol, which included an *in-vitro* transcription step, had been used. For this protocol the cDNA synthesis was modified to contain 1 mM T7-PlugOligo (5'-C AATT GTAA TAC GAC TCA CTA TAGG GAGAACGGGGG-3') comprising a T7 promotor sequence instead of 1 mM PlugOligo-3M in combination with CDS-3M adapter for the 1st strand synthesis and 0.1 mM T7-primer (5'- AATT GTAA TAC GAC TCA CTA TAGG -3') plus 0.1 mM M1-primer instead of 0.2 mM M1-primer for the 2nd strand synthesis. Amplified cDNA was purified using the peqGOLD Cycle-Pure Kit (peqlab), digested with *SfiI* and sizefractioned using CHROMA SPIN-1000 (Clontech). Purified cDNA was vacuum-concentrated to 15.5 Sl and 13 Sl were used for the generation of mRNA by *in vitro*-transcription (over night; 37°C) employing T7 RNA polymerase (reaction conditions: 40 Sl with 0.075 mM of each NTP, 1 u/Sl RNase inhibitor, 0.5 mM DTT and 5u/Sl T7 RNA polymerase (Invitrogen)). Messenger RNA was purified using peqGOLD Total RNA kit (peqlab).
The amplified cDNA libraries prepared from platyhelminths and *L. inermis* were sequenced by GENterprise GmbH (Mainz) or the Max Planck Institute for Molecular Genetics (Berlin) by 454 pyrosequencing using standard protocols. Illumina sequencing libraries for Nemertea, Gnathostomulida and Gastrotricha were prepared with double indices following the protocol described by Meyer and Kircher (2010) and Kircher et al (2011) starting either with totalRNA (Nemertea) or amplified mRNA (Gnathostomulida and Gastrotricha) as described by Hering et al (2012). The libraries were sequenced at the Max Planck Institute of Evolutionary Anthropology (Leipzig), using an Illumina Genome Analyzer IIx (GAIIx) with 76 cycles paired end. Total RNA of *M. hirudinaceus* and amplified mRNA of *R. rotatoria* were sequenced using an Illumina HiSeq 2000 (100 bp paired end) at the Institute of Molecular Genetics, Johannes Gutenberg University (Mainz). The sequencing library of *M. hirudinaceus* was additionally run on an Illumina MiSeq machine (150 bp paired end) by GENterprise GmbH (Mainz). Publically available transcriptomes (ESTs and RNASeq) and genomic data from 49 spiralian species complemented these data (Supplementary Table S2). For the choice of outgroup taxa different considerations have to be taken into account given that platyzoan taxa are eventually affected by long-branch attraction. First of all, the outgroup taxa should not introduce additional long branches themselves (Bergsten 2005). Hence, distantly related outgroup taxa should be avoided as well as outgroups exhibiting increased substitution rates (Milinkovitch et al. 1996; Philippe et al. 2011). Therefore, we used only representatives of Ecdysozoa, the sister group of Spiralia, and did not consider nematodes and nematomorphs, which are known to possess long branches themselves. Moreover, more than a single outgroup taxon should be used and the diversity of

outgroup taxa should be reflected (Milinkovitch et al. 1996; Bergsten 2005). Thus, we chose representative species of priapulids, kinorhynchs, and pancrustaceans as it has been previously shown that three to four outgroup taxa are sufficient to resolve difficult phylogenies when one also takes into account the computational limitations of phylogenomic studies (Rota-Stabelli and Telford 2008). Finally, the properties of the outgroup taxa sequence data should be similar to the ones of the ingroup taxa (Rota-Stabelli and Telford 2008) and in the case of long branch attraction being more similar to short branched ingroup taxa than to the long branched ones. The LB scores show that the chosen ecdsyozoan species are similar to the short-branched spiralian taxa (Fig. 2). For other properties, such as proportion of missing data and especially base composition heterogeneity, ecdysozoan taxa are similar to the ingroup taxa (Supplementary Fig. S5).

Data assembly.
Processing of *M. hirudinaceus* and *R. rotatoria* data was performed using the FastX toolkit and included trimming of (I) 12 base pairs at the 5'-end, (II) adapter sequences and (III) low quality bases (cutoff 25). Reads longer than 20 base pairs after trimming were sorted into intact pairs and singletons using a custom perl script and were subsequently assembled using the CLC Genomics Workbench 5.5 (CLC Bio).
For the GAIIx data bases were called with IBIS 1.1.2 (Kircher et al. 2009), adaptor and primer sequences removed and reads with low complexity as well as mispaired indices discarded. Raw data of all libraries were trimmed, discarding all reads with more than 5 bases below a quality score of 15. For 454 pyrosequencing data, sequences were thinned and quality filtered as implemented by Roche. In contrast to those data that were retrieved from the NCBI nr database (i.e., *Moniezia expansa*) as well as the genomic data present in the lophotrochozoan core ortholog set of HaMStR (i.e., *Schistosoma mansoni*, *Lottia gigantea*, *Helobdella robusta*, *Capitella teleta* and *Apis mellifera*) the other data were further trimmed, quality-filtered and assembled as described in either Hausdorf et al (2007) or in Riesgo et al (2012) using the CLC Genomics Workbench with 0.05 as the limit for thinning and the scaffolding option in the assembly.
Sets of orthologous genes were determined using a profile hidden Markov model-based, reciprocal hit triangulation search using a modified version of HaMStR version 8 (Ebersberger et al. 2009) (called HaMStRad and the modified files are available at https://github.com/mptrsen/HaMStRad). As a core set we used the Lophotrochozoa set of 1,253 genes derived from the Inparanoid database (http://inparanoid51.sbc.su.se) for the primer-taxa *Capitella teleta*, *Helobdella robusta*, *Lottia gigantea*, *Schistosoma mansoni*, *Daphnia pulex*, *Apis mellifera,* and *Caenorhabditis elegans*. Modifications of HaMStR included the usage of Exonerate (Slater and Birney 2005) instead of Genewise (Birney et al. 2004) to provide frame shift-corrected, corresponding nucleotide sequences. We used the representative option with all primer taxa, the relaxed option and a cutoff e value of e-05. Using the representative option might result in the assignment of the same sequence into different sets of orthologous genes. Such redundantly assigned sequences were removed using custom perl scripts, and the responsible bug in HaMStR fixed for future analyses. Each set of orthologous genes was individually aligned using MAFFT-Linsi (Katoh et al. 2005) followed by the determination of questionably aligned positions with AliScore (Kuck et al. 2010) and masking with AliCut using default parameters. The 1,253 genes were concatenated into a super-matrix using FASconCAT (Kuck and Meusemann 2010) and the super-matrix was reduced based on the phylogenetic signal in a gene by assessing the tree-likeness by quartetmapping using extended geometry mapping as implemented in MARE (Meusemann et al. 2010). We excluded the species of the core ortholog set *Schistosoma mansoni*, *Lottia gigantea*, *Helobdella robusta*, *Capitella teleta*, *Daphnia pulex*, and *Apis mellifera* prior to matrix reduction and used a d value of 0.5 generating the large dataset d01 (Supplementary Fig. S6).

Paralogy and contamination screening.

The 559 genes present in dataset d01 were further screened for paralogous sequences and contamination within single-gene datasets. For this purpose, a screening based on bootstrap Maximum Likelihood (ML) analyses of the individual genes (Philippe et al. 2011; Struck 2013a) was conducted using TreSpEx (www.annelida.de). Initially, ML analyses were conducted for the unmasked individual genes (Supplementary Fig. S6). All bipartitions supported by a bootstrap value equal to or larger than 95 were extracted from the resulting topologies. As a first step all bipartitions congruent with clades for which independent *a priori* evidence of monophyly exist were masked for the following steps (Struck 2013a). The columns "group" and "subgroup" in Supplementary Table S2 as well as genera with more than one representative indicate these *a priori* clades. To be conservative, only sequences of bipartitions that exhibited a conflict with these *a priori* clades were pruned (Supplementary Tables S4 & S5). A conflict in this case meant that species of an *a priori* clade as well as other species were present in both groups of the bipartition. For example, Platyhelminthes was such an *a priori* clade and, if in a bipartition platyhelminth as well as other spiralian and/or ecdysozoan species were present in both clades of the bipartition, this was regarded as a strong conflict. Thus, there was a strong conflict in these cases regarding the monophyly of a clade with *a priori* independent evidence of monophyly. At the group level all, but one clade fulfilled this criterion, i.e. showed strong conflicts. The single exception was a clade comprising only all gnathiferan species in that dataset eventually reflecting true phylogenetic signal. Previous studies have shown that such a pattern is characteristic for phylogenies of paralogous sequences reflecting the gene tree rather than the species tree (Rodriguez-Ezpeleta et al. 2007; Philippe et al. 2009; Philippe et al. 2011; Struck 2013a). However, other sources of artificial signal like shared missing data, compositional biases, contamination, or LBA (Bergsten 2005; Lemmon et al. 2009; Simmons and Freudenstein 2011; Simmons 2012a, b; Struck 2013a) can also result in such a pattern. In any case, potentially strong misleading signal with significant bootstrap support in single gene analyses has been masked by this procedure.

The paralogy screening was followed by a screening procedure for contamination in the libraries of our study. Therefore, the 18S rRNA sequence of *Lineus bilineatus* (DQ279932) was blasted against each assembled library (Supplementary Fig. S6) using blastn and a cutoff value of e-20. All detected contigs were then blasted against the NCBI nr database using blastn. If the best hit represented a species from a different supra-specific taxon with the traditional rank of a phylum than the query species, this was taken as an indication of possible contamination (Supplementary Table S6). For example, for some of the contigs of the *Alvinella pompejana* (Annelida) library, blast searches resulted in best hits linking the query sequence to the nematod *Tripylella* sp., the arthropod *Ptinus fur* or an uncultured acaulosporan fungus. To prune eventually contaminated sequences from the sets of 559 genes, reference databases were specifically generated for each affected species based on the blast results against the NCBI database. For the *Alvinella* example, a reference database consisted of the non-redundant proteome information retrieved from the genomes of *Apis mellifera* (Arthropoda), *Caenorhabditis elegans* (Nematoda), *Schizosaccharomyces cervesiae* (Fungi), and the transcriptome of *Daphnia pulex* (Arthropoda) as negative references as well as from the genomes of *Capitella teleta*, *Helobdella robusta* (Annelida), *Lottia gigantea* (Mollusca) and *Schmidtea mediterranea* (Platyhelminthes) as positive references. Each of the 559 genes present for that species (e.g., *A. pompejana*) was blasted against this species-specific reference database. Three pruning strategies were tested: a sequence was pruned when (I) the best hit was a negative reference sequence, (II) the best hit was a negative reference sequence and in addition the E value was at least one order of a magnitude better than that of the best hit for a positive reference, or (III) the best hit was a negative reference sequence and in addition the E value was at least four orders better than that of the best hit for a positive reference. As ML analyses of the dataset d01 with 65 taxa and 82,162 amino acid positions using the three different pruning

strategies resulted in no significant differences of the topologies inferred we chose the most conservative first pruning strategy for subsequent analyses. Custom Perl scripts were written for all these steps.

Phylogenetic analyses.
The most appropriate substitution model was LG+I+Γ as determined using the ProteinModelSelection script for RAxML (Stamatakis 2006). Before the time-consuming Bayesian Inference (BI) we conducted a series of ML analyses as part of the sensitivity analyses and screening procedures (see Supplementary Fig. S6). In total, 1,129 ML analyses were conducted with RAxML 7.3 (Stamatakis 2006) using 300 and 100 bootstrap replicate searches for concatenated and individual gene datasets, respectively. The bootstrap searches were followed by a search of the best tree. Preliminary analyses using the automatic bootstopping option (Pattengale et al. 2009) (-# autoMRE) in RAxML obtained a maximum of 240 bootstrap replicates for different tested concatenated datasets and, hence, we used 300 replicates for all analyses for reasons of comparability. Moreover, these preliminary analyses showed that a bootstrap search followed by a best tree search always found a tree with an equal or better likelihood score than independent searches for the best tree using 100 replicate searches starting from randomized maximum parsimony trees.
For the BI analysis we used PhyloBayes MPI 1.4f (Lartillot and Philippe 2004; Lartillot et al. 2013) using the GTR+CAT model and the dataset d02 generated by excluding genes with high degrees of missing data (see sensitivity analyses below). For the analysis four chains ran in parallel for 13,669 cycles on average (ranging from 12,164 to 14,217). Convergence of likelihood values, alpha parameter, and tree length of the four chains was assessed using Tracer v1.5 (http://beast.bio.ed.ac.uk/Tracer). Upon convergence the average standard deviation of split frequencies was below 0.1 with a value of 0.055. The first 6,000 cycles of each chain were discarded as burnin and the majority rule consensus tree containing the posterior probabilities was calculated from the remaining trees of the chain with the best average likelihood score sampling every 2nd tree.

Sensitivity analyses.
Leaf stability indices of species were determined using Phyutility (Smith and Dunn 2008) and the bootstrap trees of ML analyses of the dataset d01 comprising all species sampled. To assess the branch length heterogeneity we used the herein newly developed LB score (see main text) using TreSpEx (www.annelida.de), which we also used to calculate classical tip-to-root distances. Taxa were excluded from the datasets d01-d10 in accordance with these results and the phylogenetic reconstructions repeated.
To objectively assess the branch length heterogeneity in a tree we developed a new treebased measurement, which we call the LB score. The score utilizes patristic distances (PD), i.e. the distance between two taxa based on the connecting branches, and is based on the mean pairwise PD of a taxon *i* to all other taxa in the tree relative to the average pairwise PD over all taxa (a):

$$LB_i = \left(\frac{\overline{PDi}}{\overline{PDa}} - 1 \right) * 100$$

In specific, the score measures for each taxon the percentage deviation from the average and is independent of the root of the tree. The latter is also the reason for not using the traditional tip-to-root distance (Bergsten 2005). When using tip-to-root distances the recognition of long-branched taxa heavily depends on the root of the tree. For example, in the reconstruction of the indvidual gene with the ID 111427 in our analyses below the ecdysozoan outgroup species are not monophyletic. Whereas tip-to-root distances based on an *Apis*–rooted tree and LB scores

indicate the same taxa as long-branched, rooting the tree with either *Echinoderes* or *Priapulus* some of these species would be indicated as short-branched (Supplementary Fig. S1). Given the automatic process pipelines in phylogenomic analyses due to the vast amount of genes detection of long-branched taxa should be robust against changes in the root of the tree. Moreover, in the search for the best tree in phylogenetic reconstructions only unrooted trees are used and rooting is an *a posteriori* procedure. Thus, notwithstanding that outgroup species might be long-branched the artificial grouping of species due to LBA in phylogenetic reconstructions is not directly due to the root by itself (Bergsten 2005). Hence, detection of LBA should be independent of the root. Fortunately, either using the large dataset d01 in our analyses below or the 559 individual genes of this dataset LB scores and tip-to-root distances are highly and positively correlated with a R2 value of 0.91543 or an average R2 of 0.85684, respectively (Supplementary Fig. S2).

For data partitioning we analyzed the 559 genes of the dataset d01 generated with the MARE setting "all taxa included" and a d value of 0.5. We determined both alignment- and tree-based properties. Using BaCoCa (Kuck and Struck 2014), the proportion of hydrophobic and polar amino acids, the proportion of missing data as well as the compositional heterogeneity as measured by the RCFV values (Zhong et al. 2011) were determined from the pruned and masked alignments across all species in each gene (Supplementary Fig. S6). ML trees from these alignments were used to determine the evolutionary rate for each gene, calculated as the average pairwise patristic distance between two species in the tree, as well as the mean of the upper quartile of LB scores (i.e., the upper 25% of all LB scores) and the standard deviation of all LB scores as measurements of branch length heterogeneity with the aid of TreSpEx (www.annelida.de). Correlation studies of these properties were conducted in Excel (Supplementary Fig. S2 & S4). Principal component analyses were conducted in R with scaled values (Supplementary Fig. S3). The determination of the branch length heterogeneity within a gene was based on either the mean of the upper quartile of LB scores or the standard deviation of all LB scores within a gene. However, both approaches led to a strong linear correlation (R2 = 0.8363, Supplementary Fig. S4) and, thus, we used solely the standard deviation of LB scores as a measure of branch length heterogeneity in the principal component analysis. Similarly, the proportions of hydrophobic and polar amino acids were also strongly correlated (R2 = 0.6481, Supplementary Fig. S4) and, hence, we excluded the proportion of polar amino acids.

Genes with either high degrees of missing data or high base composition heterogeneity were excluded based on the results of heat map analyses in combination with hierarchical clustering without scaling the values in R (Bapteste et al. 2005; Susko et al. 2006). Four clusters of proportion of missing data were found ranging from low to high degrees of missing data (Supplementary Fig. S7). From dataset d01 with 559 genes belonging to the groups with medium-high to high degrees of missing data were excluded to generate dataset d02 characterized by only low degrees of missing data. We also generated a dataset d03, where we excluded only high degrees of missing data from dataset d01. Alternatively, the dataset d01 was condensed using MARE (Meusemann et al. 2010) with d values of 1.0, 1.5 or 2.0 instead of 0.5 used above resulting in the datasets d04, d05 and d06, respectively.

For compositional heterogeneity the heatmap revealed three clusters with low, medium and high compositional heterogeneity (Supplementary Fig. S8). Only genes, which were part of the cluster with low compositional heterogeneity, were kept for dataset d07. The tree-based property branch length heterogeneity was ranked and divided into three equal parts. To generate dataset d08 only the genes from the third with the lowest heterogeneity values were not excluded (Supplementary Fig. S6 and Table S9). Moreover, we excluded all genes from dataset d01, which were not part of the 70% or 95% confidence interval of the first two principal components (Supplementary Fig. S3) resulting in datasets d09 and d10, respectively. Finally, for each dataset we determined the number of single-gene trees, which find a monophyletic or

paraphyletic Platyzoa using custom perl scripts. For the latter at least one platyzoan taxon (i.e., Platyhelminthes, Gastrotricha, Gnathostomulida or Syndermata) had to be placed more closely to the outgroup than at least one other platyzoan taxon.

Acknowledgements and funding information

This work was supported by the priority program ‘‘Deep Metazoan Phylogeny’’ of the Deutsche Forschungsgemeinschaft (DFG-STR 683/5-2 and DFG-STR 683/8-1 to T.H.S., DFG-Ha2103/4 to T.H., DFG-HA 2763/5 to B.H., DFG-BL 787/5-1 to C.B.). T.H. wishes to thank the Center for Computational Sciences Mainz (CSM) for additional financial support. T.H.S. also acknowledges the support of the Regional Computing Centre of Cologne (RRZK) by, among others, providing access to the HPC cluster CHEOPS for the parallel PhyloBayes analyses. The authors gratefully acknowledge Edyta Fialkowska (Institute of Environmental Sciences, Jagiellonian University, Poland) for providing *Lecane* specimens, Laszlo Sugar (Faculty of Animal Science, Kaposvar University, Hungary) for collecting and providing *Macracanthorhynchus* specimens, Sanja Ramljak (Institute of Clinical Research and Development, Mainz, Germany) for helping with collecting *Seison* specimens and Michael Plewka (http://plingfactory.de/) for the picture of *Rotaria neptunoida*. We also thank Birgit Nickel and Matthias Meyer (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany) for their assistance in sequencing using Illumina GAIIx and Steffen Rapp (Institute of Molecular Genetics (Johannes Gutenberg University, Mainz) for operating the Illumina HiSeq 2000. Sequence data have been deposited in the NCBI short read archive. All other data and trees used in this study are available at DataDryad.

References

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

Alexe G, Vijaya Satya R, Seiler M, Platt D, Bhanot T, Hui S, Tanaka M, Levine AJ and Bhanot G. 2008. PCA and Clustering reveal alternate mtDNA phylogeny of N and M clades. J Mol Evol 67: 465–487.

Bapteste E, Susko E, Leigh J, MacLeod D, Charlebois RL and Doolittle WF. 2005. Do orthologous gene phylogenies really support tree-thinking? BMC Evol Biol 5: 33.

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

Bernt M, Bleidorn C, Braband A, et al. 2013. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. Mol Phylogenet Evol 69: 352-364.

Birney E, Clamp M and Durbin R. 2004. GeneWise and Genomewise. Genome Res 14: 988-995.

Bleidorn C. 2007. The role of character loss in phylogenetic reconstruction as exemplified for the Annelida. J Zool Syst Evol Res 45: 299–307.

Brinkman H and Philippe H. 2008. Animal phylogeny and large-scale sequencing: progress and pitfalls. J Syst Evol 46: 274–286.

Cavalier-Smith T. 1998. A revised six-kingdom system of life. Biol Rev 73: 203-266.

Chesebro JE, Pueyo JI and Couso JP. 2013. Interplay between a *Wnt*-dependent organiser and the *Notch* segmentation clock regulates posterior development in *Periplaneta americana*. Biol Open 2: 227 - 237.

Chipman AD. 2010. Parallel evolution of segmentation by co-option of ancestral gene regulatory networks. Bioessays 32: 60-70.

Couso JP. 2009. Segmentation, metamerism and the Cambrian explosion. Int J Dev Biol 53: 1305-1316.

Davidson EH and Erwin DH. 2006. Gene regulatory networks and the evolution of animal body plans. Science 311: 796-800.

De Robertis EM. 2008. The molecular ancestry of segmentation mechanisms. Proc Natl Acad Sci U S 105: 16411 16412.

Doe D. 1981. Comparative ultrastructure of the pharynx simplex in turbellaria. Zoomorphology 97: 133-193.

Dunn CW, Hejnol A, Matus DQ, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452: 745-750.

Ebersberger I, Strauss S and von Haeseler A. 2009. HaMStR: Profile hidden markov model based search for orthologs in ESTs. BMC Evol Biol 9: 157.

Edgecombe G, Giribet G, Dunn C, Hejnol A, Kristensen R, Neves R, Rouse G, Worsaae K and Sorensen M. 2011. Higher-level metazoan relationships: recent progress and remaining questions. Org Divers Evol 11: 151 - 172.

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

Giribet G. 2008. Assembling the lophotrochozoan (=spiralian) tree of life. Philosophical Transactions of the Royal Society B: Biological Sciences 363: 1513-1522.

Halanych KM. 2004. The new view of animal phylogeny. Annu Rev Ecol Evol Syst 35: 229-256.

Hannibal R and Patel N. 2013. What is a segment? EvoDevo 4: 35.

Haszprunar G. 1996. Plathelminthes and Plathelminthomorpha - paraphyletic taxa. J Zool Syst Evol Res 34: 41 - 48.

Hausdorf B, Helmkampf M, Meyer A, Witek A, Herlyn H, Bruchhaus I, Hankeln T, Struck TH and Lieb B. 2007. Spiralian phylogenomics supports the resurrection of Bryozoa comprising Ectoprocta and Entoprocta. Mol Biol Evol 24: 2723-2729.

Hausdorf B, Helmkampf M, Nesnidal MP and Bruchhaus I. 2010. Phylogenetic relationships within the lophophorate lineages (Ectoprocta, Brachiopoda and Phoronida). Mol Phylogenet Evol 55: 1121-1127.

Hejnol A, Obst M, Stamatakis A, et al. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc R Soc B: Biol Sci 276: 4261-4270.

Hering L, Henze MJ, Kohler M, et al. 2012. Opsins in Onychophora (velvet worms) suggest a single origin and subsequent diversification of visual pigments in arthropods. Mol Biol Evol 29: 3451-3458.

Herlyn H and Ehlers U. 1997. Ultrastructure and function of the pharynx of *Gnathostomula paradoxa* (Gnathostomulida). Zoomorphology 117: 135.

Herlyn H and Rohrig H. 2003. Ultrastructure and overall organization of ligament sac, uterine bell, uterus and vagina in *Paratenuisentis ambiguus* (Acanthocephala, Eoacanthocephala) – the character evolution within the Acanthocephala. Acta Zool 84: 239–247.

Heuer C, Muller C, Todt C and Loesel R. 2010. Comparative neuroanatomy suggests repeated reduction of neuroarchitectural complexity in Annelida. Front Zool 7: 13.

Huelsenbeck JP. 1997. Is the Felsenstein zone a fly trap? Syst Biol 46: 69-74.

Hyman LH. 1951. The Invertebrates, Vol. 2, Platyhelminthes and Rhynchocoela: The Acoelomate Bilateria. New York: McGraw-Hill.

Jenner RA. 2004a. Towards a phylogeny of the Metazoa: evaluating alternative phylogenetic positions of Platyhelminthes, Nemertea, and Gnathostomulida, with a critical reappraisal of cladistic characters. Cont Zool 73: 3-163.

Jenner RA. 2004b. When molecules and morphology clash: reconciling conflicting phylogenies of the Metazoa by considering secondary character loss. Evol Devel 6: 372-378.

Kapp H. 2000. The unique embryology of Chaetognatha. Zool Anz 239: 263-266.

Katoh K, Kuma K-i, Toh H and Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33: 511-518.

Kieneke A, Riemann O and Ahlrichs WH. 2008. Novel implications for the basal internal relationships of Gastrotricha revealed by an analysis of morphological characters. Zool Scr 37: 429-460.

Kircher M, Sawyer S and Meyer M. 2011. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res 2011: 1–8.

Kircher M, Stenzel U and Kelso J. 2009. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome biology 10: R83.

Koch M, Quast B and Bartolomaeus T. 2014. Coeloms and nephridia in annelids and arthropods. In: Wagele JW, Bartolomaeus T, editors. Deep Metazoan Phylogeny: The backbone of the Tree of Life - New insights from analyses of molecules, morphology, and theory of data analysis. Berlin: De Gruyter. p. 173-284.

Kuck P, Mayer C, Wagele J-W and Misof B. 2012. Long branch effects distort Maximum Likelihood phylogenies in simulations despite selection of the correct model. PLoS ONE 7: e36593.

Kuck P and Meusemann K. 2010. FASconCAT: Convenient handling of data matrices. Mol Phylogenet Evol 56: 1115-1118.

Kuck P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wagele JW and Misof B. 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front Zool 7: 10.

Kuck P and Struck TH. 2014. BaCoCa - A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol Phylogenet Evol 70: 94–98.

Kvist S and Siddall ME. 2013. Phylogenomics of Annelida revisited: a cladistic approach using genome-wide expressed sequence tag data mining and examining the effects of missing data. Cladistics 29: 435–448.

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

Lartillot N and Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21: 1095-1109.

Lartillot N, Rodrigue N, Stubbs D and Richer J. 2013. PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol 62: 611-615.

Lemmon AR, Brown JM, Stanger-Hall K and Lemmon EM. 2009. The effect of ambiguous data on phylogenetic estimates obtained by Maximum Likelihood and Bayesian Inference. Syst Biol 58: 130-145.

Loesel R. 2014. Brain complexity in protostomes. In: Wagele JW, Bartolomaeus T, editors. Deep Metazoan Phylogeny: The backbone of the Tree of Life - New insights from analyses of molecules, morphology, and theory of data analysis. Berlin: De Gruyter. p. 79-91.

Marletaz F, Martin E, Perez Y, et al. 2006. Chaetognath phylogenomics: a protostome with deuterostome-like development. Curr Biol 16: R577-R578.

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

Meusemann K, von Reumont BM, Simon S, et al. 2010. A phylogenomic approach to resolve the arthropod Tree of Life. Mol Biol Evol 27: 2451-2464.

Meyer M and Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor protocols 2010: pdb prot5448.

Milinkovitch MC, LeDuc RG, Adachi J, Farnir F, Georges M and Hasegawa M. 1996. Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on DNA sequence data in Cetaceans. Genetics 144: 1817-1833.

Nesnidal M, Helmkampf M, Meyer A, et al. 2013. New phylogenomic data support the monophyly of Lophophorata and an Ectoproct-Phoronid clade and indicate that Polyzoa and Kryptrochozoa are caused by systematic bias. BMC Evol Biol 13: 253.

Nielsen C. 2012. Animal Evolution - Interrelationships of the Living Phyla. New York: Oxford University Press Inc.

Northcutt RG. 2012. Evolution of centralized nervous systems: Two schools of evolutionary thought. Proceedings of the National Academy of Sciences 109: 10626-10633.

Nosenko T, Schreiber F, Adamska M, et al. 2013. Deep metazoan phylogeny: When different genes tell different stories. Mol Phylogenet Evol 67: 223–233.

Paps J, Baguna J and Riutort M. 2009a. Bilaterian Phylogeny: A Broad Sampling of 13 Nuclear Genes Provides a New Lophotrochozoa Phylogeny and Supports a Paraphyletic Basal Acoelomorpha. Mol Biol Evol 26: 2397-2406.

Paps J, Baguna J and Riutort M. 2009b. Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. Proc R Soc B: Biol Sci 276: 1245-1254.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME and Stamatakis A. 2009. How many bootstrap replicates are necessary? In: Batzoglou S, editor. RECOMB 2009, LNCS 5541. Berlin Heidelberg Springer-Verlag. p. 184–200.

Perez Y, Muller CHG and Harzsch S. 2014. The Chaetognatha: An anarchistic taxon between Protostomia and Deuterostomia. In: Wagele JW, Bartolomaeus T, editors. Deep Metazoan Phylogeny: The backbone of the Tree of Life - New insights from analyses of molecules, morphology, and theory of data analysis. Berlin: De Gruyter. p. 49-77.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Worheide G and Baurain D. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. PLoS Biology 9: 3.

Philippe H, Derelle R, Lopez P, et al. 2009. Phylogenomics revives traditional views on deep animal relationships. Curr Biol 19: 706-712.

Rieger RM and Tyler S. 1995. Sister-group relationship of Gnathostomulida and Rotifera- Acanthocephala. Invertebr Biol 114: 186-188.

Riesgo A, Andrade SC, Sharma P, Novo M, Perez-Porro A, Vahtera V, Gonzalez V, Kawauchi G and Giribet G. 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Front Zool 9: 33.

Rodriguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H and Lang BF. 2007. Toward resolving the eukaryotic tree: The phylogenetic positions of jakobids and cercozoans. Curr Biol 17: 1420 1425.

Rota-Stabelli O and Telford MJ. 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: 103-111.

Rothe B and Schmidt-Rhaesa A. 2009. Architecture of the nervous system in two *Dactylopodola* species (Gastrotricha, Macrodasyida). Zoomorphology 128: 227-246.

Roure B, Baurain D and Philippe H. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol Biol Evol 30: 197-214.

Salichos L and Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497: 327–331.
Simmons MP. 2012a. Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics 28: 208-222.
Simmons MP. 2012b. Radical instability and spurious branch support by likelihood when applied to matrices with non-random distributions of missing data. Mol Phylogenet Evol 62: 472-484.
Simmons MP and Freudenstein JV. 2011. Spurious 99% bootstrap and jackknife support for unsupported clades. Molecular Phylogenetics and Evolution 61: 177-191.
Slater G and Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6: 31.
Smith SA and Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24: 715-716.
Stamatakis A. 2006. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.
Struck TH. 2013a. The impact of paralogy on phylogenomic studies – A case study on annelid relationships. PLoS ONE 8: e62892.
Struck TH. 2013b. Phylogeny. In: Westheide W, Purschke G, editors. Handbook of Zoology Annelida. Berlin: DeGruyter.
Struck TH and Fisse F. 2008. Phylogenetic position of Nemertea derived from phylogenomic data. Mol Biol Evol 25: 728-736.
Struck TH, Paul C, Hill N, et al. 2011. Phylogenomic analyses unravel annelid evolution. Nature 471: 95–98.
Susko E, Leigh J, Doolittle WF and Bapteste E. 2006. Visualizing and Assessing Phylogenetic Congruence of Core Gene Sets: A Case Study of the {gamma}-Proteobacteria. Mol Biol Evol 23: 1019-1030.
Tomer R, Denes AS, Tessmar-Raible K and Arendt D. 2010. Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. Cell 142: 800-809.
Turbeville JM. 1986. An ultrastructural analysis of coelomogenesis in the hoplonemertine *Prosorhochmus americanus* and the polychaete *Magelona* sp. J Morphol 187: 51-60.
Wey-Fabrizius AR, Herlyn H, Rieger B, Rosenkranz D, Witek A, Welch DBM, Ebersberger I and Hankeln T. 2014. Transcriptome data reveal syndermatan relationships and suggest the evolution of endoparasitism in Acanthocephala via an epizoic stage. PLoS ONE 9: e88618.
Witek A, Herlyn H, Ebersberger I, Mark Welch DB and Hankeln T. 2009. Support for the monophyletic origin of Gnathifera from phylogenomics. Mol Phylogenet Evol 53: 1037-1041.
Zhong M, Hansen B, Nesnidal MP, Golombek A, Halanych KM and Struck TH. 2011. Detecting the symplesiomorphy trap: A multigene phylogenetic analysis for terebelliform annelids. BMC Evol Biol 11: 369

5.4. Elucidating the phylogenetic position of Gnathostomulida and first mitochondrial genomes of Gnathostomulida, Gastrotricha and Polycladida (Platyhelminthes)

Anja Golombek[1], Sarah Tobergte[1,2], Torsten H. Struck[1]

[1] Zoological Research Museum Alexander Koenig, Adenauerallee 160, D-53113 Bonn, Germany; torsten.struck.zfmk@uni-bonn.de, A.Golombek@gmx.de
[2] University of Osnabrück; FB05 Biologie/Chemie; AG Zoologie; Barbarastr. 11; D-49076 Osnabrück; Germany; sarah.tobergte@freenet.de

Corresponding author:
Torsten H. Struck; Zoological Research Museum Alexander Koenig; Adenauerallee 160; D- 53113 Bonn; Germany Phone: +49-225-9122-401; Fax: +49-225-9122-212; E-Mail: torsten.struck.zfmk@unibonn.de

Abstract
Gnathostomulida is a taxon of small marine worms, which exclusively inhabit the interstitium. The evolution of Gnathostomulida has been discussed for decades. Originally regarded as primitive animals with affinities to flatworms, the phylogenetic position of Gnathostomulida has been debated. Given the lack of an anus a close relationship to Platyhelminthes has been maintained (i.e., Plathelminthomorpha hypothesis). Alternative hypotheses proposed Gnathostomulida as being close to Gastrotricha due to the presence of a monociliary epidermis (i.e., Monokonta/Neotrichozoa hypothesis) or to Syndermata based on the complicated jaw apparatus (i.e., Gnathifera hypothesis). Molecular analyses using only few genes were inconclusive. Recent phylogenomic studies brought some progress by placing Gnathostomulida as sister to Syndermata, but support for this relationship was low and depended on the analytical strategy. Herein we present the first data of complete or nearly complete mitochondrial genomes for two gnathostomulids, one gastrotrich and one ploy clad flatworm to address the uncertain phylogenetic affinity of Gnathostomulida. Our analyses found Gnathostomulida as sister to Syndermata (Gnathifera hypothesis). Thorough sensitivity analyses addressing taxon instability, branch length heterogeneity (also known as long branch attraction) and base composition heterogeneity showed that the position of Gnathostomulida is consistent across the different analyses and, hence, independent of potential misleading biases. Moreover, ameliorating these different biases nodal support values increased to maximum values. Thus, our data support the hypothesis that the different jaw apparatuses of Syndermata and Gnathostomulida are indeed homologous structures as proposed by the Gnathifera hypothesis.

Key words: Gnathostomulida, Gnathifera, Gastrotricha, Platyhelminthes, mitochondrial genomes, next generation sequencing

1. Introduction

The 100 species of Gnathostomulida inhabit the marine interstitium of usually detritus-rich sands (Nielsen, 2012). Hence, gnathostomulids are slender and dorsoventrally flattened animals, which range in length from 1 to 4 mm (Brusca and Brusca, 2003). Locomotion in the spaces between the sand grains is achieved by cilia. Interestingly, multiciliated epithelial cells are not occurring, but only monociliated cells, which is unusual for bilaterian interstitial organisms (Nielsen, 2012). A characteristic feature of Gnathostomulida is the presence of a ventral pharyngeal jaw apparatus (Herlyn and Ehlers, 1997; Kristensen and Nørrevang, 1977;

Sørensen et al., 2003; Sterrer et al., 1985). On the other hand, a permanent anus is not present, but a posterior area of the ectoderm lacks a basal membrane and is in direct contact with the hindgut. Hence, this connection might function as a temporary anal opening (Knauss, 1979; Nielsen, 2012). Gnathostomulids are hermaphrodites and the cleavage follows a spiral pattern (Nielsen, 2012; Riedl, 1969).

Traditionally Gnathostomulida has been placed within Platyhelminthes or Nemathelminthes (Kristensen and Funch, 2000; Sørensen, 2002; Sterrer et al., 1985). However, based on morphological data three hypotheses dominated the recent discussion about their placement. The first hypothesis follows the traditional view by placing Gnathostomulida as sister to Platyhelminthes (Ax, 1985, 1995; Eernisse et al., 1992; Giribet et al., 2000; Meglitsch and Schram, 1991; Peterson and Eernisse, 2001; Schram and Ellis, 1994; Zrzavy et al., 1998). This hypothesis is also known as the Plathelminthomorpha hypothesis. The second hypothesis, the Monokonta or Neotrichozoa hypothesis, places Gnathostomulida as sister to Gastrotricha (Cavalier-Smith, 1998; Zrzavy et al., 2001) and the third one, the Gnathifera hypothesis, as closely related to Syndermata (Ahlrichs, 1997; Haszprunar, 1996; Herlyn and Ehlers, 1997; Kristensen and Funch, 2000; Melone et al., 1998; Nielsen, 2012; Sørensen et al., 2000; Zrzavy, 2003). All these three hypotheses have in common that they favour a closer relationship to platyzoan taxa. Platyzoa was a taxon grouping simple-bodied bilaterians such as Platyhelminthes, Gastrotricha, Syndermata and Gnathostomulida together (e.g., Cavalier-Smith, 1998), but has recently been shown to be a paraphyletic assemblage (Struck et al., 2014). Molecular analyses using a single or few genes were inconclusive regarding the position of Gnathostomulida. These analyses found different placements of Gnathostomulida as sister to Syndermata (e.g., Zrzavy, 2003), to Gastrotricha (e.g., Paps et al., 2009a; Zrzavy et al., 2001), within Platyzoa (e.g., Giribet et al., 2000; Todaro et al., 2006) or as sister to all other spiralian taxa (e.g., Baguñà et al., 2008; Paps et al., 2009b). However, support for any position remained low. Recently, phylogenomic analyses using hundreds of genes were able to robustly resolve problematic phylogenetic relationships in different spiralian taxa (Andrade et al., 2014; Kocot et al., 2011; Smith et al., 2011; Struck et al., 2011; Weigert et al., 2014). First phylogenomic studies addressing bilaterian relationships including Gnathostomulida placed them either as sister to Acoela or to Gastrotricha, but support was low and Gnathostomulida was among the unstable taxa in these analyses (Dunn et al., 2008; Hejnol et al., 2009). Following phylogenomic studies focussing on the phylogeny of platyzoan taxa generally found a close relationship to Syndermata congruent with the Gnathifera hypothesis (Struck et al., 2014; Wey-Fabrizius et al., 2014; Witek et al., 2009). However, bootstrap support for this relationship was usually low and the position of Gnathostomulida depended on the analytical strategy. Using different datasets or analytical methods alternative positions close to Platyhelminthes or Gastrotricha were found and could not be rejected with certainty (Struck et al., 2014; Wey-Fabrizius et al., 2014; Witek et al., 2009). Hence, although phylogenomic data show some support in favour of the Gnathifera hypothesis with respect to the Monokonta or Plathelminthomorpha hypotheses further evidence is needed for the position of Gnathostomulida (Hankeln et al., 2014).

The data of complete or nearly complete mitochondrial genomes have been successfully used to provide additional evidence for the placement of problematic spiralian taxa like Myzostomidae or Diurodrilidae using both sequence and gene order data (e.g., Bleidorn et al., 2007; Golombek et al., 2013). Therefore, to address the uncertain phylogenetic affinity of Gnathostomulida with respect to Syndermata, Platyhelminthes and Gastrotricha we determined the first mitochondrial genomes of Gnathostomulida and Gastrotricha. Several mitochondrial genomes of Platyhelminthes are publically available, but the majority of these genomes belong to the highly derived Neodermata and genomes of turbellarian flatworms are sparse. Two complete genomes of the tricladidan genus *Dugesia* and a partial one of the macrostomorphan *Microstomum* were determined. Polycladida are part of the basal radiation of Platyhelminthes (Egger and Rieger, 2013) and, thus, we also determined the genome of the polycladidan

Stylochoplana maculata. For Syndermata mitochondrial genomes of the different subgroups are already present. In our analyses, the mitochondrial data strongly support the Gnathifera hypothesis even using different datasets and analytical strategies. A sistergroup relationship to either Platyhelminthes or Gastrotricha was not found in any of our analyses in contrast to the previous phylogenomic studies. In the light of these findings the morphological characters have been reevaluated.

2. Material and Methods

2.1. Material

Specimens of *Gnathostomula paradoxa* Ax, 1956 and *G. armata* Riedl, 1971 (Gnathostomulida, Bursovaginoidea, location: N 55°01.508'/E 008°26.180') were collected near List, North Sea island Sylt, Germany. Samples of *Lepidodermella squamata* (Dujardin, 1841) (Gastrotricha, Chaetonotida) were obtained from Carolina Biological Supply Company (Burlington, NC, USA). Specimens of *Stylochoplana maculata* (Quatrefages, 1845) (Platyhelminthes, Polycladida, location: N 54°11.326'/E 007°52.220') were collected from rock pools on the North Sea island Helgoland, Germany.

Animals were preserved in RNAlater (Sigma, Hamburg, Germany) or snap frozen in liquid nitrogen and then stored at -70°C. Genomic DNA was extracted using the DNeasy Tissue Kit (QIAGEN, Hilden, Germany) according to manufactures instruction.

Table 1: Sequence information of species-specific primers used to fill gaps. Closing the gaps in *Lepidodermella squamata* and *Gnathostomula paradoxa* was not successful, but the area of the stretch of N's in *G. paradoxa* could be determined. The primers, which were successful, are indicated by an asterix.

Primer	Sequence (5' -> 3')	Direction	Tm (°C)
For *Gnathostomula paradoxa*			
*Gpara_in_F	TGGCAGAGTAAGTGCATAATATTTAG	Forward	62.0
*Gpara_in_R	AGAAGATTCCAAAAATAACACTCAAAC	Reverse	61.0
Gpara_F	GGTGTATAAAGGAGGTACTAGTGGG	Forward	61.0
Gpara_R	ACGCCGATAAAGGAGGTAGAAG	Reverse	62.0
For *Gnathostomula armata*			
*Garma_F	TTTAATACAGAGAAAGACCAAGGATAC	Forward	62.0
*Garma_R	AGGAGGTAAAATTGAAGGAAAACAG	Reverse	61.0
For *Lepidodermella squamata*			
Lepido_F	CCGCTCACGCCTTTGTAATG	Forward	60.0
Lepido_Fin	CCTGACATGGCTTTTCCTCG	Forward	60.0
Lepido_R	CTGCTCTACTAACGGCAGCA	Reverse	60.0
Lepido_Rin	AAAGCTACAAGAAGGCCCCC	Reverse	60.0
For *Stylochoplana maculata*			
*Stylo_F	CCGCTACGGCCAAAAATACC	Forward	60.0
*Stylo_R	GAATGGCATTGGGTTGTAGCC	Reverse	61.0

2.2. Determination of mitochondrial genomes

The determination of mitochondrial genomes followed the protocol of Golombek et al. (2013). In specific, to increase the amount of genomic DNA the whole genomes of *G. paradoxa*, *G. armata* as well as *L. squamata* were amplified using the illustra GenomiPhi HY DNA

Amplification Kit (GE Healthcare Life Science, Freiburg, Germany). The genomic DNA of *G. paradoxa*, *G. armata*, *L. squamata* and *S. maculata* was send to Genterprise Genomics (Mainz, Germany) for genomic DNA shot-gun library paired-end sequencing on an Illumina HiSeq2000 with TruSeq v3 chemistry, with 22.8 to 30.4 million reads and an average read length of 95 to 97 bp after quality trimming. These libraries were assembled into contigs with CLC Genomic workbench (CLC bio, Aarhus, Denmark) using default parameters with a fragment size window of 200 to 600 bp as the maximum peak of the fragmented library was at 400 bp and the scaffolding option. Using protein sequence information of the 13 protein7 coding mitochondrial genes of *Platynereis dumerilii* (AF178678) as query sequences in TBLASTN searches we searched in the assembled contigs for fragments of the mitochondrial genomes. One large fragment of the mitochondrial genome was found in all libraries. Moreover, due to the scaffolding option there was a stretch of undetermined nucleotides in the fragment of *G. paradoxa*. Hence, we employed two strategies to complete the mitochondrial genomes using species-specific primers (Table 1), which were designed with the aid of the Primer3Plus web-interface (Untergasser et al., 2007). First, we tried to close the gap of undetermined nucleotides within the fragment of *G. paradoxa*. Second, as the single fragments were linear after the assembly we tried to close gap between the ends of the fragment for each species. By this latter approach we showed that the mitochondrial genomes are indeed circular. The lacking fragments were amplified from the same genomic DNA as above using the QIAGEN® Multiplex PCR Kit (QIAGEN, Hilden, Germany) (20 µl reaction: 10 µl multiplex mix, 2 µl Q solution, 1.6 µl 10 pmol/µl forward primer, 1.6 µl 10 pmol/µl reverse primer, 2.5 µl genomic DNA and 2.3 µl water) and a touch down PCR (initial denaturation: 15’ 95°C; 15 cycles: 35” 94°C, 90” 55°C or 60°C (decreasing 1°C at each cycle), 90” 72°C; 25 cycles: 35” 94°C, 90” 50°C or 55°C, 90” 72°C; final elongation: 10’ 72°C). In the case of *L. squamata* also nested PCR approaches had been used, as the fragment was especially difficult to amplify. For the nested PCR the same conditions as for either the normal or the touch down PCR above have been tested. The only differences were that instead of the genomic DNA 2.5 µl of the previous PCR reaction diluted 1:10 with water were used and a set of internal primer pairs instead of the external pair. PCR fragments of the expected sizes were excised from TBE agarose gels and purified via NucleoSpin® Gel and PCR Cleanup (MACHEREY-NAGEL, Düren, Germany). Purified fragments were sent to Macrogen Europe (Amsterdam, Netherland) for sequencing. The sequences of the amplified fragments were assembled with present contigs using SeqMan II (DNASTAR Inc., Madison, WI, USA).

The mitochondrial genomes were annotated using the MITOS webserver with the invertebrate mitochondrial code (NCBI code table 5) and manually corrected based on the raw data of the pipeline provided by the MITOS webserver (Bernt et al., 2013). Specifically the positions of the mitochondrial rRNA’s and tRNA’s were checked. Additionally, we screened the mitochondrial genomes for tRNA’s, which were lacking based on the MITOS results using ARWEN (Laslett and Canbäck, 2008) and tRNAscan-SE (Schattner et al., 2005) as well as manually. However, in any case additional screenings did not find any of the lacking tRNA’s and indeed found less than MITOS congruent with recent results about the performance of different tRNA annotation tools (Jühling et al., 2012).

2.3. Phylogenetic analyses

We obtained complete or nearly complete mitochondrial (mt) genomes from GenBank (Table 2) and annotated each genome using the MITOS webserver with the invertebrate mitochondrial code (NCBI code table 5) (Bernt et al., 2013) to ensure that the annotation procedure of mitochondrial genes was the same for all taxa. From these genomes we compiled a dataset of all mitochondrial protein-coding genes except for NADH dehydrogenase subunit 4L (nad4L), ATP synthase subunit 6 (atp6) and 8 (atp8) due to their very high variability. The nucleotide sequences were translated into amino acid (aa) sequences for each gene using the invertebrate

mitochondrial code (NCBI code table 5) with the aid of GeneDoc (Nicholas and Nicholas, 1997). The aa sequences of each gene were aligned using MAFFT (Katoh et al., 2005). MAFFT was used without any specific prior setting, so that the program by itself did chose the most appropriate alignment method, which was L-INSI in each case. For each alignment ambiguously aligned position were masked using AliScore and AliCut (Kück et al., 2010; Misof and Misof, 2009). Finally, the alignments were concatenated into one dataset using FASconCAT (Kück and Meusemann, 2010).

Maximum Likelihood (ML) analyses were conducted with RAxML 7.3.1 (Stamatakis, 2006). Besides non-partitioned analyses partitioned analyses applying individual substitution models and branch lengths for each gene were also conducted. The best fitting substitution model was determined using the ProteinModelSelection script of RAxML 7.3.1. For the concatenated dataset as well as for each gene individually the MTZOA+Γ+I+F substitution model was determined, except for the genes cytochrome oxidase subunit 3 (cox3) and NADH dehydrogenase subunit 3 (nad3), for which no additional empirical frequencies (+F) had to be applied. Confidence values for the edges of the ML tree were computed using the automatic bootstopping option (-# autoMRE) in RAxML to a maximum of 1,000 bootstrap (BP) replicates (Felsenstein, 1985).

As previous studies (e.g., Struck et al., 2014; Zhong et al., 2011) had shown that Gnathostomulida, Platyhelminthes or Syndermata as well as mitochondrial data in general can be affected by misleading biases we also conducted a sensitivity analysis to assess the influence of biased taxa on the monophyly of Gnathostomulida, Platyhelminthes or Syndermata and the position of Gnathostomulida. First, we determined leaf stability indices of species using Phyutility (Smith and Dunn, 2008) and the bootstrap trees of both the nonpartitioned and partitioned ML analyses. Second, to assess the branch length heterogeneity among species of the non-partitioned ML tree we used the LB score (Struck et al., 2014) using TreSpEx (Struck, 2014). We did not use the partitioned ML tree, which included unlinked branch length between different partitions. In this case the terminal branch length leading to a species is not only based on differences in substitution rates, but also on the presence or absence of partitions in the species (Golombek et al., 2013). Third, base composition heterogeneity for each species was calculated based on the taxon-specific RCFV value (Zhong et al., 2011) using BaCoCa (Kück and Struck, 2014). Given these criteria, taxa were excluded from the dataset and the ML analyses repeated (Tables 3 & 4).

3. Results

3.1. Determination of the mitochondrial genomes

The small genomic shotgun libraries assembled into 38,352 to 84,916 contigs larger than 300 bp and 4,496 to 38,032 larger than 1,000 bp. The N50 values of the libraries were between 687 to 2,393 bp and the average contig length between 780 to 1,530 bp. Moreover, each library contained one larger fragment of the mitochondrial genome. For *Gnathostomula paradoxa* the fragment had a length of 14,197 bp with 36.4x average coverage with a stretch of 64 undetermined nucleotides, which were determined using species-specific primers. For *G. armata* the fragment had a length of 14,230 bp with 112.0x coverage, for *Lepidodermella squamata* 14,558 bp with 42.0x coverage and for *Stylochoplana maculata* 15,351 with 112.2x coverage. Using species-specific primers the "gaps" within *G. armata* and *S. maculata* could be closed showing the ring-like structure of these mitochondrial genomes. The total size of the mitochondrial genome of *G. armata* is 14,030 bp showing that the assembly process generated ends, which actually overlapped by 200 bp (Fig. 1A). For *S. maculata* the size is 15,329 bp and again the ends overlapped by 21 bp (Fig. 1C). In *G. paradoxa* and *L. squamata* we were not successful to close the gap or show the ring-like structure (Fig. 1B & D) probably due to the

unknown or control region, which is not unusual for mitochondrial genomes due to its secondary structure and the presence of long nucleotide homopolymers (Halanych and Janosik, 2006; Vallès and Boore, 2006).

Table 2: Sequence accession numbers of mitochondrial genomes used in this study. Sequences that were newly obtained are indicated in boldface.

Taxon	Species	Accession
Annelida	*Diurodrilus subterraneus* Remane, 1934	KC790350
	Phascolopsis gouldii (De Pourtalès, 1851)	AF374337
	Phascolosoma arcuatum (Gray, 1828)	NC_012618
	Sipunculus nudus Linnaeus, 1766	FJ422961
	Endomyzostoma sp.	FJ975144
	Myzostoma seymourcollegiorum Rouse & Grygier, 2005	EF506562
	Nephtys sp.	NC_010559
	Platynereis dumerilii Audouin &Milne-Edwards, 1833	NC_000931
	Orbinia latreillii Audouin &Milne-Edwards, 1833	AY961084
	Questa ersei Jamieson & Webb, 1984	FJ612452
	Scoloplos cf. *armiger*	DQ517436
	Galathealinum brachiosum Ivanov, 1961	AF178679
	Riftia pachyptila Jones, 1981	AY741662
	Pectinaria gouldii (Verrill, 1874)	FJ976040
	Pista cristata (Müller, 1776)	EU239688
	Auchenoplax crinita Ehlers, 1887	FJ976041
	Eclysippe vanelli (Fauvel, 1936)	EU239687
	Terebellides stroemi Sars, 1835	EU236701
	Paralvinella sulfincola Desbruyères & Laubier, 1993	FJ976042
	Clymenella torquata (Leidy, 1855)	AY741661
	Urechis caupo Fisher and MacGinitie, 1928	AY619711
	Urechis unicinctus (von Drasche, 1881)	EF656365
	Lumbricus terrestris Linnaeus, 1758	NC_001673
	Perionyx excavatus Perrier, 1872	NC_009631
	Helobdella robusta Shankland, et al. 1992	AF178680
Ectoprocta	*Bugula neritina* (Linnaeus, 1758)	NC_010197
	Flustra foliacea (Linnaeus, 1758)	NC_016722
	Flustrellidra hispida (Fabricius, 1780)	DQ157889
	Watersipora subtorquata (d'Orbigny, 1852)	NC_011820
Mollusca	*Albinaria caerulea* (Deshayes, 1835)	NC_001761
	Aplysia dactylomela Rang, 1828	NC_015088
	Aplysia vaccaria Winkler, 1955	DQ991928
	Cristaria plicata Leach, 1815	GU944476
	Ilyanassa obsoleta (Say, 1822)	NC_007781
	Graptacme eborea (Conrad, 1846)	AY484748
	Hyriopsis cumingii Lea, 1852	HM347668
	Hyriopsis schlegelii von Martens, 1861	NC_015110
	Katharina tunicata (W. Wood, 1815)	NC_001636
	Loligo opalescens Berry, 1911	GQ225110
	Mytilus galloprovincialis Lamarck, 1819	NC_006886
	Mytilus trossulus Gould, 1850	NC_007687
	Nassarius reticulatus (Linnaeus, 1758)	NC_013248
	Sepia officinalis Linnaeus, 1758	NC_007895
	Sthenoteuthis oualaniensis (Lesson, 1830 in 1830-1831)	NC_010636
Brachiopoda	*Terebratalia transversa* (Sowerby, 1846)	NC_003086
	Laqueus rubellus (Sowerby, 1846)	NC_002322
Phoronida	*Phoronis psammophila* Cori, 1889	AY368231
Nemertea	*Cephalothrix simula* Iwata, 1952	FJ594739
	Cephalothrix sp.	NC_014869
	Emplectonema gracile (Johnston, 1837)	NC_016952
	Lineus viridis (Müller, 1774)	FJ839919

Table 2 (continued)

Taxon	Species	Accession
	Nectonemertes cf. *mirabilis*	NC_017874
	Paranemertes cf. *peregrina*	NC_014865
	Zygeupolia rubens (Coe, 1895)	NC_017877
Entoprocta	*Loxocorone allax* Iseto, 2002	NC_010431
	Loxosomella aloxiata Iseto, 2001	AB264800
Syndermata	*Adineta vaga* Cuvier, 1817	JX184001 (CO1), JX184009 (CO2), JX184017 (CO3), JX184025 (Cytb), JX184033 (NAD1), JX184041 (NAD2), JX184048 (NAD3), JX184056 (NAD4), JX184072 (NAD5), JX184080 (NAD6)
	Brachionus plicatilis Mueller, 1786	NC_010472
	Echinorhynchus truttae Schrank, 1788	NC_019805
	Leptorhynchoides thecatus (Linton, 1891)	NC_006892
	Macracanthorynchus hirudinaceus Pallas, 1781	NC_019808
	Macrotrachela quadricornifera (Milne, 1886)	JX183998 (CO1), JX184005 (CO2), JX184013 (CO3), JX184021 (Cytb), JX184036 (NAD1), JX184037 (NAD2), JX184044 (NAD3), JX184058 (NAD4), JX184075 (NAD5), JX184076 (NAD6)
	Oncicola luehei (Travassos, 1917)	NC_016754
	Philodina citrina Ehrenberg, 1832	NC_019806
	Rotaria rotatoria (Pallas, 1766)	NC_013568
Platyhelminthes	*Benedenia hoshinai* Ogawa, 1984	NC_014591
	Dugesia ryukyuensis Kawakatsu, 1976	AB618488
	Dugesia japonica Ichikawa & Kawakatsu, 1964	AB618487
	Echinococcus multilocularis Leuckart, 1863	NC_000928
	Gyrodactylus thymalli Zitnan, 1960	NC_009682
	Gyrodactylus salaris Malmberg, 1957	NC_008815
	Microcotyle sebastis Goto, 1894	NC_009055
	Microstomum lineare (Müller OF, 1773)	AY228756
	Opisthorchis viverrini (Poirier 1886)	JF739555
	Pseudochauhanea macrorhis Lin, Liu & Zhang in Zhang, Yang & Liu, 2001	NC_016950
	Schistosoma mansoni Sambon, 1907	HE601612
	Schistosoma japonicum (Katsurada, 1904)	NC_002544
	Stylochoplana maculata (Quatrefages, 1845)	**??_??????**
	Taenia crassiceps (Zeder, 1800)	NC_002547
	Taenia solium (Linnaeus, 1758)	NC_004022
Gnathostomulida	*Gnathostomula paradoxa* Ax, 1956	**??_??????**
	Gnathostomula armata Riedl, 1971	**??_??????**
Gastrotricha	*Lepidodermella squamata (Dujardin, 1841)*	**??_??????**
Ecdysozoa	*Priapulus caudatus* Lamarck, 1816	NC_008557
	Bothropolys sp.	NC_009458
	Onisimus nanseni G. O. Sars, 1900	FJ555185
	Epiperipatus biolleyi (Bouvier, 1902)	NC_009082

The partial genome of *G. paradoxa* contains all 13 protein-coding genes, the 2 ribosomal RNA genes and 20 of 22 mitochondrial tRNA. tRNA-L2 and tRNA-K are lacking (Fig. 1B). It remains uncertain if these are actual losses or if they occur in the part of the genome we could not determine. Another option is that although we thoroughly searched for the lacking tRNA's and used the best tools to date (Jühling et al., 2012) we might still have missed these tRNA's in the existing data. Even further improved methods in the future might show if this is the case. Similarly, all genes are present in the complete genome of *G. armata* except for tRNAC, tRNA-Q and tRNA-T. In both gnathostomulid genomes the small ribosomal RNA and the tRNA's L1, M and A are on the opposite strand and in *G. armata* additionally the tRNA's D and K. With respect to tRNA-K, which it is presently lacking in *G. paradoxa*, it is interesting to notice that the position of the tRNA-K in *G. armata* is at a position, where the gap in the genome of *G. paradoxa* is occurring.

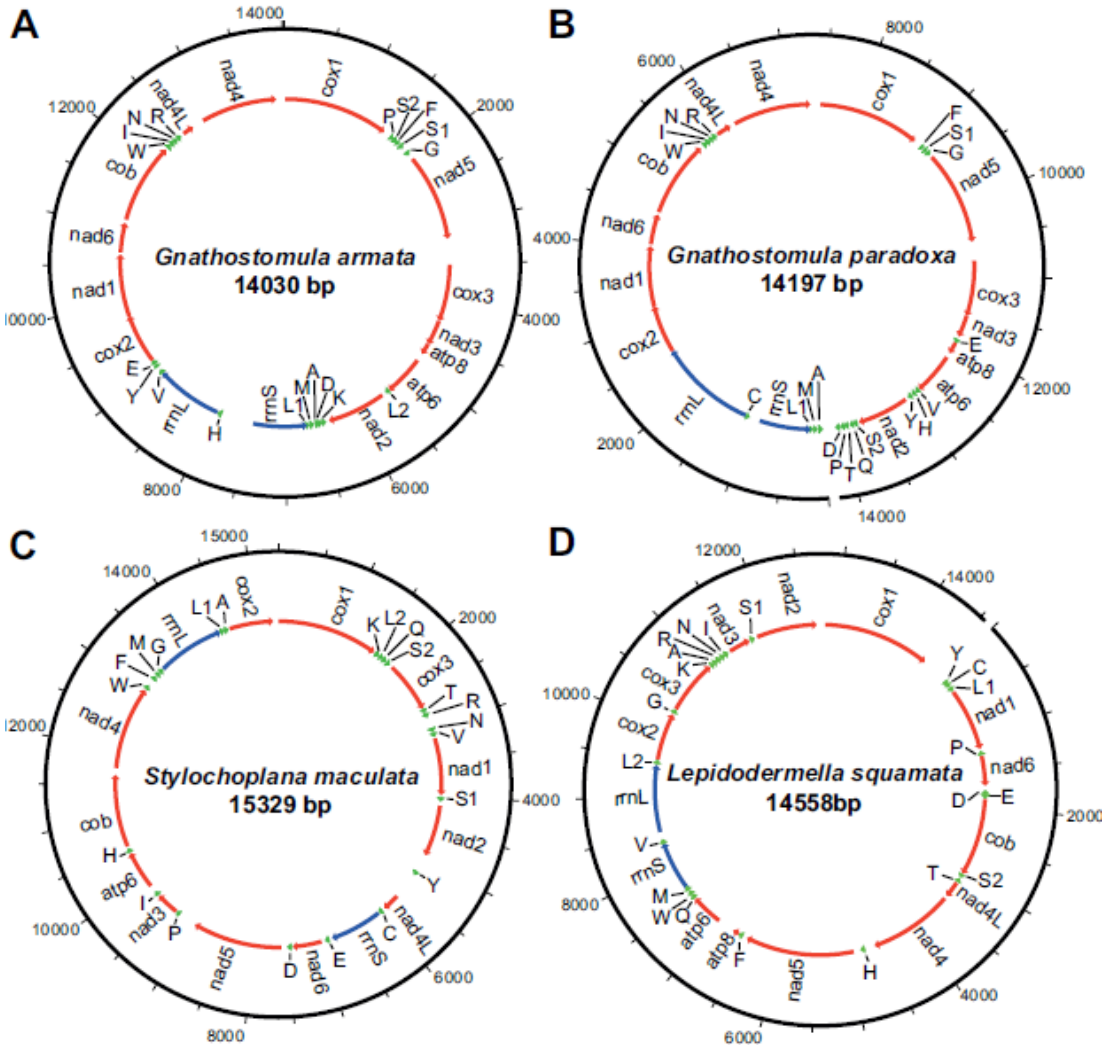


Fig. 1: Gene order of the mitochondrial genomes of (A) *Gnathostomula armata*, (B) *G. paradoxa*, (C) *Stylochoplana maculata* and (D) *Lepidodermella squamata*. Incomplete rings indicate partial mitochondrial genomes. Gene names within the circle indicate that the genes are on the opposite strand. nad1-6, 4L = NADH dehydrogenase subunits 1-6 and 4L; cox1-3 = cytochrome oxidase subunits 1-3; cob = cytochrome b; atp6, 8 = ATP synthase subunits 6 and 8; rrnS = small ribosomal RNA (also known as 12S); rrnL = large ribosomal RNA (also known as 16S); single letters = mitochondrial tRNA's coding for alanine (A), cysteine (C), aspartic acid (D), glutamatic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L1 (CUA), L2 (UUA)), methionine (M), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S1 (AGC), S2 (UCA)), threonine (T), valine (V), tryptophan (W), tyrosine (Y).

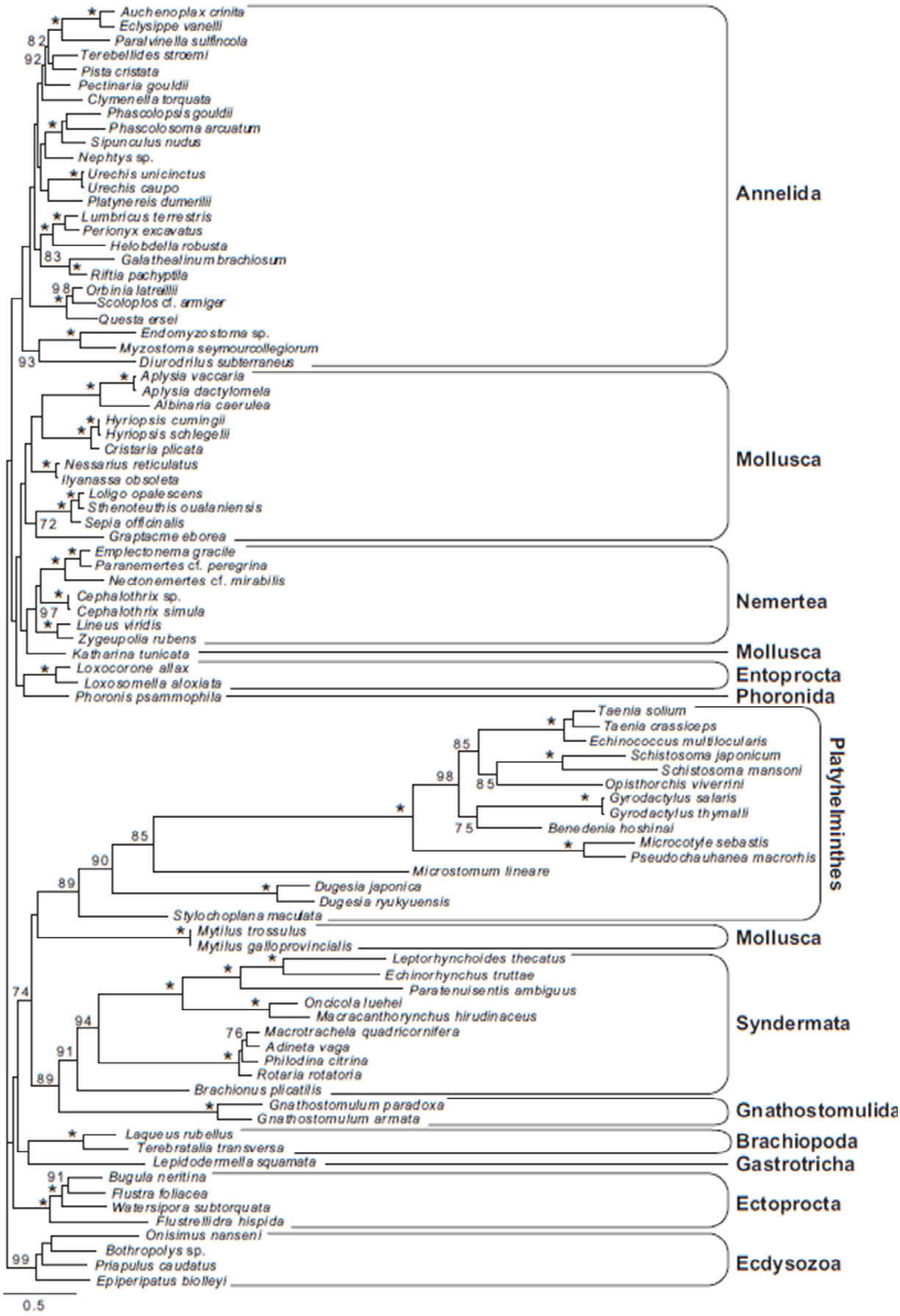


Fig. 2: Phylogram of the non-partitioned ML analysis with all taxa (#2 in Table 3, 88 taxa, -ln L = 148,624.48). Only bootstrap values ≥ 70 are shown. * = bootstrap value of 100. Higher taxonomic units are indicated.

Table 3: Results of the different analyses with respect to monophyly of (A) Platyhelminthes, (B) Syndermata and (C) Gnathostomulida as well as the different hypotheses for the position of Gnathostomulida with (D) Syndermata (Gnathifera hypothesis), (E) Gastrotricha (Monokonta hypothesis) or (F) Platyhelminthes (Plathelminthomorpha hypothesis). Taxa were excluded based on the specified criterion. Abbreviations: part. = partitioned, y = yes, n = no, rep. = bootstrap replicates, ls = leaf stability index, LB = long branch score, RCFV = relative compositional frequency variation, na = not applicable as only one representative of the taxon is present, ns = not supported.

	Strategy				Hypotheses					
#	criterion	# taxa	part.	# rep.	A	B	C	D	E	F
1	none	88	y	250	96	93	100	96	ns	ns
2			n	200	89	91	100	89	ns	ns
3	ls ≤ 0.580 (part.)	83	y	200	94	90	100	94	ns	ns
4			n	300	89	93	100	89	ns	ns
5	ls ≤ 0.629 (non-part.)	76	y	150	97	94	100	97	ns	ns
6			n	250	92	94	100	92	ns	ns
7	ls ≤ 0.707 (non-part.)[1]	70	y	200	98	94	100	98	ns	ns
8			n	250	91	95	100	91	ns	ns
9	LB ≥ 66.76 (non-part.)	77	n	400	99	98	100	99	ns	ns
10			n	250	99	98	100	99	ns	ns
11	LB ≥ 40.27 (non-part.)	73	y	400	100	91	100	100	ns	ns
12			n	300	100	93	100	100	ns	ns
13	LB ≥ 8.31 (non-part.)[2]	64	y	300	na	na	na	99	ns	ns
14			n	300	na	na	na	99	ns	ns
15	RCFV ≥ 0.0022 (non-part.)[2]	71	y	200	94	-	na	96	ns	ns
16			n	250	91	74	na	93	ns	ns
17	RCFV ≥ 0.0012 (non-part.)[3]	30	y	450	na	na	na	99	ns	ns
18			n	300	na	na	na	100	ns	ns

[1]except for *Lepidodermella squamata*
[2]except for *Gnathostomula armata*
[3]except for *Gnathostomula armata*, *Lepidodermella squamata* and *Stylochoplana maculata*

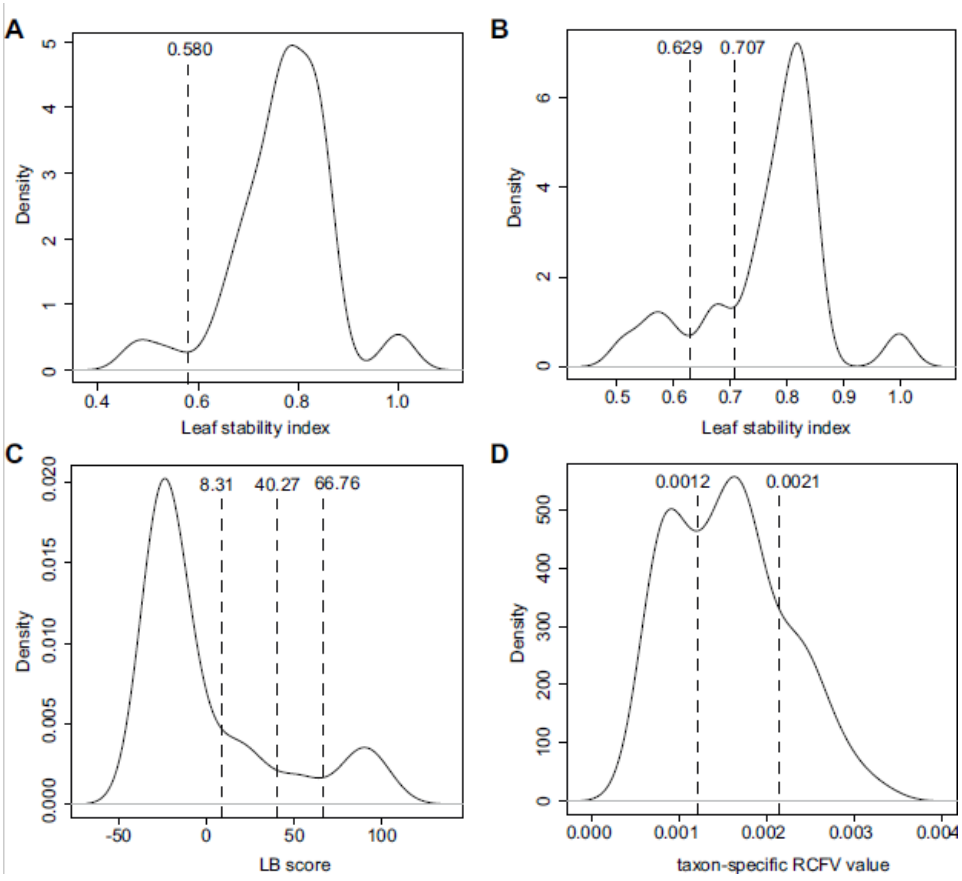


Fig. 3: Density plots of the different taxon-specific measurements. (A) Leaf stability index based on the bootstrap trees from the partitioned ML analysis (#1 in Table 3), (B) leaf stability index based on the bootstrap trees from the non-partitioned ML analysis (#2 in Table 3), (C) LB score based on the best tree of the non-partitioned ML analysis (Fig. 2, #1 in Table 3), and (D) RCFV values based on the dataset with all 88 taxa.

Hence, future completion of the mitochondrial genome of *G. paradoxa* will show if tRNA-K, should it be present, is also on the opposite strand or not. However, tRNA-D is definitely not on the opposite strand in *G. paradoxa* (Fig. 1B). All genes except for the protein-coding gene atp8 were found in the mitochondrial genome of *S. maculata* and the genes were all on the same strand. All mitochondrial protein-coding, ribosomal rRNA and tRNA genes were present in the partial genome of *L. squamata* with the tRNA's P, D and T being on the opposite strand (Fig. 1D).

3.2. Phylogenetic analyses of mitochondrial protein-coding genes

The concatenated dataset of the 10 protein-coding genes consisted of 88 taxa and 2,168 amino acid positions with 88.81% coverage. In both the partitioned and non-partitioned ML analyses of this dataset *Stylochoplana maculata* was part of monophyletic Platyhelminthes with bootstrap support (BS) values of 96 and 89, respectively, and the two gnathostomulid species grouped together with BS values of 100 (Fig. 2 and Table 3). In both analyses Platyhelminthes and Gnathostomulida were part of a clade, which also comprised Syndermata, Gastrotricha, Brachiopoda, Ectoprocta and *Mytilus* (Mollusca), but with weak bootstrap support (BS < 70, Fig. 2) only. Within this clade Gnathostomulida was sister to Syndermata (BS = 96 and 89, Fig. 2 and Table 3). Monophyly of Syndermata was also supported by BS values of 93 and 91 (Fig. 2 and Table 3) and Gastrotricha was sister to Brachiopoda with weak bootstrap support (Fig. 2). Finally, monophyly of Annelida, Nemertea, Entoprocta, Brachiopoda, and Ectoprocta, but not of Mollusca, was found in both analyses (Fig. 2).

The leaf stability index of the partitioned ML analysis revealed that the majority of the taxa had a leaf stability index around 0.8 (Fig. 3A & Table 4). One local optimum occurred at a leaf stability index below 0.580 (Fig. 3A) comprising a total of five species (i.e., the molluscs *Cristaria plicata*, *Hyriopsis cumingii* and *H. schlegelii*, and the two entoprocts; Table 4). Hence, a dataset was generated excluding these five species (Table 3). Using the leaf stability indices of the non-partitioned ML analysis resulted in a slightly different distribution (Fig. 3B). The global optimum is still at around 0.8, but two local optima occurred at lower values. One comprised 12 species with values below 0.629. Besides the three molluscan species from above these were the annelids *Diurodrilus subterraneus*, *Endomyzostoma* sp. and *Myzostoma seymourcollegiorum*, the molluscs *Albinaria caerulea*, *Aplysia dactylomela*, *A. vaccaria* and *Graptacme eborea* as well as the two brachiopod species (Table 4). The other one had seven additional species with values below 0.707 (i.e., the gastrotrich *Lepidodermella squamata*, the four ectoprocts and the two entoprocts). Except for the only gastrotrich *L. squamata* all other taxa were excluded based on these two thresholds generating two datasets with 76 and 70 species, respectively (Table 3).

The majority of the species in the non-partitioned ML analysis had a LB scores around -25 (Table 4), but a long tail of the distribution could be observed at higher LB scores (Fig. 3C). An additional local optima could be observed at values higher than 66.76. Moreover, slight increases in the distribution curve could be observed at values higher than 40.27 and 8.31. These three areas covered 11, 15 and 25 species respectively. Among others this included the two gnathostomulid species with values of 16.52 for *G. armata* and 19.61 for *G. paradoxa* (Table 4). As *G. armata* had the lower value of the two gnathostomulids all other species except *G. armata* were excluded based on these three thresholds resulting in three datasets with 77, 73 and 64 species, respectively (Table 3).

Table 4: Results of the sensitivity analyses for each species with respect to leaf stability, branch length heterogeneity and base composition heterogeneity. Values below or above the highest or lowest threshold, respectively (Table3), are in bold. Abbreviations: p. = partitioned, n-p. = non-partitioned, ls = leaf stability index, LB = long branch score, RCFV = relative compositional frequency variation.

Taxon	**Species**	**ls p.**	**ls n-p.**	**LB n-p.**	**RCVF n-p.**
Annelida	*Auchenoplax crinita*	0.785	0.807	-20.04	0.0008
	Clymenella torquata	0.779	0.802	-28.60	**0.0016**
	Diurodrilus subterraneus	0.752	**0.623**	-11.71	**0.0021**
	Eclysippe vanelli	0.785	0.807	-20.40	0.0007
	Endomyzostoma sp.	0.738	**0.565**	-11.94	**0.0028**
	Galathealinum brachiosum	0.778	0.801	-19.32	**0.0019**
	Helobdella robusta	0.779	0.801	-22.13	**0.0017**
	Lumbricus terrestris	0.779	0.801	-29.82	**0.0014**
	Myzostoma seymourcollegiorum	0.738	**0.565**	-17.69	**0.0019**
	Nephtys sp.	0.766	0.792	-31.16	**0.0016**
	Orbinia latreilli	0.767	0.794	-28.17	**0.0017**
	Paralvinella sulfincola	0.785	0.807	-21.37	**0.0017**
	Pectinaria gouldii	0.784	0.807	-32.19	**0.0016**
	Perionyx excavatus	0.779	0.801	-29.59	**0.0016**
	Phascolopsis gouldii	0.771	0.794	-23.85	**0.0020**
	Phascolosoma arcuatum	0.771	0.794	-22.47	**0.0017**
	Pista cristata	0.785	0.806	-30.79	**0.0018**
	Platynereis dumerilii	0.767	0.793	-28.54	**0.0015**
	Questa ersei	0.767	0.794	-23.83	**0.0019**
	Riftia pachyptila	0.778	0.801	-27.09	**0.0016**
	Scoloplos cf. *armiger*	0.767	0.794	-24.36	**0.0017**
	Sipunculus nudus	0.771	0.794	-28.06	**0.0018**
	Terebellides stroemi	0.785	0.807	-30.01	**0.0017**
	Urechis caupo	0.777	0.800	-28.49	**0.0014**
	Urechis unicinctus	0.777	0.800	-28.43	**0.0014**
Ectoprocta	*Bugula neritina*	0.779	**0.674**	-20.24	**0.0017**
	Flustra foliacea	0.779	**0.674**	-19.05	0.0008
	Flustrellidra hispida	0.780	**0.674**	-7.05	0.0009
	Watersipora subtorquata	0.779	**0.674**	-18.84	**0.0015**
Mollusca	*Albinaria caerulea*	0.666	**0.512**	-8.79	0.0009
	Aplysia dactylomela	0.666	**0.512**	-12.84	0.0007
	Aplysia vaccaria	0.666	**0.512**	-13.21	0.0007
	Cristaria plicata	**0.479**	**0.585**	-23.21	0.0008
	Graptacme eborea	0.617	**0.608**	-20.55	0.0009
	Hyriopsis cumingii	**0.479**	**0.585**	-22.91	0.0009
	Hyriopsis schlegelii	**0.479**	**0.585**	-22.68	0.0009
	Ilyanassa obsoleta	0.703	0.753	-34.11	0.0011
	Katharina tunicata	0.715	0.757	-31.09	0.0008
	Loligo opalescens	0.688	0.749	-26.68	0.0007
	Mytilus galloprovincialis	0.769	0.725	0.85	0.0011
	Mytilus trossulus	0.769	0.725	0.83	0.0010
	Nessarius reticulatus	0.703	0.753	-33.62	**0.0012**
	Sepia officinalis	0.688	0.749	-27.79	0.0008
	Sthenoteuthis oualaniensis	0.688	0.749	-26.75	0.0006
Brachiopoda	*Laqueus rubellus*	0.649	**0.553**	-16.96	**0.0012**
	Terebratalia transversa	0.649	**0.553**	-13.50	0.0009
Phoronida	*Phoronis psammophila*	0.670	0.712	-29.46	0.0008

Table 4 (continued)

Taxon	Species	ls p.	ls n-p.	LB n-p.	RCVF n-p.
Nemertea	*Cephalothrix simula*	0.718	0.764	-30.66	0.0008
	Cephalothrix sp.	0.718	0.764	-30.85	0.0008
	Emplectonema gracile	0.718	0.764	-25.12	**0.0014**
	Lineus viridis	0.718	0.764	-30.41	**0.0014**
	Nectonemertes cf. *mirabilis*	0.718	0.764	-21.31	**0.0012**
	Paranemertes cf. *peregrina*	0.718	0.764	-24.81	**0.0013**
	Zygeupolia rubens	0.718	0.764	-29.70	**0.0015**
Entoprocta	*Loxocorone allax*	**0.543**	**0.668**	-29.26	0.0007
	Loxosomella aloxiata	**0.543**	**0.668**	-27.33	0.0006
Syndermata	*Adineta vaga*	0.840	0.833	**14.85**	**0.0019**
	Brachionus plicatilis	0.839	0.832	-1.28	0.0011
	Echinorhynchus truttae	0.840	0.834	**48.36**	**0.0025**
	Leptorhynchoides thecatus	0.840	0.834	**50.25**	**0.0021**
	Macracanthorynchus hirudinaceus	0.840	0.834	**29.31**	**0.0030**
	Macrotrachela quadricornifera	0.840	0.833	**14.63**	**0.0018**
	Oncicola luehei	0.840	0.834	**25.63**	**0.0033**
	Paratenuisentis ambiguus	0.840	0.834	**55.30**	**0.0026**
	Philodina citrina	0.840	0.833	**14.19**	**0.0017**
	Rotaria rotatoria	0.840	0.833	**12.87**	**0.0017**
Platyhelminthes	*Benedenia hoshinai*	0.840	0.835	**71.86**	**0.0017**
	Dugesia japonica	0.838	0.832	**27.70**	**0.0027**
	Dugesia ryukyuensis	0.838	0.832	**29.03**	**0.0023**
	Echinococcus multilocularis	0.840	0.836	**82.81**	**0.0024**
	Gyrodactylus salaris	0.840	0.836	**88.52**	**0.0024**
	Gyrodactylus thymalli	0.840	0.836	**88.72**	**0.0024**
	Microcotyle sebastis	0.841	0.837	**99.42**	**0.0023**
	Microstomum lineare	0.838	0.831	**53.92**	**0.0019**
	Opisthorchis viverrini	0.840	0.836	**88.44**	**0.0020**
	Pseudochauhanea macrorhis	0.841	0.837	**97.07**	**0.0022**
	Schistosoma japonicum	0.840	0.836	**94.18**	**0.0024**
	Schistosoma mansoni	0.840	0.836	**103.53**	**0.0029**
	Stylochoplana maculata	0.838	0.829	-8.39	**0.0015**
	Taenia crassiceps	0.840	0.836	**86.53**	**0.0026**
	Taenia solium	0.840	0.836	**85.06**	**0.0024**
Gnathostomulida	*Gnathostomulum armata*	0.839	0.830	**16.52**	**0.0022**
	Gnathostomulum paradoxa	0.839	0.830	**19.61**	**0.0025**
Gastrotricha	*Lepidodermella squamata*	0.715	**0.689**	-8.16	**0.0013**
Ecdysozoa	*Bothropolys* sp.	1.000	0.999	-22.82	0.0009
	Epiperipatus biolleyi	1.000	0.999	-24.52	**0.0012**
	Onisimus nanseni	1.000	0.998	-10.63	0.0009
	Priapulus caudatus	0.999	0.998	-25.36	**0.0012**

Finally, the distribution of the taxon-specific RCFV values showed two optima, which were relatively close to each other, with a more or less similar density (Fig. 3D). One optimum had values below 0.0012 and the other above this value. Additionally, a shoulder occurred in the distribution at values higher than 0.021. A total of 61 species had values higher than 0.0012 including 17 species exhibiting values of higher than 0.021 (Table 4). These included among others the two gnathostomulids (RCFV = 0.0022 and 0.0025), the only gastrotrich (RCFV = 0.0013) and all Platyhelminthes (RCFV ≥ 0.0015). Based on these two thresholds all taxa were excluded except for one representative for Gnathostomulida (*G. armata*), Gastrotricha (*L. squamata*) and Platyhelminthes (*S. maculata*) with the lowest RCFV value resulting in two datasets with 71 and 30 species, respectively (Tables 3 & 4).

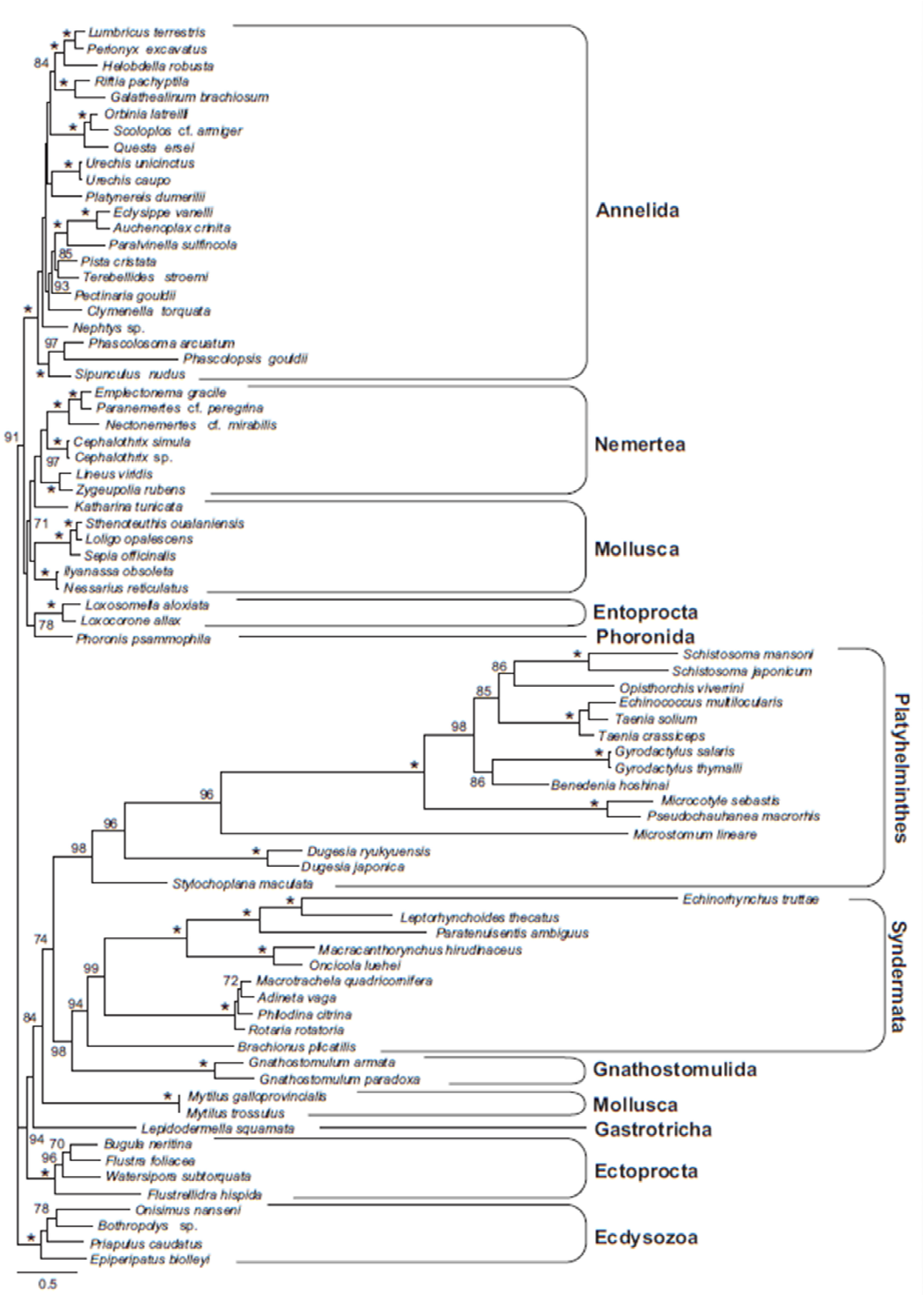

Lumbricus terrestris
Perionyx excavatus
Helobdella robusta
Riftia pachyptila
Galathealinum brachiosum
Orbinia latreilli
Scoloplos cf. armiger
Questa ersei
Urechis unicinctus
Urechis caupo
Platynereis dumerilii
Eclysippe vanelli
Auchenoplax crinita
Paralvinella sulfincola
Pista cristata
Terebellides stroemi
Pectinaria gouldii
Clymenella torquata
Nephtys sp.
Phascolosoma arcuatum
Phascolopsis gouldii
Sipunculus nudus
Annelida
Emplectonema gracile
Paranemertes cf. peregrina
Nectonemertes cf. mirabilis
Cephalothrix simula
Cephalothrix sp.
Lineus viridis
Zygeupolia rubens
Nemertea
Katharina tunicata
Sthenoteuthis oualaniensis
Loligo opalescens
Sepia officinalis
Ilyanassa obsoleta
Nessarius reticulatus
Mollusca
Laxosomella aloxiata
Laxocorone allax
Entoprocta
Phoronis psammophila
Phoronida
Schistosoma mansoni
Schistosoma japonicum
Opisthorchis viverrini
Echinococcus multilocularis
Taenia solium
Taenia crassiceps
Gyrodactylus salaris
Gyrodactylus thymalli
Benedenia hoshinai
Microcotyle sebastis
Pseudochauhanea macrorhis
Microstomum lineare
Dugesia ryukyuensis
Dugesia japonica
Stylochoplana maculata
Platyhelminthes
Echinorhynchus truttae
Leptorhynchoides thecatus
Paratenuisentis ambiguus
Macracanthorynchus hirudinaceus
Oncicola luehei
Macrotrachela quadricornifera
Adineta vaga
Philodina citrina
Rotaria rotatoria
Brachionus plicatilis
Syndermata
Gnathostomulum armata
Gnathostomulum paradoxa
Gnathostomulida
Mytilus galloprovincialis
Mytilus trossulus
Mollusca
Lepidodermella squamata
Gastrotricha
Bugula neritina
Flustra foliacea
Watersipora subtorquata
Flustrellidra hispida
Ectoprocta
Onisimus nanseni
Bothropolys sp.
Priapulus caudatus
Epiperipatus biolleyi
Ecdysozoa
0.5

Fig. 4: Phylogram of the partitioned ML analysis with species excluded with a leaf stability index lower than or equal to 0.707 based on the bootstrap trees from the non-partitioned ML analysis (#7 in Table 3, 70 taxa, -ln L = 129,366.20). Only bootstrap values ≥ 70 are shown. * = bootstrap value of 100. Higher taxonomic units are indicated.

The results of the partitioned and non-partitioned ML analyses of these 8 additional taxonreduced datasets were relatively similar to results of the complete dataset. If applicable, monophyly of Platyhelminthes, Gnathostomulida, Syndermata, Annelida, Nemertea, Entoprocta, Brachiopoda, and Ectoprocta, but not of Mollusca, was generally found (Table 3, Figs. 4-6). Bootstrap support in the partitioned analyses was usually higher than in nonpartitioned analyses. More specifically, monophyly of Gnathostomulida was maximally supported throughout all analyses (hypothesis C in Table 3). Bootstrap support for the monophyly of Platyhelminthes ranged from 89 to 100 (hypothesis A in Table 3). For the monophyly of Syndermata bootstrap support had at least values of 91 or higher except for the dataset excluding species with RCFV values above or equal to 0.0022 (hypothesis B in Table 3). In this case the non-partitioned ML analysis supported the monophyly of Syndermata with only a bootstrap support of 74 and the partitioned analysis found *G. armata* nested within Syndermata (#15 and #16 in Table 3). Finally, with respect to the position of Gnathostomulida the Gnathifera hypothesis (hypothesis D in Table 3) was support by all analyses with bootstrap support values ranging from 89 to 100. Especially the partitioned ML analyses significantly supported this hypothesis with values of 96 or higher. The only exception is the analysis of the dataset in which all species with a leaf stability index of 0.580 or lower were excluded (#3 in Table 3). This analysis had a bootstrap value of 94. In contrast, none of the analyses reconstructed a close relationship of Gnathostomulida to Gastrotricha, the Monokonta hypothesis (hypothesis E in Table 3), or to Platyhelminthes, the Plathelminthomorpha hypothesis (hypothesis F in Table 3).

3.3. Mitochondrial gene order

The order of the protein-coding and rRNA genes of the two gnathostomulid species is the same (Fig. 7). Differences occur only in the position of the tRNA's (Fig. 1). However, the gene order of Gnathostomulida is not similar to either the known gene orders of the Syndermata, Gastrotricha or Playthelminthes (Fig. 7). The most prominent difference to the other gene orders is the position of the rrnS gene on the opposite strand. Hence, in contrast to sequence data the gene order data do neither support nor reject a closer relationship of Gnathostomulida to Syndermata, to Gastrotricha or to Platyhelminthes.

The order of the polycladidan *S. maculata* is different from the known orders of Platyhelminthes (i.e., the neodermatans, the tricladidan *Dugesia* and the partial one of the macrostomorphan *Microstomum*) (Fig. 7). However, *S. maculata* has the lack of atp8 in common with the other platyhelminths.

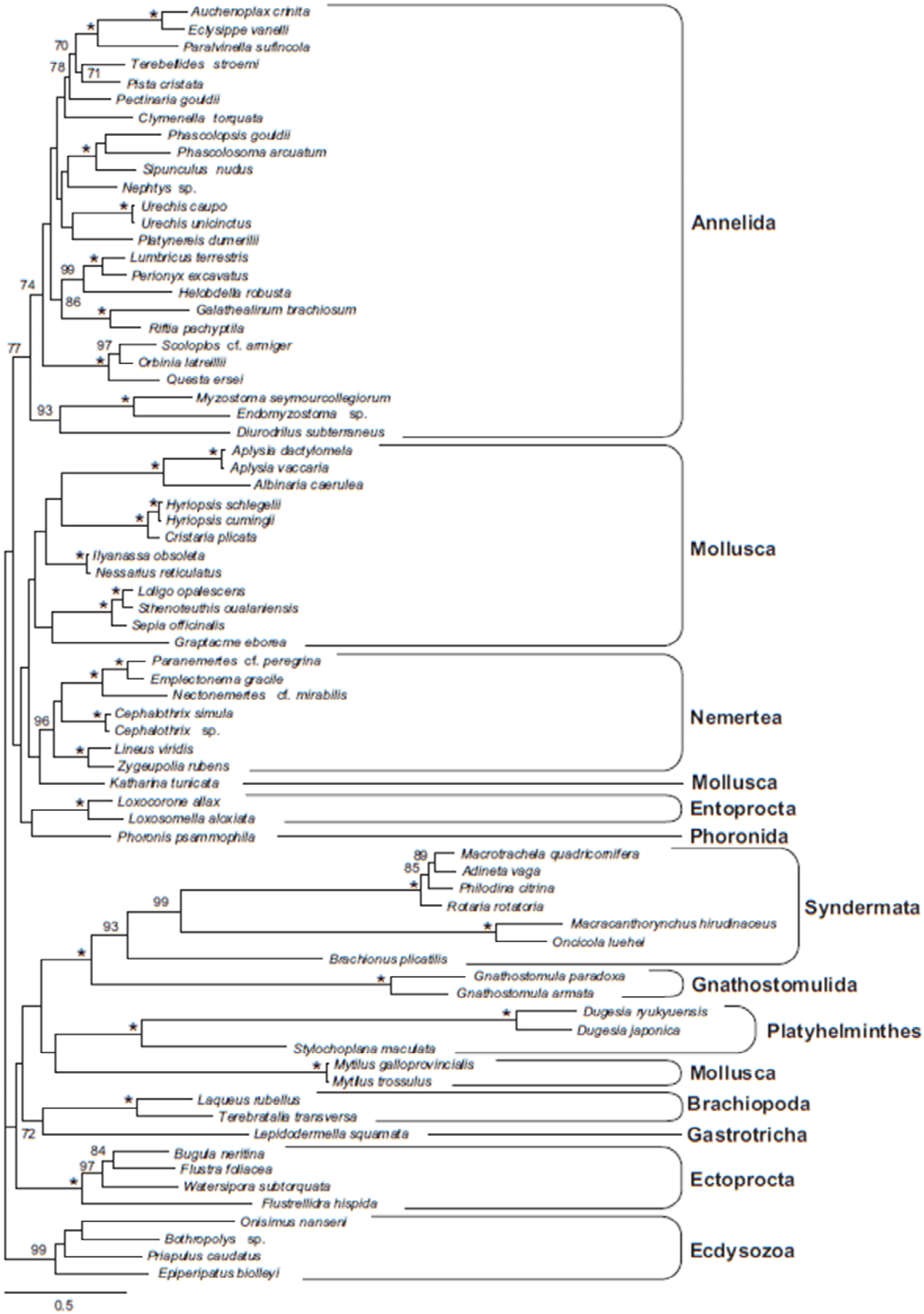


Fig. 5: Phylogram of the non-partitioned ML analysis with species excluded with a LB score higher than or equal to 40.27 based on the best tree from the non-partitioned ML analysis (#12 in Table 3, 73 taxa, -ln L = 113,716.42). Only bootstrap values ≥ 70 are shown. * = bootstrap value of 100. Higher taxonomic units are indicated.

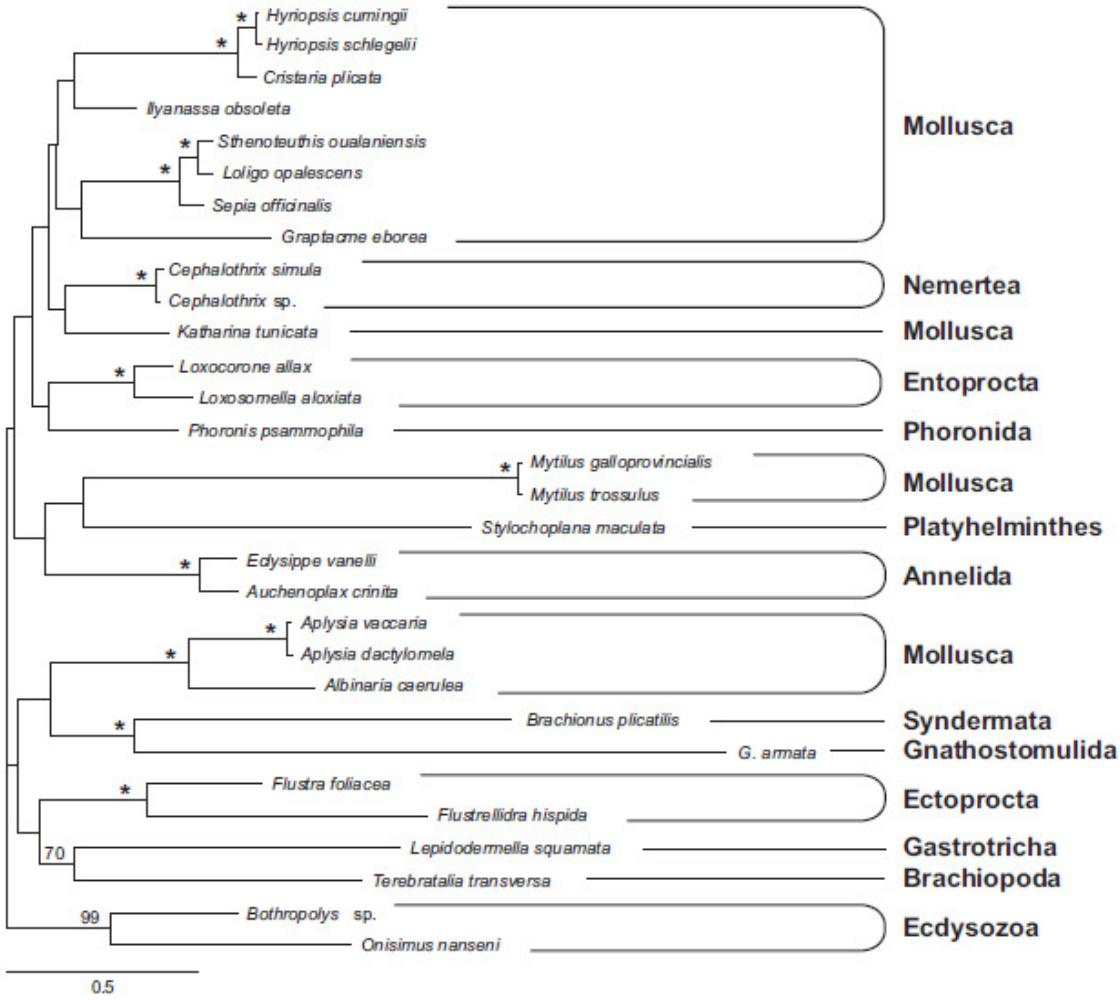


Fig. 6: Phylogram of the non-partitioned ML analysis with species excluded with a taxonspecific RCFV value higher than or equal to 0.0012 (#18 in Table 3, 30 taxa, -ln L = 53,058.12). Only bootstrap values ≥ 70 are shown. * = bootstrap value of 100. Higher taxonomic units are indicated.

4. Discussion

All our analyses herein provide further support for the Gnathifera hypothesis. Even addressing different misleading biases such as unstable taxa, branch length heterogeneity or base composition heterogeneity the results are not altered and support values for the Gnathifera hypothesis actually increase. Recent phylogenomic studies also generally found a close relationship of Gnathostomulida to Syndermata, but were also affected by differences in the datasets and the chosen analytical methods (Struck et al., 2014; Wey-Fabrizius et al., 2014; Witek et al., 2009). The strength and weaknesses of different morphological characters substantiating the Gnathifera, Plathelminthomorpha or Monokonta hypothesis in morphological-cladistic analyses have been discussed in great detail in the comprehensive review of Jenner (2004) about the phylogenetic positions of Platyhelminthes, Nemertea and Gnathostomulida. Therefore, we will concentrate our discussion in the following on the major morphological traits put forward in favor of one or the other hypothesis.

The complicated jaw apparatuses, which can be found in gnathostomulids and the rotiferan syndermatans, are the most prominent and name-giving character supporting the Gnathifera hypothesis. The homology of these jaws had been suggested based on similarities such as the

presence of tube-like support rods composed of electron lucent material surrounding an electron-dense core, the presence of pincers, which caudally articulate into unpaired pedicles, and the presence of cross-striated pharyngeal muscles attaching to the jaw elements through epithelial cells (Ahlrichs, 1997; Haszprunar, 1996; Herlyn and Ehlers, 1997; Kristensen and Funch, 2000; Rieger and Tyler, 1995; Sørensen, 2000, 2003; Sørensen and Sterrer, 2002). These three features of the jaws have been regarded as separate characters in some morphological-cladistic analyses (Nielsen, 2001; Sørensen et al., 2000), but it is not certain if they constitute truly independent identities allowing independent character coding (Jenner, 2004). Nonetheless, if coded separately or not, these characters increase the probability of the homology of the jaw apparatuses based on a homology complexity test (e.g., Scholtz, 2010). Another pharyngeal character, which supported the Gnathifera hypothesis in morphologicalcladistic analyses, was an extended non-contractile region of the pharyngeal musculature (Zrzavy, 2003). Additionally, the location of the protonephridial lumen within the protonephridial cells also supported the monophyly of Gnathifera (Ahlrichs, 1995; Haszprunar, 1996; Zrzavy, 2003). This is in contrast to the typical situation in which the cells enfold the lumen. The difference between these two states can be recognized by the presence of intercellular junctions that close off the extracellular lumen, when the cells enfold the lumen, and the lack of such junctions, when the lumen is inside the cells (Jenner, 2004). Besides these characters several other ones like the presence of buccal ganglia or mouth regions with a chitinous membrane have been proposed to substantiate the Gnathifera hypothesis (e.g., Ahlrichs, 1997; Nielsen, 2001), but are either lacking in the ground pattern of Gnathostomulida or Syndermata or wide-spread across Bilateria indicating most likely a plesiomorphic condition (Jenner, 2004). In summary, our analyses herein strongly support the homology of the complex jaw apparatuses and of the structure of the protonephridial canal and that they are autapomorphic characters of Gnathifera.

Given the sheer amount of supporting characters the Plathelminthomorpha hypothesis is seemingly better substantiated by morphological characters than the Gnathifera hypothesis (Jenner, 2004); a total of 21 different morphological characters have been put forward in the different cladistic analyses supporting the Plathelminthomorpha hypothesis (Ax, 1985, 1995; Eernisse et al., 1992; Giribet et al., 2000; Meglitsch and Schram, 1991; Peterson and Eernisse, 2001; Schram and Ellis, 1994; Zrzavy et al., 1998) in contrast to 12 characters for the Gnathifera hypothesis (Ahlrichs, 1997; Haszprunar, 1996; Herlyn and Ehlers, 1997; Kristensen and Funch, 2000; Melone et al., 1998; Nielsen, 2001; Sørensen et al., 2000; Zrzavy, 2003). The character most often found to substantiate the Plathelminthomorpha hypothesis was the lack of a permanent anus in Platyhelminthes and Gnathostomulida (Eernisse et al., 1992; Giribet et al., 2000; Meglitsch and Schram, 1991; Zrzavy et al., 1998). Gnathostomulids might possess a temporary anus as a posterior area of the ectoderm lacks a basal membrane and is in direct contact with the hindgut (Knauss, 1979). Similar conditions have been found in several flatworms, but also in Micrognathozoa, some gastrotrichs, and even in Arthrotardigrada (Jenner, 2004; Sørensen, 2003). Hence, this character is not unique for Gnathostomulida and at best homoplastic. However, this also makes convergent evolution more likely. Similarly, most other characters such as the presence of protonephridia, direct internal fertilization and hermaphroditism supporting the Plathelminthomorpha hypothesis are also widespread across Metazoa and hence homoplastic as well (Jenner, 2004). For example, protonephridia can also be found in, among others, several annelid taxa, Gastrotricha or Syndermata (Nielsen, 2012; Rouse and Pleijel, 2001). Internal fertilization and direct sperm transfer can be found in several other interstitial taxa as well (Giere, 2009). As Gnathostomulida as well as several flatworms also inhabit the interstitial realm, this might rather be indicative of the adaptation to the interstitium than of common ancestry (Westheide, 1984).

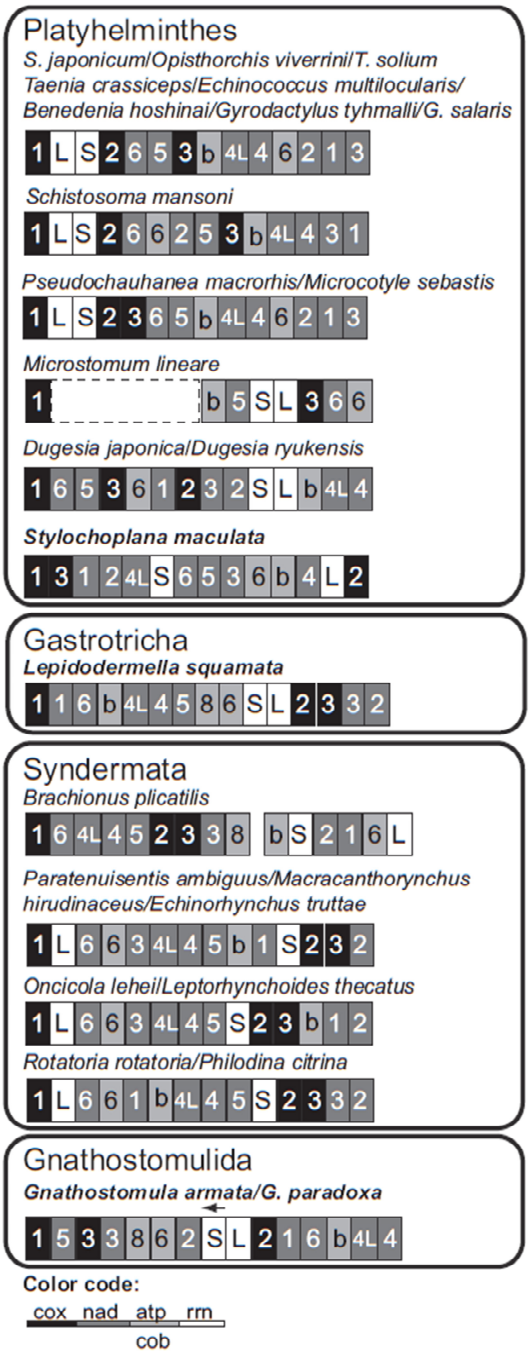


Fig. 7: The orders of the protein-coding and rRNA gene of all published mitochondrial genomes from Platyhelminthes, Gastrotricha, Syndermata and Gnathostomulida. Cytochrome oxidase subunits are in black, NADH dehydrogenase subunits in dark grey, ATP synthase and cytochrome b (= b) in light grey and rRNA genes (L = 16S, S = 12S) in white. Numbers indicate specific subunit (e.g., dark box with 1 = cytochrome oxidase subunit 1). Arrows indicate that the genes are on the opposite strand. The dashed box in *M. lineare* indicates the lacking region of this genome.

Finally, misscorings of characters of crucial taxa including Gnathostomulida and Platyhelminthes resulted in erroneous support for the Platyhelminthomorpha hypothesis (Jenner, 2004). In contrast to the Plathelminthomorpha or Gnathifera hypotheses, the Monokonta hypothesis was only supported by a single morphological-cladistic analysis (Zrzavy et al., 2001). Characters supporting the monophyly of Monokonta were the loss of multiple

protonephridial terminal cells, protonephridial filter formed by weir-like fenestrations of the terminal cell's wall and serial protonephridia. However, the three characters can be found in other metazoan taxa as well and are thus homoplastic (Jenner, 2004). For example, protonephridia with a single terminal cell can also be found in Entoprocta, Cycliophora, Mollusca, *Lobatocerebrum* and Annelida (Jenner, 2004; Zrzavy et al., 2001). Moreover, as for internal fertilization and direct sperm transfer above the presence of protonephridia with a single terminal cell might be a convergent adaptation to the requirement of small body sizes in the interstitium (Smith and Ruppert, 1988; Westheide, 1986). Another character that has been proposed in favour of the Monokonta hypothesis is the presence of monociliated epidermal cells (Cavalier-Smith, 1998; Zrzavy et al., 1998). However, monociliated epidermal cells also occur in Phoronida, Brachiopoda and some annelid taxa (Bartolomaeus, 1995; Gardiner, 1978; Nielsen, 2002; Westheide, 1997). Besides being a symplesiomorphic condition (Gardiner, 1978) it has also been argued that monociliarity can be caused by a truncated ciliogenesis (Bartolomaeus, 1995; Hausen, 2005). In any case, as for the other three characters this name-giving character is also homoplastic and thus convergent evolution is more likely. In conclusion, with respect to the morphological data supporting the three hypotheses, "the quality of the unique similarities in pharyngeal morphology … lead to the conclusion that Gnathifera is the most robustly supported clade" (Jenner, 2004). This conclusion as well as the interpretation of the characters supporting the Plathelminthomorpha or Monokonta hypotheses as being homoplastic characters, which evolved independently, is strongly reinforced by our analyses herein.

In addition to Syndermata and Gnathostomulida Micrognathozoa has also been assigned to Gnathifera (Sørensen et al., 2006). Like Gnathostomulida and Syndermata Micrognathozoa possess a complicated jaw apparatus in the pharynx (Kristensen and Funch, 2000). This jaw apparatus shows many ultrastructural similarities to the jaws of the syndermatan taxa Monogononta, Bdelloidea and Seisonidae, as well as to one of Gnathostomulida (Kristensen, 2002; Sørensen, 2003). Moreover, both Micrognathozoa and Syndermata possess an apical intracytoplasmic lamina in the cellular integument, which is unusual in Bilateria (Funch et al., 2005; Nielsen, 2012). However, molecular data and total evidence analyses so far were not able to resolve the phylogenetic position of Micrognathozoa yet (Giribet et al., 2004; Paps et al., 2009b; Sørensen and Giribet, 2006; Sørensen et al., 2006; Worsaae and Rouse, 2008) and more comprehensive molecular data are needed to achieve this goal (Hankeln et al., 2014). Other taxa, which have been discussed to belong to Gnathifera, based on morphological and/or molecular data were Cycliophora and Myzostomida (for review see Funch et al., 2005). However, recent analyses of molecular data supported the traditional views of a close relationship of Entoprocta and Cycliophora as well as of Myzostomida and Annelida (e.g., Bleidorn et al., 2014; Hejnol et al., 2009; Struck et al., 2014; Weigert et al., 2014).

In summary, we herein report the sequence data of the first mitochondrial genomes for Gnathostomulida and Gastrotricha. Analyses of the mitochondrial data strongly support a sistergroup relationship of Gnathostomulida to Syndermata and hence the Gnathifera hypothesis. Ameliorating potentially misleading biases does not deteriorate, but further increases the phylogenetic signal for this relationship. Therefore, homology of the complex jaw apparatuses of these two taxa is further substantiated and can most likely be extended to the micrognathozoan jaw apparatus.

5. Acknowledgments

We would like to thank Maximilian P. Nesnidal (University of Hamburg) and Günter Purschke (University of Osnabrück) in their assistance to collect gnathostomulids and flatworms. We also appreciate the comments of two anonymous reviewers, which improved the manuscript. This study was funded by DFG STR-683/5-2 as part of the priority program 1174 „Deep

Metazoan Phylogeny“ as well as DFG STR-683/8-1 of the Deutsche Forschungsgemeinschaft to THS.

6. References

Ahlrichs, W.H., 1995. Zur Ultrastruktur und Phylogenie von Seison nebaliae Grube,1859 und Seison annulatus Claus, 1876 - Hypothesen zu phylogenetischen Verwandtschaftsverhältnissen innerhalb der Bilateria. Cuvillier, Göttingen.

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

Andrade, S.C.S., Montenegro, H., Strand, M., Schwartz, M., Kajihara, H., Norenburg, J.L., Turbeville, J.M., Sundberg, P., Giribet, G., 2014. A transcriptomic approach to ribbon worm systematics (Nemertea): resolving the Pilidiophora problem. Mol. Biol. Evol.

Ax, P., 1985. The position of the Gnathostomulida and Platyhelminthes in the phylogenetic system of the Bilateria. In: Conway Morris, S., George, J.D., Gibson, R., Platt, H.M. (Eds.), The origins and relationships of lower invertebrates. Clarendon Press, Oxford, pp. 168-180.

Ax, P., 1995. Das System der Metazoa I. Gustav Fischer Verlag, Stuttgart.

Baguñà, J., Martinez, P., Paps, J., Riutort, M., 2008. Back in time: a new systematic proposal for the Bilateria. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 1481-1491.

Bartolomaeus, T., 1995. Secondary monociliarity in the Annelida: monociliated epidermal cells in larvae of *Magelona mirabilis* (Magelonida). Microfauna Mar. 10, 327–332.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313-319.

Bleidorn, C., Eeckhaut, I., Podsiadlowski, L., Schult, N., McHugh, D., Halanych, K.M., Milinkovitch, M.C., Tiedemann, R., 2007. Mitochondrial genome and nuclear sequence data support Myzostomida as part of the annelid radiation. Mol. Biol. Evol. 24, 1690-1701.

Bleidorn, C., Helm, C., Weigert, A., Eeckhaut, I., Lanterbecq, D., Struck, T.H., Tiedemann, R., Hartmann, S., 2014. From morphology to phylgenomics: Placing the enigmatic Myzostomida in the tree of life. In: Wägele, J.W., Bartolomaeus, T. (Eds.), Deep Metazoan Phylogeny: The backbone of the Tree of Life - New insights from analyses of molecules, morphology, and theory of data analysis. De Gruyter, Berlin, pp. 161-172.

Brusca, R.C., Brusca, G.J., 2003. Invertebrates, Second edition. Sinauer Associates, Inc., Sunderland, Massachusetts.

Cavalier-Smith, T., 1998. A revised six-kingdom system of life. Biol. Rev. 73, 203-266.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H.D., Schmidt- Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745-750.

Eernisse, D.J., Albert, J.S., Anderson, F.E., 1992. Annelida and Arthropoda are not sister taxa: a phylogenetic analysis of spiralian metazoan morphology. Syst. Biol. 41, 305-330.

Egger, B., Rieger, R.M., 2013. Plathelminthes, Plattwürmer. In: Westheide, W., Rieger, G. (Eds.), Spezielle Zoologie - Teil 1: Einzeller und Wirbellose Tiere. Springer Verlag, Berlin, Heidelberg, pp. 186-228.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783-791.

Funch, P., Sorensen, M.V., Obst, M., 2005. On the phylogenetic position of Rotifera – have we come any further? Hydrobiologia 546, 11–28.

Gardiner, S.L., 1978. Fine structure of the ciliated epidermis on the tentacles of *Owenia fusiformis* (Polychaeta, Oweniidae). Zoomorphologie 91, 37–48.

Giere, O., 2009. Meiobenthology - The microscopic motile fauna of aquatic sediments. Springer-Verlag, Berlin Heidelberg.

Giribet, G., Distel, D.L., Polz, M., Sterrer, W., Wheeler, W.C., 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.

Giribet, G., Sørensen, M.V., Funch, P., Kristensen, R.M., Sterrer, W., 2004. Investigations into the phylogenetic position of Micrognathozoa using four molecular loci. Cladistics 20, 1–13.

Golombek, A., Tobergte, S., Nesnidal, M.P., Purschke, G., Struck, T.H., 2013. Mitochondrial genomes to the rescue – Diurodrilidae in the myzostomid trap. Mol. Phylogenet. Evol. 68, 312–326.

Halanych, K.M., Janosik, A.M., 2006. A review of molecular markers used for Annelid phylogenetics. Integr. Comp. Biol. 46, 533-543.

Hankeln, T., Wey-Fabrizius, A.R., Herlyn, H., Witek, A., Weber, M., Nesnidal, M.P., Struck, T.H., 2014. Phylogeny of platyzoan taxa based on molecular data. In: Wägele, J.W., Bartolomaeus, T. (Eds.), Deep Metazoan Phylogeny: The backbone of the Tree of Life – New insights from analyses of molecules, morphology, and theory of data analysis. De Gruyter, Berlin, pp. 105-125.

Haszprunar, G., 1996. Plathelminthes and Plathelminthomorpha - paraphyletic taxa. J. Zool. Syst. Evol. Res. 34, 41-48.

Hausen, H., 2005. Comparative structure of the epidermis in polychaetes (Annelida). Hydrobiologia V535-536, 25-35.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E.G., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W., 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. R. Soc. B: Biol. Sci. 276, 4261-4270.

Herlyn, H., Ehlers, U., 1997. Ultrastructure and function of the pharynx of *Gnathostomula paradoxa* (Gnathostomulida). Zoomorphology 117, 135-145.

Jenner, R.A., 2004. Towards a phylogeny of the Metazoa: evaluating alternative phylogenetic positions of Platyhelminthes, Nemertea, and Gnathostomulida, with a critical reappraisal of cladistic characters. Cont. Zool. 73, 3-163.

Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., Stadler, P.F., 2012. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 40, 2833-2845.

Katoh, K., Kuma, K.-i., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518.

Knauss, E.B., 1979. Indication of an anal pore in Gnathostomulida. Zool. Scr. 8, 181–186.

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B., Halanych, K.M., 2011. Phylogenomics reveals deep molluscan relationships. Nature 477, 452–456.

Kristensen, R.M., 2002. An introduction to Loricifera, Cycliophora, and Micrognathozoa. Integr. Comp. Biol. 42, 641-651.

Kristensen, R.M., Funch, P., 2000. Micrognathozoa: a new class with complicated jaws like those of Rotifera and Gnathostomulidae. J. Morphol. 246, 1-49.

Kristensen, R.M., Nørrevang, A., 1977. On the fine structure of *Rastrognathia macrostoma* gen. et sp. n. placed in Rastrognathiidae fam. n. (Gnathostomulida). Zool. Scr. 6, 27–41.

Kück, P., Meusemann, K., 2010. FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. 56, 1115-1118.

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front. Zool. 7, 10.

Kück, P., Struck, T.H., 2014. BaCoCa - A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. 70, 94–98.

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

Meglitsch, P.A., Schram, F.R., 1991. Invertebrate Zoology. Oxford University Press, Oxford.

Melone, G., Ricci, C., Segers, H., Wallace, R.L., 1998. Phylogenetic relationships of phylum Rotifera with emphasis on the families of Bdelloidea. Hydrobiologia 387/388, 101-107.

Misof, B., Misof, K., 2009. A Monte Carlo Approach Successfully Identifies Randomness in Multiple Sequence Alignments : A More Objective Means of Data Exclusion. Syst. Biol. 58, 21-34.

Nicholas, K.B., Nicholas, H.B.j., 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author.

Nielsen, C., 2001. Animal Evolution. Interrelationships of the Living phyla. Second Edition. Oxford University Press, Oxford.

Nielsen, C., 2002. The phylogenetic position of Entoprocta, Ectoprocta, Phoronida, and Brachiopoda. Integr. Comp. Biol. 42, 685–691.

Nielsen, C., 2012. Animal Evolution - Interrelationships of the Living Phyla. Oxford University Press Inc., New York.

Paps, J., Baguna, J., Riutort, M., 2009a. Bilaterian phylogeny: A broad sampling of 13 nuclear genes provides a new Lophotrochozoa phylogeny and supports a paraphyletic basal Acoelomorpha. Mol. Biol. Evol. 26, 2397-2406.

Paps, J., Baguñà, J., Riutort, M., 2009b. Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. Proc. R. Soc. B: Biol. Sci. 276, 1245-1254.

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

Riedl, R.J., 1969. Gnathostomulida from America. Science 163, 445–462.
Rieger, R.M., Tyler, S., 1995. Sister-group relationship of Gnathostomulida and Rotifera- Acanthocephala. Invertebr. Biol. 114, 186-188.
Rouse, G.W., Pleijel, F., 2001. Polychaetes. University Press, Oxford.
Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686-W689.
Scholtz, G., 2010. Deconstructing morphology. Acta Zool. 91, 44-63.
Schram, F.R., Ellis, W.N., 1994. Metazoan relationships: a rebuttal. Cladistics 10, 331- 337.
Smith, P.R., Ruppert, E.E., 1988. Nephridia. In: Westheide, W., Hermans, C.O. (Eds.), The ultrastructure of Polychaeta. Gustav Fischer Verlag, Stuttgart, pp. 231-262.
Smith, S., Wilson, N., Goetz, F., Feehery, C., Andrade, S., Rouse, G., Giribet, G., Dunn, C., 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480, 364 - 367.
Smith, S.A., Dunn, C.W., 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24, 715-716.
Sørensen, M.V., 2000. An SEM study of the jaws of *Haplognathia rosea* and *Rastrognathia macrostoma* (Gnathostomulida), with a preliminary comparison with the rotiferan trophi. Acta Zool. 81, 9-16.
Sørensen, M.V., 2002. Phylogeny and jaw evolution in Gnathostomulida, with a cladistic analysis of the genera. Zool. Scr. 31, 461-480.
Sørensen, M.V., 2003. Further structures in the jaw apparatus of *Limnognathia maerski* (Micrognathozoa), with notes on the phylogeny of the Gnathifera. J. Morphol. 255, 131-145.
Sørensen, M.V., Funch, P., Willerslev, E., Hansen, A.J., Olesen, J., 2000. On the phylogeny of the metazoa in the light of Cycliophora and Micrognathozoa. Zool. Anz. 239, 297-318.
Sørensen, M.V., Giribet, G., 2006. A modern approach to rotiferan phylogeny: Combining morphological and molecular data. Mol. Phylogenet. Evol. 40, 585-608.
Sørensen, M.V., Sterrer, W., 2002. New characters in the gnathostomulid mouth parts revealed by scanning electron microscopy. J. Morphol. 253, 310-334.
Sørensen, M.V., Sterrer, W., Giribet, G., 2006. Gnathostomulid phylogeny inferred from a combined approach of four molecular loci and morphology. Cladistics 22, 32-58.
Sørensen, M.V., Tyler, S., Hoge, M.D., Funch, P., 2003. Organization of pharyngeal hard parts and musculature in *Gnathostomula armata* (Gnathostomulida: Gnathostomulidae). Can. J. Zool. 81, 1463–1147.
Stamatakis, A., 2006. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Sterrer, W., Mainitz, M., Rieger, R.M., 1985. Gnathostomulida: enigmatic as ever. In: Conway Morris, S., George, J.D., Gibson, R.H., Platt, M. (Eds.), The Origin and Relationships of Lower Invertebrates. Clarendon Press, Oxford, pp. 181–191.
Struck, T.H., 2014. TreSpEx – Detection of misleading signal in phylogenetic reconstructions based on tree information. Evolutionary Bioinformatics 10, 51-67.
Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., Bleidorn, C., 2011. Phylogenomic analyses unravel annelid evolution. Nature 471, 95–98.
Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kück, P., Herlyn, H., Hankeln, T., 2014. Platyzoan paraphyly based on phylogenomic data supports a non-coelomate ancestry of Spiralia. Mol. Biol. Evol. 31, 1833-1849.
Todaro, M.A., Telford, M.J., Lockyer, A.E., Littlewood, D.T.J., 2006. Interrelationships of the Gastrotricha and their place among the Metazoa inferred from 18S rRNA genes. Zool. Scr. 35, 251-259.
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 35, W71-W74.
Vallès, Y., Boore, J.L., 2006. Lophotrochozoan mitochondrial genomes. Integr. Comp. Biol. 46, 544-557.
Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C., Struck, T.H., 2014. Illuminating the base of the annelid tree using transcriptomics. Mol. Biol. Evol. 31, 1391-1401.
Westheide, W., 1984. The concept of reproduction in polychaetes with small body size: adaptations in interstitial species. In: Pfannenstiel, H.-D., Fischer, A. (Eds.), Polychaete Reproduction. Fortschritte der Zoologie, Stuttgart, pp. 265-287.
Westheide, W., 1986. The nephridia of the interstitial polychaete *Hesionides arenaria* and their phylogenetic significance (Polychaeta, Hesionidae). Zoomorphology 106, 35-43.
Westheide, W., 1997. The direction of evolution within the Polychaeta. J. Nat. Hist. 31, 1-15.
Wey-Fabrizius, A.R., Herlyn, H., Rieger, B., Rosenkranz, D., Witek, A., Welch, D.B.M., Ebersberger, I., Hankeln, T., 2014. Transcriptome data reveal syndermatan relationships and suggest the evolution of endoparasitism in Acanthocephala via an epizoic stage. PLoS ONE 9, e88618.
Witek, A., Herlyn, H., Ebersberger, I., Mark Welch, D.B., Hankeln, T., 2009. Support for the monophyletic origin of Gnathifera from phylogenomics. Mol. Phylogenet. Evol. 53, 1037-1041.

Worsaae, K., Rouse, G.W., 2008. Is *Diurodrilus* an annelid? J. Morphol. 269, 1426-1455.
Zhong, M., Hansen, B., Nesnidal, M.P., Golombek, A., Halanych, K.M., Struck, T.H., 2011. Detecting the symplesiomorphy trap: A multigene phylogenetic analysis for terebelliform annelids. BMC Evol. Biol. 11, 369.
Zrzavy, J., 2003. Gastrotricha and metazoan phylogeny. Zool. Scr. 32, 61-81.
Zrzavy, J., Hypsa, V., Tietz, D.F., 2001. Myzostomida Are Not Annelids: Molecular and Morphological Support for a Clade of Animals with Anterior Sperm Flagella. Cladistics 17, 170-198.
Zrzavy, 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.

5.5. Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids

Min Zhong[1], Benjamin Hansen[2], Maximilian Nesnidal[2], Anja Golombek[2], Kenneth M. Halanych[1], Torsten H. Struck[2],†

[1]Auburn University, Department of Biological Sciences, 101 Life Science Building, Auburn, AL 36849, USA
[2]University of Osnabrück, FB05 Biology/Chemistry, AG Zoology, Barbarastr. 11, 49069 Osnabrück, Germany

† Corresponding author
E-Mail addresses:
MZ: zhongmi@auburn.edu
BH: Ben.Hansen@web.de
MN: nesnidal@gmail.com
AG: A.Golombek@gmx.de
KMH: ken@auburn.edu
THS: struck@biologie.uni-osnabrueck.de

Abstract

Background
For phylogenetic reconstructions, conflict in signal is a potential problem for tree reconstruction. For instance, molecular data from different cellular components, such as the mitochondrion and nucleus, may be inconsistent with each other. Mammalian studies provide one such case of conflict where mitochondrial data, which display compositional biases, support the Marsupionta hypothesis, but nuclear data confirm the Theria hypothesis. Most observations of compositional biases in tree reconstruction have focused on lineages with different composition than the majority of the lineages under analysis. However in some situations, the position of taxa that lack compositional bias may be influenced rather than the position of taxa that possess compositional bias. This situation is due to apparent symplesiomorphic characters and known as "the symplesiomorphy trap".

Results
Herein, we report an example of the sympleisomorphy trap and how to detect it. Worms within Terebelliformia (sensu Rouse & Pleijel 2001) are mainly tube-dwelling annelids comprising five 'families': Alvinellidae, Ampharetidae, Terebellidae, Trichobranchidae and Pectinariidae. Using mitochondrial genomic data, as well as data from the nuclear 18S, 28S rDNA and elongation factor-1α genes, we revealed incongruence between mitochondrial and nuclear data regarding the placement of Trichobranchidae. Mitochondrial data favored a sister relationship between Terebellidae and Trichobranchidae, but nuclear data placed Trichobranchidae as sister to an Ampharetidae/Alvinellidae clade. Both positions have been proposed based on morphological data.

Conclusions
Our investigation revealed that mitochondrial data of Ampharetidae and Alvinellidae exhibited strong compositional biases. However, these biases resulted in a misplacement of Trichobranchidae, rather than Alvinellidae and Ampharetidae. Herein, we document that Trichobranchidae was apparently caught in the symplesiomorphy trap suggesting that in certain situations even homologies can be misleading.

Background

The amount of data used in phylogenetic reconstructions has been steadily increasing during the past decade [e.g., 1, 2-4], and phylogenies based on multiple datasets (i.e., partitions) are now common. However, analyses based on different partitions do not always result in congruent phylogenetic reconstructions. Molecular evolutionary events such as gene duplication, horizontal gene transfer, heterotachy, gene extinction, long-branch attraction, saturation and model misspecifications can cause inferred gene trees to differ from species trees. For example, incongruence regarding phylogenetic placement of taxa can occur between mitochondrial and nuclear data [e.g., 5]. In the case of mammals, mitochondrial data strongly support the Marsupionta hypothesis placing Marsupialia as sister to Monotremata (Fig. 1A) [6-11], whereas the Theria hypothesis, which places Marsupialia with Placentalia, has been strongly supported by both morphological and nuclear data [e.g., 12, 13, 14]. Phillips and Penny [15] showed that strong compositional biases in pyrimidine and purine frequencies in mitochondrial genomes of Marsupialia and Monotremata provided support for the Marsupionta hypothesis. However, both partitioning the dataset and to a lesser degree RY coding were able to effectively minimize artificial signal. In general, taxa affected by biases such as increased substitutions rates, heterotachy, etc., are the ones misplaced in phylogenetic analyses. However, biases may also influence the placement of unbiased taxa. In the case of the symplesiomorphy trap [16], a paraphyletic assemblage of taxa is grouped together as monophyletic based on the possession of symplesiomorphic characters, which are mistakenly assumed to be apomorphic. The symplesiomorphy trap has been characterized as a special class of long-branch attraction by Wägele & Mayer [17].

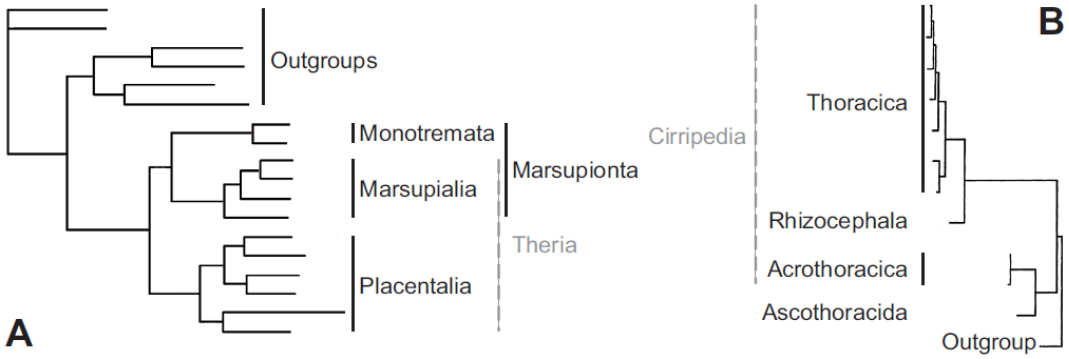


Fig. 1: Examples of misplacements.
(A) Marsupialia within Mammalia based on mitochondrial data [modified from 15] and (B) Ascothoracida within Cirripedia [modified from 74]. Only more inclusive taxonomic units are indicated for reasons of simplicity.

This problem is common for morphological data and several instances are known. One well-known annelid example is the position of Clitellata as sister to Polychaeta due to the lack of typical polychaete characters such as parapodia and nuchal organs [18]. However, molecular data clearly place Clitellata within polychaetes [e.g., 2, 3, 19]. In theory, the symplesiomorphy trap is not restricted to morphological data, but can also apply to sequence data [16]. However, studies addressing this problem in molecular data are scarce because detection of the trap is not straightforward. First, the misplaced taxa are not themselves affected by compositional biases or increased substitution rates. Second, support for monophyly of misplaced taxa is based on apomorphies for a higher taxonomic unit and hence not artificial. Third, knowledge of the 'true' phylogeny is needed to directly detect the symplesiomorphy trap. Typically, detection of the trap occurs indirectly by excluding other possibilities of incongruence and revealing characteristic signatures in the data. For example, Wägele and Mayer's [17] study showed that misplacement of Acrothoracica barnacles in a 18S parsimony analysis was due to

symplesiomorphic characters shared exclusively by Ascothoracida (a non-barnacle outgroup) and Acrothoracica (Fig. 1B). These characters overwhelmed the phylogenetic signal for the monophyly of Cirripedia. This phenomenon is known as the symplesiomorphy trap.
Here we report another instance of the symplesiomorphy trap in molecular data discovered while examining Terebelliformia (Annelida) phylogeny. Terebelliform worms [sensu 20] are typically tube-dwelling annelids, found in diverse marine habitats, including intertidal, deep-sea and even hydrothermal vent areas. Terebelliformia include about 800 species within five 'families': Alvinellidae, Ampharetidae, Terebellidae, Trichobranchidae and Pectinariidae [20-22]. Based on thorough investigations using data partitioning, topology tests, removal and addition of taxa, spectral analyses, detection of compositional biases, models of non-stationary sequence evolution, and recoding of characters, we were able to pinpoint the source of the incongruence between mitochondrial and nuclear data and relate it to the symplesiomorphy trap. Ampharetidae and Alvinellidae exhibit strong compositional biases in their mitochondrial genomes. However, these biases affect placement of Trichobranchidae and Terebellidae rather than Ampharetidae and Alvinellidae.

Methods

Sample and Data Collection

Table 1 lists taxa, gene sequences, GenBank accession numbers and sample locations used in this study. Upon collection, tissue samples were preserved in >70% non-denatured ethanol or frozen at -80oC. Genomic DNA was extracted using the DNeasy Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Mitochondrial genomes were amplified following Zhong et al. [23] in four overlapping segments using species-specific primers (for more details see Additional File 1). Amplification and sequencing of nuclear 18S and 28S genes was carried out using protocols described by Struck et al. [24]. Presence of PCR products were confirmed on a 1% agarose gel and purified with the QIAquick PCR Purification or QIAquick Gel Extraction kit (Qiagen, Hilden, Germany). When necessary, PCR products were size-selected on agarose gels and/or cloned using pGEM®-T Easy Vector System (Promega, Madison, WI, USA) or StrataCloneTM PCR Cloning Kit (Stratagene, La Jolla, CA, USA). A CEQTM 8000 Genetic Analysis System (Beckman Coulter, Fullerton, CA, USA) or ABI Prism 377 Automatic Sequencer (Perkin Elmer, Shelton, CT, USA) was used for bidirectional sequencing of all PCR products.

Genomic Assembly and Gene Identification

Sequences were edited and aligned using DNASTAR™ Lasergene programs SeqMan and MegAlign [25]. Protein-coding genes and ribosomal RNA genes were identified by BLAST [26]. All tRNA genes were identified using tRNAscan-SE web server [http://lowelab.ucsc.edu/tRNAscan-SE/, 27] under default settings and source = "mito/chloroplast", or by hand based on their potential secondary structures and anticodon sequences.

Datasets

Datasets consisted of mitochondrial and nuclear data. All alignments are available at TreeBASE (www.treebase.org). Seventeen available annelid mitochondrial genomes with about 50% coverage or greater were used for the phylogenetic analyses (Table 1). The alignment of Zhong et al. [23] was employed with the addition of *Nephtys* sp.*, Pectinaria gouldi, Paralvinella sulfincola* and *Auchenoplax crinita*. Because we were interested in relationships within Terebelliformia, we deleted the mitochondrial data of *Katharina* (Mollusca) and *Terebratalia* (Brachiopoda) and used all other annelids as outgroup taxa.

Both nucleotide and amino acid datasets were created for mitochondrial phylogenetic analyses. In the nucleotide dataset, all protein-coding genes (except for *atp6*, *atp8* and *nad6* genes which exhibit high variability) and the two rRNA genes (*mLSU* and *mSSU*) were included. Clustal X [28] under default settings was used to align rRNA genes. Gblocks 0.91b [29] was used to identify ambiguous aligned regions in the rRNA genes. These regions and the 3rd positions of protein-coding genes, which are saturated with substitutions for family7 level analyses, were excluded from the analyses with the aid of MacClade4.08 [30] and Se-Al v2.0a11 [31]. The amino acid dataset was created from the aligned nucleotide dataset by translation of protein-coding genes with the *Drosophila* mitochondrial genetic code and exclusion of rRNA genes. The mitochondrial nucleotide and amino acid datasets comprised 6,287 and 2,990 positions, respectively.

Table 1: Taxa used in phylogenetic analyses with 17 taxa.
Accession numbers of determined sequences in bold. For locality information on available data see original source. mtDNA = mitochondrial genome; EF1α = elongation factor 1α.

Taxon	Species	mtDNA	18S rDNA	28S rDNA	EF1α	Locality
Terebelliformia						
Ampharetidae						
	Eclysippe vanelli	EU239687	**JN936467**	**JN936489**		63°30.84'N/ 10°25.01'E Storgrunnen (Norway)
	Auchenoplax crinite	**FJ976041**	DQ790077	DQ790026	DQ813352	39°53.88'N/ 69°39.64'W Southern New England (USA)
	Pectinaria gouldi	**FJ976040**	DQ790091		DQ790054	41°37.91'N/ 70°53.34'W Egypt Lane, Fairhaven, MA (USA)
Pectinariidae						
	Pectinaria koreni				DQ813388	
Terebellidae						
	Pista cristata	EU239688	AY611461	DQ790057	DQ813391	
Trichobranchidae						
	Terebellides stroemi	EU236701	DQ790094	DQ790066		
	Terebellides sp.				DQ813404	
Alvinellidae						
	Paralvinella sulfincola	**FJ976042**	**JN936461**			47°57.001'N/ 129°05.851'W Juan de Fuca (Canada)
	Paralvinella palmiformis			**JN936479**		47°56.947'N/ 129°05.878'W Juan de Fuca (Canada)
	Paralvinella hessleri				DQ813385	
Outgroups						
Siboglinidae						
	Galathealinum brachiosum	AF178679	AF168738			
	Siboglinum fiordicum			DQ790061	DQ813398	
	Riftia pachyptila	AY741662	AF168739	Z21534	DQ813394	
Clitellata						
	Helobdella robusta	AF178680				
	Helobdella triserialis	AY962435				
	Hirudo medicinalis			AY364866	U90063	

Table 1 (continued)

Taxon	Species	mtDNA	18S rDNA	28S rDNA	EF1α	Locality
	Lumbricus terrestris	NC_001673	AJ272183			
	Lumbricus sp.			DQ790041	DQ813373	
Nereididae						
	Platynereis dumerilii	NC_000931	EF117897			
	Nereis succinea			AY210464		
	Nereis virens				U90064	
Echiura						
	Urechis caupo	AY619711	AF342805		DQ813410	
	Arhynchite pugettensis			AY210455		
Nephtyidae						
	Nephtys longosetosa		DQ790082	DQ790042		
	Nephtys sp.	EU293739			DQ813376	
Sipuncula						
	Phascolopsis gouldi	AF374337	AF123306	AF342795	AF063421	
Orbiniidae						
	Orbinia latreilliid	AY961084	AY532355			
	Orbinia swani		DQ790087	DQ790048		
	Orbinia michaelseni				DQ813381	
	Scoloplos cf. *armiger*	DQ517436	AY340443	AY366515		
Maldanidae						
	Clymenella torquata	AY741661		DQ790030	DQ813356	
	Clymenura clypeata		AF448152			

Additionally, a combined data matrix was constructed with the addition of 18S, 28S and EF-1α sequences to the mitochondrial data for the above 17 taxa (Table 1). Because we employed data from GenBank and collected data in two different laboratories (Univ. of Osnabrück and Auburn Univ.), in some cases we concatenated data from as closely related species as possible to generate Operational Taxonomic Units (OTUs) with a more complete coverage (see Table 1). Sequences were aligned as above. Due to the addition of nuclear data, the combined datasets comprised 11,813 nucleotide and 3,331 amino acid positions. The amino acid dataset comprised only the protein-coding genes.

Moreover, we also constructed a nuclear dataset comprising only 18S, 28S and EF-1α sequences at the nucleotide level for these 17 taxa (Table 1). The nuclear dataset comprised 5,526 nucleotide positions. Analyses of nuclear ribosomal gene datasets were also based on 32 and 61 taxa to reveal if taxon sampling had a substantial impact on the phylogenetic reconstruction of the nuclear data. By comparison, taxon sampling was far more limited for mitochondrial genome sequences. Additional File 2 provides a summary of the construction of these datasets with more than 17 taxa.

Phylogenetic Analyses

Maximum likelihood (ML) and Bayesian inference (BI) approaches were employed for all mitochondrial, nuclear and combined datasets. For all nucleotide datasets with 17 taxa, ML analyses were performed in PAUP4.0b10 [32] with a GTR+Γ+I model as determined by Modeltest v3.7 based on the Akaike information criterion (AIC) [33, 34]. Heuristic searches were run with random-taxon addition (10 replicates) using Tree-Bisection-Reconnection (TBR) swapping. All model parameters used fixed values as determined by Modeltest v3.7. Bootstrap analyses employed 1,000 iterations using heuristic searches with 10 random taxa addition replicates. Partitioned ML analyses were conducted with RAxML 7.2.8 [35] using a GTR+Γ+I model for each individual gene and 200 bootstrap replicates followed by a best tree search.

Partitioned BI invoked independent substitution models for each gene in MrBayes version 3.1.2 [36] and ran for 5*106 (mitochondrial and nuclear) or 2*106 (combined) generations, respectively, with 2 runs of 4 chains (3 heated and 1 cold). Trees were sampled every 100 generations. The implemented diagnosis feature comparing the 2 runs by average standard deviation of split frequencies was determined every 10,000 generations. GTR+Γ+I models were selected under the AIC in MrModeltest [37, 38] for 18S and 28S rDNA, EF-1α, *cox1*, *cox2*, *cob*, *nad1*, *nad3*, and *nad4*, GTR+I models for both 12S and 16S rDNA, GTR+Γ model for *cox3*, and HKY+Γ model for *nad2, nad4L* and *nad5*. Convergence of -ln likelihood scores and tree length was determined using Tracer v1.4.1 [39] to identify the *burnin* point at which all estimated parameters reached equilibrium (*burnin* = 100 trees). The majority-rule consensus tree containing posterior probabilities (PP) was determined from the remaining trees. Additional File 2 provides a more detailed description of the analyses and results for the datasets with more than 17 taxa.

For both amino acid datasets (mitochondrial and combined data with 17 taxa), nonpartitioned and partitioned ML, and partitioned BI analyses were run. For ML analyses, model selection was performed in RAxML 7.2.8 [35] and the MtZOA+Γ+I+F model was chosen as the best-fitting one for both non-partitioned datasets. For individual genes, MtZOA+Γ+I models were selected for *cox1*, *cox2* (additionally +F), *cox3* and *cob*, and DAYHOFF+Γ+I for *nad1*, *nad2, nad3*, *nad4*, *nad4L*, *nad5* and EF-1α. Maximum likelihood searches were implemented with 200 bootstrap replicates using RAxML [35] followed by a ML tree search for both non-partitioned and partitioned ML analyses. For partitioned BI of amino acid datasets, the mixed amino acid substitution model option plus a Γ distribution and a proportion of invariant sites was assigned to each partition individually and unlinked in MrBayes v3.1.2. BI ran for 2*106 generations and trees sampled every 500 generations (*burnin* = 20 trees). In the mixed model option, a specific model is not specified *a priori*, but each model is chosen during the run based on its posterior probability.

Non-stationary sequence evolution

To analyze data in a non-stationary Bayesian framework, we used PHASE 2.0 [40] to allow usage of different compositional vectors along branches of the tree. As in stationary Bayesian inferences using MrBayes, we conducted partitioned analyses for nucleotide datasets with 17 taxa of both mitochondrial and nuclear data invoking previously mentioned substitution models for each gene (except that the proportion-of-invariant-sites parameter is not available in PHASE 2.0). We performed analyses based on 3, 6 or 9 different compositional vectors. For each number of compositional vectors, we ran 4 independent runs, with one cold chain each and different random seeds (i.e., 3, 11, 88, and 1000), in parallel. Each run ran for 12*106 generations and trees were sampled every 1,000 generations. The first 2*106 generations were discarded as *burnin* as convergence of -ln likelihood scores and tree length was indicated by Tracer v1.4.1[39].

Topology testing

To further understand congruence and incongruence in our datasets, the Approximately Unbiased (AU) topology test of CONSEL [41, 42] was employed to assess support for alternative hypotheses. More specifically under the ML criterion, AU tests compared the three possible terebelliform hypotheses with respect to incongruence for each possible combination of partitions in the 17-taxa case (i.e., 18S, 28S, mtDNA, 18S/28S, 18S/EF-1α, 18S/mtDNA, 28S/EF-1α, 28S/mtDNA, EF-1α/mtDNA, 18S/28S/EF-1α, 18S/28S/mtDNA, 18S/EF-1α/mtDNA, 28S/EF-1α/mtDNA, and 18S/28S/EF-1α/mtDNA). Based on initial results, the following hypotheses were tested: 1) Trichobranchidae as sister to Alvinellidae/Ampharetidae (TriAA), 2) Trichobranchidae as sister to Terebellidae (TriTer), and 3) Terebellidae as sister to Alvinellidae/Ampharetidae (TerAA). PAUP analyses were constrained to obtain only the best

trees congruent with the particular hypothesis. Settings for the analyses were as described above.

Spectral Analyses

We conducted spectral analyses to gain further insights into the support for specific bipartitions (or splits) [43, 44] because they have been useful in the detection of the symplesiomorphy trap [17]. A bipartition splits a set of OTUs into two groups. In the context of spectral analyses, we use the term *ingroup* (italicized here to distinguish its usage in spectral analyses from common systematic usage) to define the group of the bipartition we are interested in, and *outgroup* for the other group of that bipartition. For example, Trichobranchidae, Alvinellidae and Ampharetidae in one group of the bipartition, the *ingroup*, and all others including Terebellidae in the other, the *outgroup*, would be congruent with the TriAA hypothesis. To calculate and visualize the bipartition support, we used Splits Analyses MethodS [SAMS, 17] and Microsoft Excel for mitochondrial, nuclear and combined datasets with 17 taxa. SAMS is a split-decomposition tool that does not require Hadamard conjugations. Hence, there is no need to consider the complete split space. SAMS differentiates support for a bipartition into three categories: 1) binary, both groups exhibit only one character state each, but different from each other; 2) noisy *outgroup* (i.e., while the *ingroup* exhibits only one state the *outgroup* exhibits more than one state, though a majority state within the group can still be identified); 3) noisy *ingroup* and *outgroup* [17]. Because we were only interested in bipartitions regarding relationships within Terebelliformia, we only retrieved bipartitions from the results that were relevant regarding these relationships. The PERL script to retrieve these bipartitions is available from THS upon request.

Determination of Compositional Biases

We also analyzed our nuclear and mitochondrial datasets for compositional biases, which can mislead phylogenetic analyses [e.g., 15, 45-53]. First, we employed relative composition variability (RCV), which is the average variability in composition between taxa for a dataset [15]. Phillips and Penny [15] used absolute numbers of nucleotide occurrence for calculation of RCV. However, this means that the RCV value does not only reflect composition variability, but also sequence length variability in the dataset. Therefore, we created a measure of relative composition frequency variability (RCFV) by modifying the RCV calculation to use base frequencies instead of absolute numbers:

$$RCFV = \sum_{i=1}^{n} \frac{\left|\mu_{A_i} - \tilde{\mu}_A\right| + \left|\mu_{Ci} - \tilde{\mu}_C\right| + \left|\mu_{Gi} - \tilde{\mu}_G\right| + \left|\mu_{Ti} - \tilde{\mu}_T\right|}{n}$$

where μA*i* is the base frequency of A for the *i*th taxon and μ is the mean base frequency across *n* taxa. Besides the RCFV for complete datasets, we also report herein taxon-specific RCFV values (i.e., $\left(\left|\mu_{Ai} - \tilde{\mu}_A\right| + \left|\mu_{Ci} - \tilde{\mu}_C\right| + \left|\mu_{G_i} - \bar{\mu}_G\right| + \left|\mu_{Ti} - \tilde{\mu}_T\right|\right) / n$), taxonspecific absolute deviations of each nucleotide (e.g.,$\left|\mu_{Ai} - \tilde{\mu}_A\right|$), and combinations of nucleotides (i.e. AT or GC and Y or R). Second, we determined different skew values to determine if strong biases between two nucleotide frequencies exist. Perna and Kocher [54] introduced the A-T and G-C skews for an individual strand of nucleic acids. Herein, we additionally propose A-G and C-T skews, because for mitochondrial genomes, major mutational biases are within purine and pyrimidine frequencies, respectively [55]. A-G and C-T skews for a taxon are calculated the same way as A-T and G-C skews are:

$$\text{A} - \text{G skew} = \frac{\mu_A - \mu_G}{\mu_A + \mu_G} \; ; \; \text{C} - \text{T skew} = \frac{\mu_C - \mu_T}{\mu_C + \mu_T}$$

Results

Phylogenetic Analyses

Mitochondrial datasets

ML and partitioned BI analyses of 17-taxa mitochondrial datasets based on either nucleotides or amino acids inferred identical topologies, with one exception, regarding terebelliform relationships with strong nodal support (Fig. 2b & Additional File 3). Monophyly of Terebelliformia is well supported (BS: 100 for nonpartitioned nucleotide (nNuc) and partitioned nucleotide (pNuc) analyses, 93 for nonpartitioned amino acid (nAA), and 94 for the partitioned amino acid (pAA) analyses; PP: 1.00 for both BI analyses). Mitochondrial datasets infer a sister relationship between Trichobranchidae and Terebellidae, the TriTer hypothesis (BS: 95 for nNuc, 100 for pNuc, 62 for nAA and 84 for pAA; PP: 1.00 for both). Furthermore, topology testing significantly rejected a sistergroup relationship of Trichobranchidae to Alvinellidae/Ampharetidae, the TriAA hypothesis ($p = 0.003$), as well as Terebellidae as sister to Alvinellidae/Ampharetidae, the TerAA hypothesis ($p = 0.028$). Two Ampharetidae taxa were close to Alvinellidae in the analyses of both mitochondrial datasets (BS: 100 for all four; PP: 1.00 for both). Pectinariidae was shown to be the basal lineage in Terebelliformia except in the partitioned ML analysis of the nucleotide dataset, which placed Pectinaridae as sister to Trichobranchidae/Terebellidae (BS: 72, data not shown).

Nuclear datasets

ML and partitioned BI of the 17-taxa, three-nuclear-gene (i.e., 18S, 28S and EF-1α) dataset inferred an identical topology with respect to terebelliform relationships (Fig. 2a). Interestingly, monophyly of Terebelliformia was not recovered as *Pectinaria gouldi* was placed as sister to the sipunculid *Phascolopsis gouldi*, albeit with weak support (Fig. 2a). The other four terebelliform taxa formed a clade with stronger nodal support (BS: 86 for nNuc, 100 for pNuc; PP 1.00) than in mitochondrial analyses (BS: 69 for nNuc, <50 for pNuc; PP: 0.92, Fig. 2b). As for the mitochondrial analyses, a sistergroup relationship of Alvinellidae and Ampharetidae is well corroborated (BS: 98 for nNuc, 99 for pNuc; PP: 1.00). Moreover, the TriAA hypothesis was supported (BS: 96 for nNuc, 92 for pNuc; PP: 1.00) and topology testing significantly rejects the alternative TriTer (favored by the mitochondrial data) and TerAA hypotheses ($p = 0.038$ and $p = 0.006$, respectively).

Combined datasets

Phylogenetic trees from combined analyses (Fig. 2c & Additional File 3) were similar to the ones from mitochondrial data (Fig. 2b) with differences occurring in outgroup relationships. Monophyly of Terebelliformia is significantly supported in these analyses (BS: 99 for nNuc, 100 for pNuc, 98 for nAA and 93 for pAA; PP: 1.00 for both; Fig. 2c, Additional File 3). Pectinariidae branched off first within terebelliforms (BS: 95 for nNuc, 100 for pNuc, 96 for nAA and 72 for pAA; PP: 1.00 for both). Alvinellidae was recovered as sister to Ampharetidae (BS: 100 for all four; PP: 1.00 for both). Trichobranchidae was placed as sister to Terebellidae, the TriTer hypothesis, in all analyses. However, bootstrap support for the TriTer hypothesis in the combined analyses was generally lower than in mtDNA alone analyses (83 in nNuc, 95 in pNuc, 41 in nAA, and 74 in pAA compared to 95, 100, 62, and 84, respectively; Fig. 2 & Additional File 3). Furthermore in contrast to the mitochondrial Nuc dataset, topology testing

did not significantly reject the alternative TriAA hypothesis favored by the nuclear dataset (p = 0.184), though the TerAA hypothesis is still significantly rejected (p = 0.012).

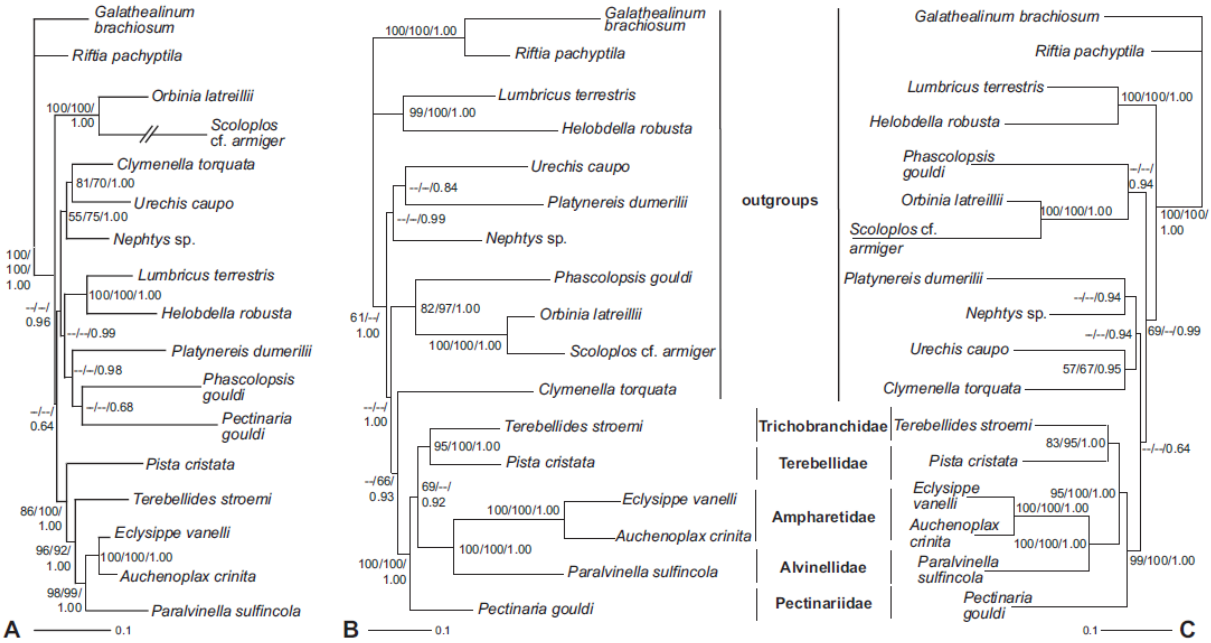


Fig. 2: Phylogenetic reconstructions using nuclear, mitochondrial and combined nucleotide datasets with 17 taxa. (A) Nuclear ML tree. The branch leading to *Scoloplos* cf. *armiger* was reduced by 75%. (B) Mitochondrial ML tree. (C) Combined ML tree. All trees represent identical topologies regarding terebelliform relationships for both ML and partitioned BI. Nodal support values are given at branches in the order: non-partitioned ML bootstrap, partitioned ML bootstrap and PP of the BI. A dash indicates < 50%.

Congruence and Incongruence between Partitions regarding Terebelliformia

Due to these results, we further explored conflict regarding the TriTer and TriAA hypotheses indicated by mtDNA (Fig. 2b) or nuclear partitions (Fig. 2a), respectively. Therefore, we conducted phylogenetic analyses and topology testing for all possible combinations of the four partitions (18S, 28S, EF-1α, mtDNA) when using 17 taxa. These analyses showed that when the mitochondrial data partition was added, the TriTer hypothesis was supported, whereas all possible combinations of the three nuclear genes, excluding mtDNA data, recovered the TriAA hypothesis. With an increasing amount of nuclear data (mitochondrial partition excluded) bootstrap support for the TriAA hypothesis steadily increased (black circles in Fig. 3a), while bootstrap support for the TriTer hypothesis remained low (grey circles in Fig. 3a). Furthermore, the p value of the AU test for the TriTer hypothesis decreased with an increasing amount of nuclear data from a non-significant value of 0.447 to a significant one of 0.041 (Fig. 3b, grey circles and trend line). On the other hand, in all datasets including mitochondrial data bootstrap support for the TriTer hypothesis was high, though it slightly decreased with an increasing amount of nuclear data (grey triangles in Fig. 3a), and, vice versa, the bootstrap support for the TriAA hypothesis was low, but slightly increased with increasing nuclear data (black triangles in Fig. 3a). However, as the proportion of nuclear data combined with mtDNA data increased, the p value of the AU test for the TriAA hypothesis became less significant (Fig. 3b, black triangles and trend line; p values change from 0.004 to 0.184). Comparatively and independent of the inclusion of mitochondrial data, the p value for the TerAA hypothesis decreased with an increasing amount of nuclear data (open triangles and circles in Fig. 3b). Hence, topology tests clearly revealed that nuclear data favor the TriAA hypothesis, whereas mitochondrial data favor the TriTer hypothesis.

Spectral Analyses

Spectral analyses revealed that 160 positions of the 17-taxon nuclear dataset support the TriAA hypothesis (Fig. 4a) recovered in the best tree (Fig. 2a). One hundred and five positions are consistent with the TriTer hypothesis favored by the mtDNA data and 91 with the TerAA hypothesis. This is congruent with the results of the topology tests based on the 17-taxon nuclear dataset, where the TriTer hypothesis had a higher p value than the TerAA hypothesis (0.038 > 0.006). However for the mitochondrial dataset with 17 taxa, similar numbers of positions, 103 and 102, support the TerAA and TriAA hypothesis, respectively. On the other hand, only 49 positions are consistent with the TriTer hypothesis, which was recovered by the best tree of the mitochondrial dataset (Fig. 2b).

Besides the number of positions, the quality of supporting positions is different for these three alternative hypotheses in both 17-taxon datasets. For the nuclear dataset, two binary positions support the TriAA hypothesis (black color in Fig. 4a) and no binary positions support the TriTer and TerAA hypotheses. In contrast, no binary positions are found to support any of the three hypotheses in the mitochondrial dataset. All other positions consistent with the TriAA or TerAA hypothesis are either noisy only in the *outgroup* (dark grey in Fig. 4) or in both *ingroup* and *outgroup* (light grey in Fig. 4), with more positions belonging to the latter class. Conversely, positions consistent with the TriTer hypothesis are exclusively based on a single class of positions, noisy in the *outgroup* only (Fig. 4).

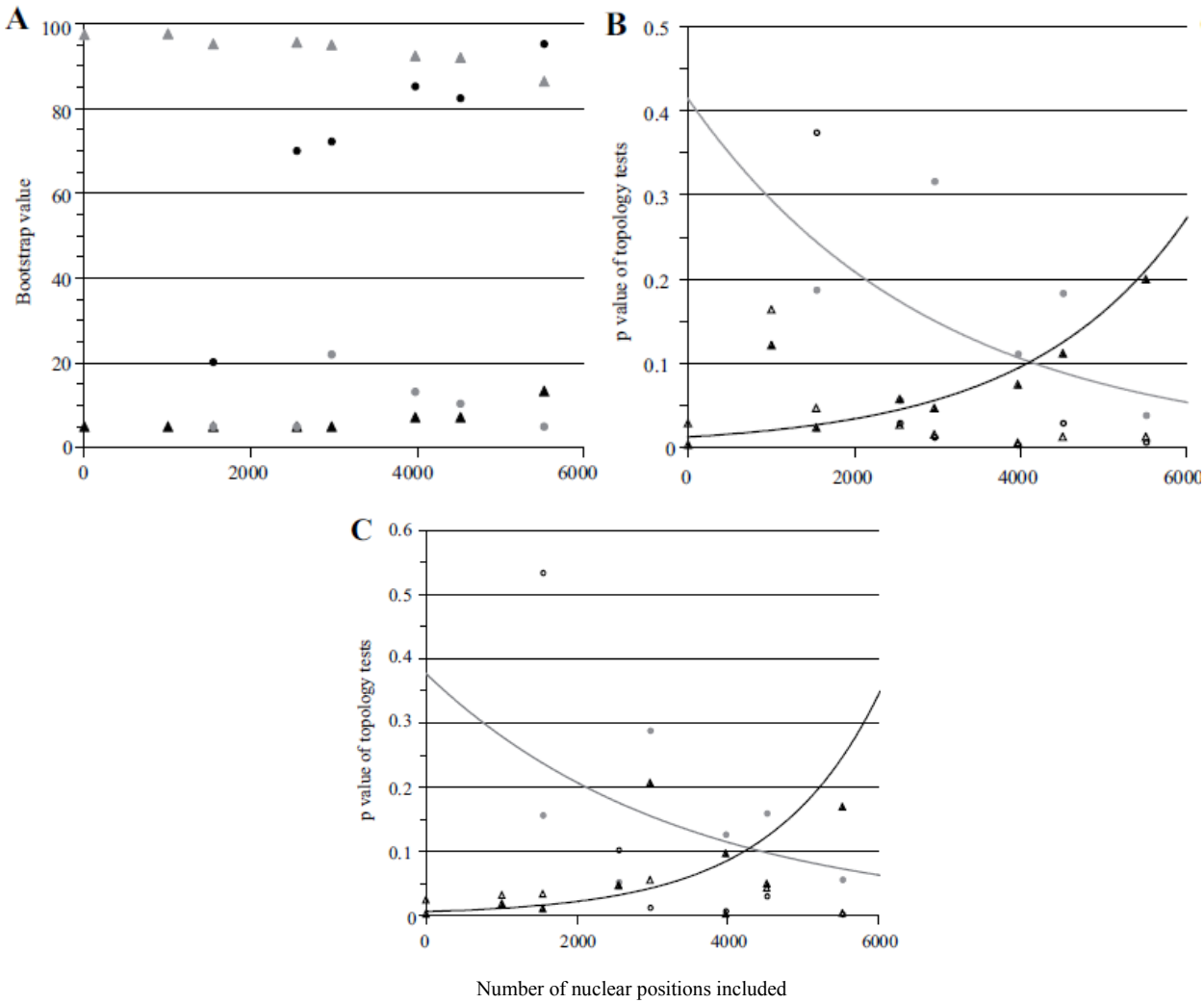


Fig. 3: Analyses evaluating incongruence of mitochondrial and nuclear data concerning placement of Trichobranchidae. (A) Bootstrap support. (B) Results of the topology tests against the best tree. (C) Same as B, but with the long-branched taxa *Pectinaria gouldi*, *Phascolopsis gouldi* and *Scoloplos* cf. *armiger* excluded from the analyses. Black symbols indicate TriAA, grey symbols the TriTer and open symbols the TerAA hypothesis. Circles stand for all possible combinations of only the nuclear partitions and triangles for mitochondrial data plus all possible combinations of the nuclear partitions.

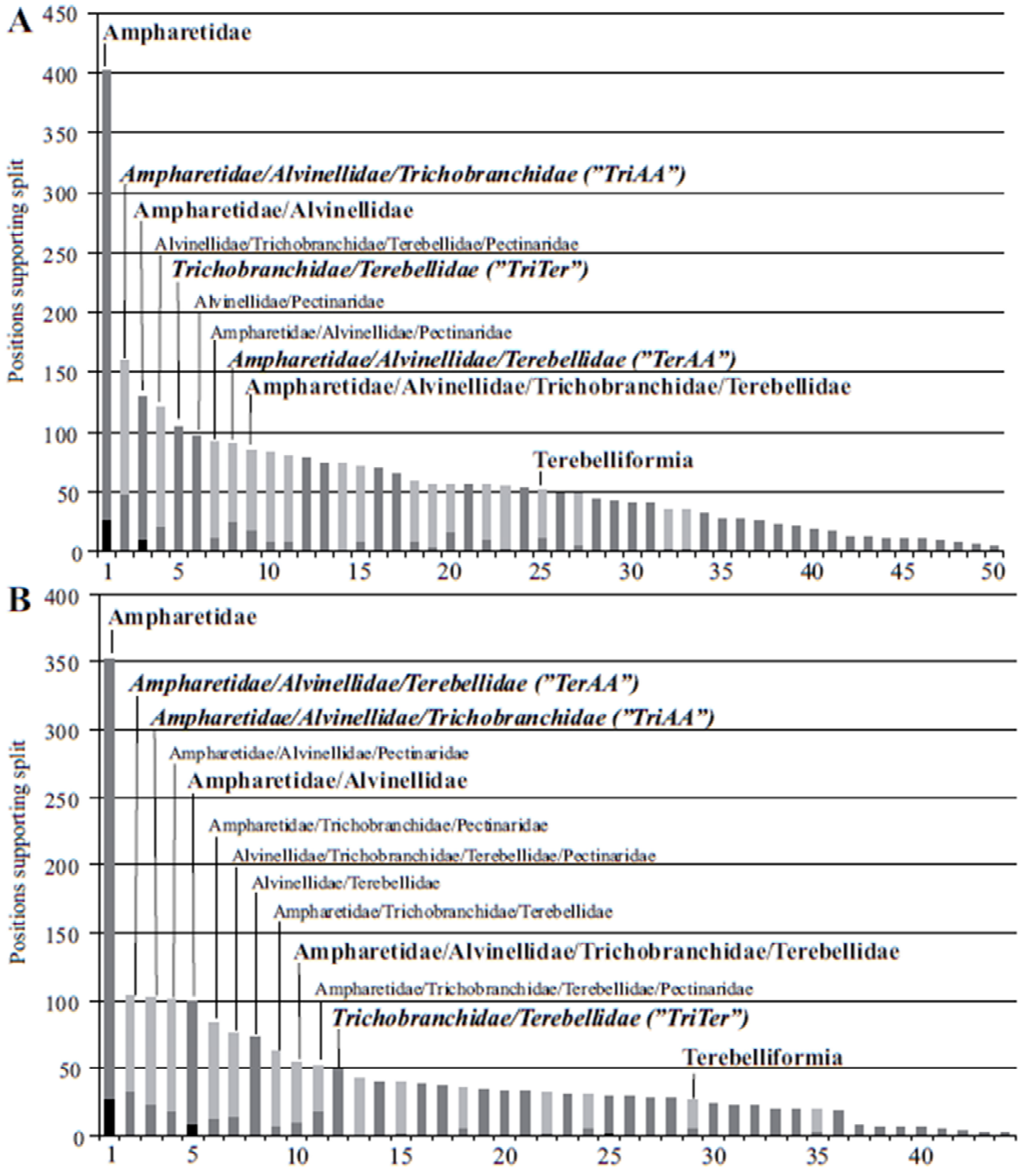


Fig. 4: Results of spectral analyses for all splits recovered by SAMS concerning terebelliform relationships. (A) Nuclear and (B) mitochondrial datasets with 17 taxa. Only supporting positions for the *ingroup* of the split are shown and not the *outgroup*. Names in bold at splits were recovered in our analyses. Additionally, the TriAA, TriTer and TerAA hypotheses are in italic. Black indicates binary positions, that is both groups exhibit only one character state each, but different from each other; dark grey noisy *outgroup* positions, while the *ingroup* exhibits only one state, the *outgroup* exhibits more than one state, though a majority state within the group can still be identified; light grey noisy *ingroup* and *outgroup* positions supporting a split.

Source of Incongruence

Based on analyses herein, placement of Trichobranchidae is incongruent between mitochondrial and nuclear data. To further investigate possible sources of incongruence with regards to Trichobranchidae placement, we examined two properties known to mislead placement of taxa, placement of the root and base composition heterogeneity.

Placement of the root

With respect to the relationships of Trichobranchidae, Terebellidae, Alvinellidae and Ampharetidae to each other, mitochondrial and nuclear partitions yield identical subtrees that were rooted differently (Fig. 5). Effects of longbranched outgroups and basal taxa misleading placement of the root have been long known [for review see 56]. *Pectinaria gouldi*, as well as *Phascolopsis gouldi*, exhibit long branches in nuclear rRNA data [19, 57 and see also Additional File 2]. However, Pectinariidae is placed as sister to the other terebelliform taxa and may influence placement of Trichobranchidae within the nuclear dataset (Fig. 2, Additional File 2). Nuclear data of *Scoloplos* cf. *armiger* also exhibited a long branch on the reconstructed topology (Fig. 2a). Therefore, we excluded these taxa (*Pectinaria gouldi*, *Phascolopsis gouldi*, *Scoloplos* cf. *armiger*) to examine the possibility of long-branch attraction, but found that they did not influence placement of the root or Trichobranchidae. All combinations of nuclear genes still favored the TriAA hypothesis, whereas the addition of the mitochondrial data always rendered Trichobranchidae being sister to Terebellidae in ML reconstructions. Correspondingly, results of topology tests are not altered substantially by excluding these three long branched taxa (compare Fig. 3c with Fig. 3b).

Poor taxon sampling can also influence taxon placement and rooting [58, 59]. As we could not easily increase the available number of mitochondrial genomes for Terebelliformia, we focused on adding more nuclear data and included 18 new 18S and 13 28S sequences for Terebelliformia and one cirratulid to the available data (Additional File 2). Phylogenetic analyses of this dataset comprising 32 taxa also recovered a sistergroup relationship of Trichobranchidae to Alvinellidae/Ampharetidae (BS: 80; PP: 0.95) within a monophyletic Terebelliformia. Additionally, the 61-taxon dataset based only on 18S rRNA data failed to provide resolution within Terebelliformia (Additional File 2); thus, neither exclusion of longbranched taxa nor an increased taxon sampling had an influence on the placement of the root for the nuclear data.

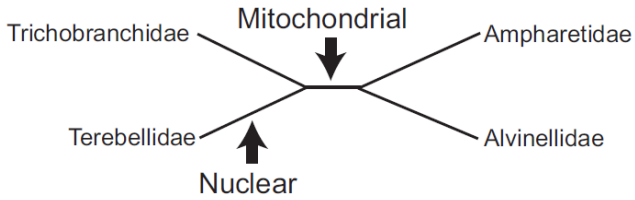


Fig. 5: The unrooted subtree of Trichobranchidae, Terebellidae, Alvinellida and Ampharetidae. Arrows indicate the position of the root by either nuclear or mitochondrial data.

Base composition

Evaluations of base composition heterogeneity showed a strong difference between nuclear and mitochondrial data. The RCFV value for mitochondrial data (0.0494) was much greater than for nuclear data (0.0159). Thus, mitochondrial data exhibit a stronger compositional heterogeneity. For mitochondrial data, taxon-specific RCFV values (Fig. 6a) showed that Alvinellidae, and especially Ampharetidae, had much higher values than the other terebelliforms or the average outgroup value indicating strong compositional biases in

Alvinellidae and Ampharetidae. No obvious biases were observed in nuclear data. Similar results were obtained for absolute deviations from mean frequency for individual nucleotides as well as combinations of nucleotides (Fig. 6b). For pyrimidines (cytosine and thymine), Ampharetidae and Alvinellidae deviated more from the mean than other terebelliform taxa. In addition, Ampharetidae also showed a much stronger deviation from the mean in guanine. Binning nucleotides as AT and GC did not alleviate these differences in deviation (and even made it more pronounced for Alvinellidae), but recoding pyrimidines (Y) and purines (R) reduced the biases between terebelliform taxa (Fig. 6b).

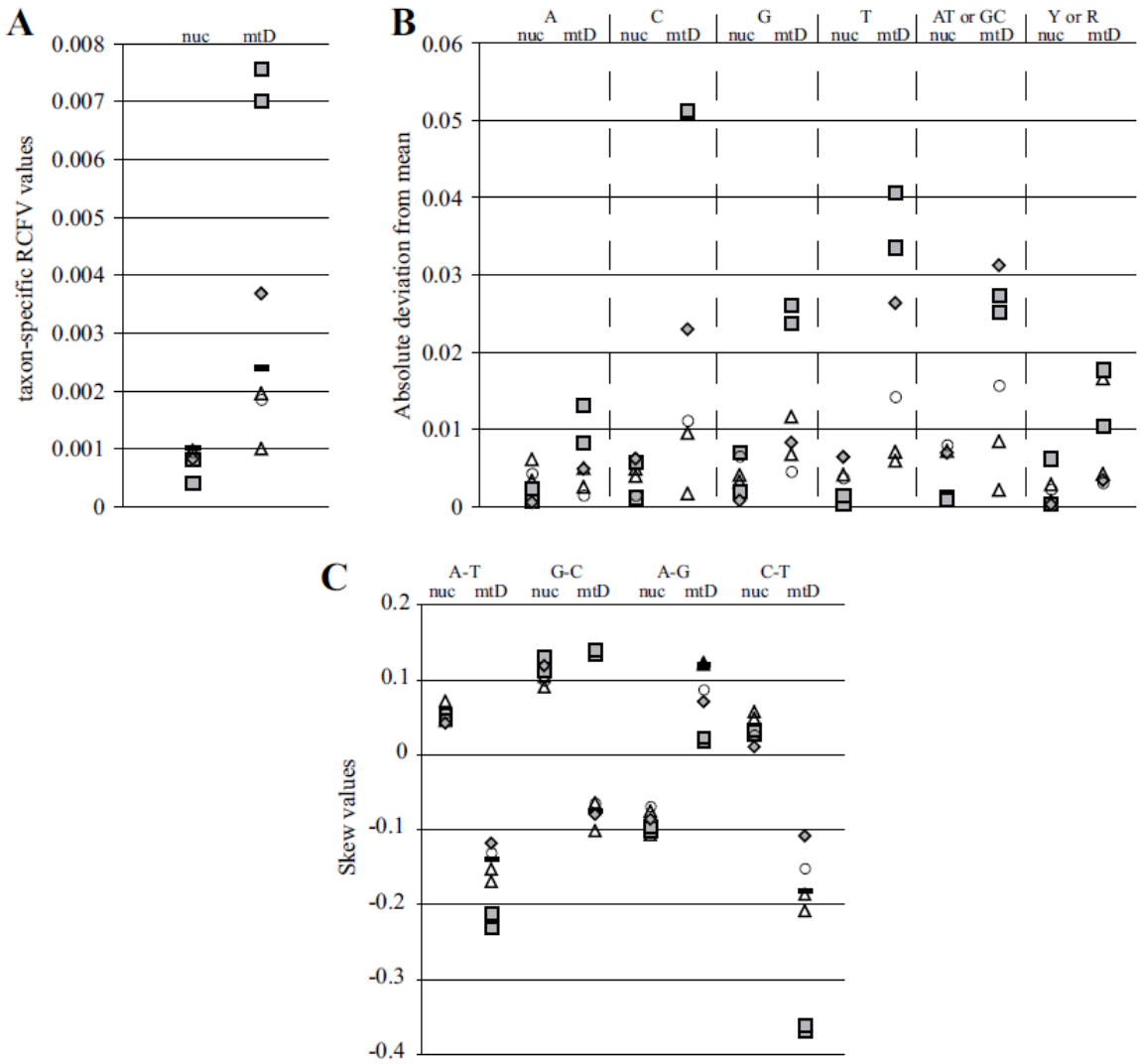


Fig. 6: Analyses of compositional heterogeneity in nuclear and mitochondrial datasets in the mitochondrial protein-coding genes. (A) Taxon-specific relative composition frequency variability (RCFV). (B) Absolute deviation from mean frequency for adenine (A), cytosine (C), guanine (G), and thymine (T) as well as combinations of adenine/thymine (AT) or guanine/cytosine (GC) and of pyrimidines (Y) or purines (R). Only one absolute value is provided for AT and GC or Y and R as only two character states are now present and any change in one state has the exact opposite negative or positive value in the other. (C) Skew values within the combinations adenine/thymine (A-T), guanine/cytosine (G-C), purines (A-G) and pyrimidines (C-T). Ampharetidae (grey squares), Alvinellidae (grey diamonds), Pectinariidae (open circles), Trichobranchidae and Terebellidae (both open triangles), mean values of outgroup taxa (black bar), nuclear (nuc), mitochondrial (mtD) .

Ampharetidae exhibited a strong G-C skew value towards guanine relative to cytosine (Fig. 6c). Moreover for mitochondrial data, C-T skews indicated that Ampharetidae was biased towards thymine, and Alvinellidae away from it, relative to other taxa. The same pattern could be observed in A-T skews driven by the differences in thymine frequencies. Thus, Ampharetidae and Alvinellidae showed strong-but opposite-biases in frequencies of pyrimidines, and Ampharetidae also a strong skew towards guanine. These evaluations were based on the mitochondrial dataset, we used for phylogenetic analyses (i.e., excluding 3rd positions), but examining either 3rd positions alone or with 3rd positions included resulted in similar patterns (Additional File 4). Codon usage reflected biases in base frequencies with deviations in Ampharetidae and Alvinellidae compared to the other taxa (Additional File 1).

Amelioration of Incongruence

Non-stationary sequence evolution

Using models of non-stationary sequence evolution has successfully ameliorated misleading effects of compositional biases in mitochondrial genomes of beetles [60]. Therefore, we also employed such models for both our mitochondrial and nuclear datasets using PHASE 2.0 [40]. For both datasets and each number of different compositional vectors, 4 independent chains starting from different random seeds failed to converge upon the same score indicating a structured tree-space with several local optima. Nonetheless for mitochondrial data, the majority-rule consensus topology derived from the best run (i.e, -lnL values) for each number of different compositional vectors (i.e., 3, 6, or 9) were identical except for the position of the outgroup taxon *Clymenella torquata* (Additional File 5). As before with mitochondrial data, Terebellidae and Trichobranchidae were sister to each other (PP: 1.00 for all three; Additional File 5). For nuclear data, the three topologies derived from the best runs invoking 3, 6 or 9 different vectors placed Trichobranchidae as sister to Alvinellidae/Ampharetidae (PP: 1.00 for all three; Additional File 5). Thus, using different compositional vectors along the branches did not reduce incongruence between datasets.

RY coding

For mitochondrial genomes, RY coding strategies can ameliorate biases within pyrimidines and purines, because they do not distinguish between transition or transversion classes [15, 61]. The best ML tree based on RY coding of the nuclear partition (Fig. 7) is similar to the ML tree using standard nucleotide coding (Fig. 2a; with the exception of *Scoloplos* cf. *armiger*/*Orbinia latreillii* placement). However, bootstrap support for Trichobranchidae as sister to Alvinellidae/Ampharetidae dropped.

In contrast, RY coding of the mitochondrial partition and combined dataset (inset in Fig. 7) yielded different ingroup relationships (see Figs. 2b & c for standard nucleotide coding) with Terebellidae as sister to Ampharetidae/Alvinellidae rather than Trichobranchidae. Notably, bootstrap support for this clade was below 50 in the analyses of both mitochondrial and combined data and all previous topology tests clearly rejected this relationship (Figs. 3b&c). Besides this difference in ingroup relationships, RY coding of mitochondrial and combined data also differed in several outgroup relationships.

Discussion

Biases in nucleotide frequencies influenced placement of Trichobranchidae and Terebellidae in both mitochondrial and combined analyses. Misplacement of these taxa is interesting because the taxa themselves did not exhibit compositional biases, but Alvinellidae and Ampharetidae

biases influenced their placement. This misplacement was apparently due to biases in Ampharetidae and Alvinellidae and can be related to the "symplesiomorphy trap" for which few molecular examples have been elucidated [16, 17]. In the Cirripedia example by Wägele and Mayer [17] (Fig. 1B), Acrothoracica and Ascothoracida grouped together due to symplesiomorphic characters because of the long branch uniting the remaining Cirripedia. Though no long branches could be observed in our analyses based on mitochondrial data regarding terebelliform taxa, biases in base composition and codon usage detected in Ampharetidae and Alvinellidae pointing in opposite directions appear to have had a similar effect. These directional biases affected nucleotides in all three coding positions of mitochondrial genes in Ampharetidae and Alvinellidae presumably due to differences in substitution rate or pattern.

In our case the symplesiomorphy trap appears to have misrooted a terebelliform subtree rendering a paraphyletic assemblage as a monophyletic group. The misinterpretation appears due to basal homologies, or symplesiomorphies, rather than an artificial signal due to homoplasy (e.g., long branches). First of all, though Alvinellidae and Ampharetidae are affected by opposite biases in mitochondrial nucleotide frequencies their sistergroup relationship, which is independently confirmed by the nuclear data, is still strongly supported by mitochondrial data as judged by bootstrap and spectral analyses. Hence, these two taxa appear unaffected by the opposite biases. Second, we could exclude that the nuclear partition is affected by an artificial signal; the nuclear data exhibited no biases with respect to terebelliform taxa. The root of the subtree comprising Terebellidae, Trichobranchidae and Ampharetidae/Alvinellidae, which was supported by all our analyses as well as several previous ones [e.g., 19, 57, 62], was not placed differently by the inclusion or exclusion of taxa [56]. Moreover, the spectral analysis of the nuclear partition is in agreement with the reconstructed nodes regarding the relations of these three taxa to each other. The number of supporting positions in the spectral analysis is in agreement with support by bootstrap and topology test p values for nuclear data. Third and contrasting with the nuclear data, the spectral analyses of the mitochondrial data are not congruent with tree reconstructions. Whereas the TriTer hypothesis was recovered in all best trees that included mtDNA data and was strongly supported by bootstrap and topology test results, spectral analyses revealed that this hypothesis was consistent with the fewest numbers of positions in the mitochondrial data. Using mitochondrial data, these characters overwhelmed the larger numbers of positions supporting the alternative placement of Trichobranchidae.

In the case of the symplesiomorphy trap, the phylogenetic signal for a certain relationship can be eroded along internal branches leading to subgroups without affecting the subgroups themselves. In the Cirripedia example [17], this erosion occurred along the branch leading to all Cirripedia but Acrothoracica (Fig. 1B). In our case, there are more possibilities; the branch leading to Ampharetidae/Alvinellidae as well as the branches within this clade could be relevant. For the Terebellidae/Trichobranchidae/Ampharetidae/Alvinellidae clade, differences in substitution processes of Alvinellidae and Ampharetidae obscured signal for this clade by exhibiting a state different from the apomorphic state of this clade in one or both of these two taxa (Fig. 8). Hence, a large proportion of the data would still exhibit the original character-state only in Terebellidae and Trichobranchidae, but not in Ampharetidae/Alvinellidae. As only four character states are exhibited in nucleotide data and because of skews in mitochondrial nucleotide frequencies, the likelihood is high that, in this case, states exhibited in Ampharetidae, Alvinellidae, or both, are also present in either Terebellidae or Pectinaridae. Accordingly, results of spectral analyses showed that 1) most of the positions in mitochondrial data supporting the split of Trichobranchidae/Ampharetidae/Alvinellidae are noisy within *ingroup* and *outgroup*, and 2) equal in numbers to the splits of Terebellidae/Ampharetidae/Alvinellidae and Pectinaridae/Ampharetidae/Alvinellidae (Fig. 4b). Therefore, as with the Cirripedia example, strong support for the sistergroup relationship of Terebellidae and Trichobranchidae by mitochondrial data is due to symplesiomorphic characters rather than apomorphic ones.

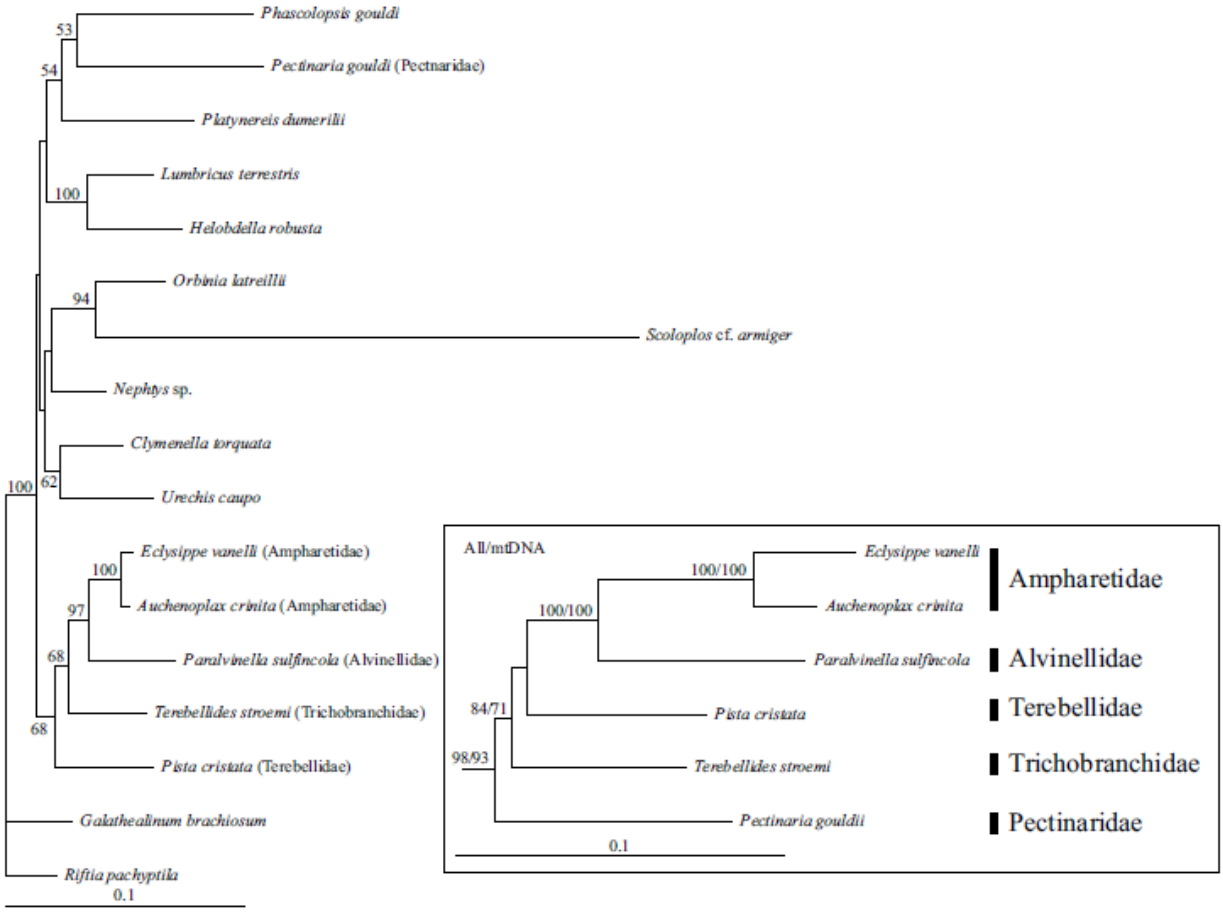


Fig. 7: Phylogenetic reconstructions using nuclear, mitochondrial and combined datasets based on RY coding. Only the nuclear ML tree is completely shown. With respect to terebelliform relationships, analyses of the mitochondrial and combined dataset recovered the same topology. Therefore, in the inlet only this part of the mitochondrial ML tree is shown and no outgroups. Only bootstrap values above 50 are shown. In the inlet, bootstrap values of the mitochondrial analysis are given at the first position and of the combined analysis at the second.

The process of deamination of the non-coding strand may be responsible for biases observed herein for pyrimidines and purines [55]. Compositional biases in our mitochondrial data were greater within pyrimidines than in purines; guanine had the lowest average frequency (16%) of all nucleotides. This is similar to the situation found in mammals though their guanine frequency can be considerably lower [15, 55, 63, 64]. In mammals, this is due to spontaneous deamination of cytosine to uracil and adenine to hypoxanthine on the complementary strand during replication of mitochondrial genomes [55]. The former deamination occurs more often than the latter [65] explaining the low level of guanines in mammals on the coding strand and the stronger bias observed in pyrimidines than in purines, because the low guanine frequency allows for little variation [15].

The best strategy to ameliorate the effect of the symplesiomorphy trap is to increase ingroup taxon sampling [17]. However, increasing the taxon sampling might not always be easily achieved or possible. For example, sampling of nearly complete mitochondrial genomes in annelids is time consuming and expensive, but new sequencing technologies are changing this. In other cases, taxon sampling will be limited by number of extant taxa from which genetic material can be obtained. Therefore, we tested different strategies with respect to their capabilities to ameliorate the effect of the symplesiomorphy trap given a limited taxon sampling. In the Cirripedia example, using appropriate methods such as ML and increased

outgroup sampling ameliorated the symplesiomorphy problem because this misplacement was due to long branches [66]. In the Mammalia example, the problem could be solved by the RY coding strategy and partitioned analyses, which resulted in weak support for the Theria hypothesis even using mitochondrial data [15]. Moreover, usage of nonstationary models of sequence evolution were able to adjust for compositional biases in mitochondrial genomes in the reconstruction of the beetle phylogeny [60].

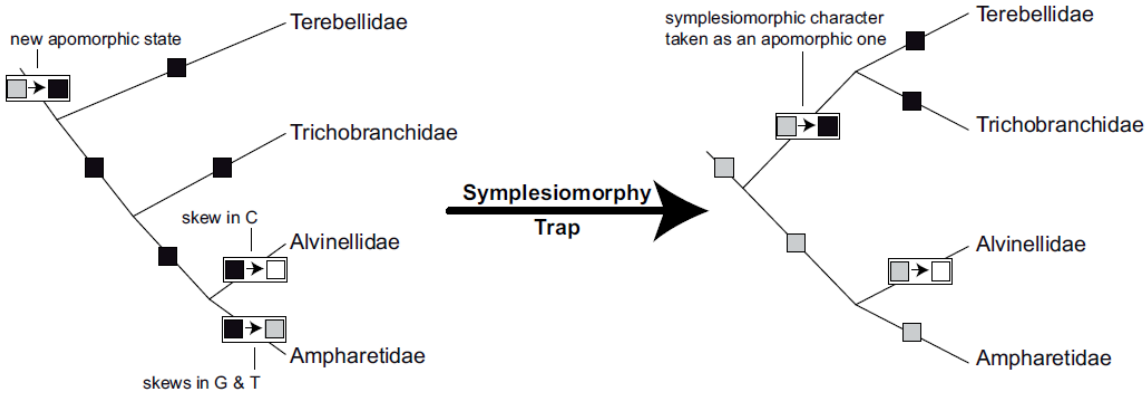


Fig. 8: Schematic representation of the effect of biases with respect to the symplesiomorphy trap in our terebelliform example. White, grey and black boxes indicate different character states as well as the possible change of one state to another along a branch.

In our case, the most effective strategy was RY coding, which reduced the effects of compositional biases within pyrimidines and purines. However, we still did not recover strong support for Trichobranchidae as sister to Ampharetidae/Alvinellidae with either mitochondrial or combined data. Moreover, phylogenetic signal in all datasets was substantially decreased by RY coding. Addition of nuclear data was only able to slightly minimize the effects of the symplesiomorphy trap as indicated, for example, by the slight decrease in bootstrap support for the presumed 'incorrect' hypothesis. Therefore, substantially more unbiased nuclear data would have been necessary to turn the tides. On the other hand, herein partitioned analyses always obtained the same topology as non-partitioned ML analyses, and PHASE analyses did not resolve incongruence either. The poor performance of non-stationary models of sequence evolution in our analyses, in comparison to Sheffield et al. [60], might be due to the limited sampling of ingroup taxa. Increased sampling may allow better adjustment to biases along the branches [58, 59]. Finally, we also tested if exclusion of biased taxa in turn would alter the results [56], but there was no noticeable effect. Thus, though several approaches were tried, none completely ameliorated the influence of the symplesiomorphy trap.

Interestingly, results based on combined data seem to be congruent with morphological and mitochondrial gene order data and, therefore, the underlying incongruence in the data was not apparent at first. Trichobranchidae strongly resemble Terebellidae and, thus, were placed as sister to or within Terebellidae [18, 20, 67]. However, only one non-homoplastic character supports their common origin: prostomium on peristomium with fused frontal edges. In contrast, others did not support a sister relationship of Terebellidae and Trichobranchidae [68, 69]. The position of two adjacent *trnM* genes also seemed to support such a relationship of Terebellidae and Trichobranchidae [23]. However, two adjacent *trnM* genes are also found in the pectinarid *P. gouldi* (Additional File 1) and in some but not all sipunculids [70-72]. Thus, no unequivocal character supports a sistergroup relationship of Terebellidae and Trichobranchidae. Analyses herein revealed that support by mitochondrial and combined data was only due to symplesiomorphic characters. On the other hand, although a close relationship

between alvinellids and ampharetids has been long suspected based on morphology [e.g., 18, 69, 73], until now strong support by molecular data [e.g., 19, 68] has been lacking.

Conclusions

Herein we report the detection of the symplesiomorphy trap in molecular data, one of a few known examples to date. Mitochondrial data placed Trichobranchidae as sister to Terebellidae in contrast to the nuclear data, which placed Trichobranchidae as sister to Ampharetidae and Alvinellidae. These latter two taxa exhibited strong compositional biases in the mitochondrial data as shown by spectral analyses as well as skew and RCFV values. However, Ampharetidae and Alvinellidae themselves were not misplaced but caused Trichobranchidae to be erroneously placed. This taxon exhibits no obvious compositional bias. Unfortunately, several state-of-the-art approaches (i.e., partitioning the dataset, performing ML analyses and partitioned analyses, use of several outgroup taxa, exclusion of biased taxa, use of different numbers of compositional vectors to implement timeheterogeneous models) were not able to ameliorate the influence of the symplesiomorphy trap in the mitochondrial data. Therefore, more sophisticated substitution models have to be developed to appropriately address this peculiar tree reconstruction artifact. In the mean time, partitioned and careful analyses can be used to detect the trap and to be aware of incongruencies in the molecular data even if nodal support is high as in our case. Given the advent of next generation sequencing technologies, we hope that analyses, such as those done here, will be better able to detect artifacts due to systematic errors because much more data will be brought to bear on such issues. Hence, these approaches may add strength and confidence to results of phylogenomic studies by allowing more in depth understanding of the sources of signal and noise.

Competing interests

The author declare no competing interests.

Authors contributions

T.H.S. and K.H.M. conceived this study. B.H., A.G. and M.N. collected the nuclear data and M.Z. the mitochondrial data. T.H.S. and M.Z. performed the analyses. T.H.S., M.Z. and K.H.M. mainly contributed to writing the manuscript.

Acknowledgements

This study was funded by the NSF-WormNet grant (EAR-0120646; DEB-1036537) and the German Science Foundation DFG STR683/5-2 from the priority program 1174 “Deep Metazoan Phylogeny” and DFG STR683/6-1. Contribution #86 to the AU Marine Biology Program and #6 to the Molette Biology Laboratory for Environmental and Climate Change Studies.

References

1. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD *et al*: **Broad phylogenomic sampling improves resolution of the animal tree of life**. *Nature* 2008, **452**(7188):745-750.
2. Dordel J, Fisse F, Purschke G, Struck TH: **Phylogenetic position of Sipuncula derived from multi-gene and phylogenomic data and its implication for the evolution of segmentation.** *J Zool Syst Evol Res* 2010, **48**(3):197–207.
3. Struck TH, Paul C, Hill N, Hartmann S, Hösel C, Kube M, Lieb B, Meyer A, Tiedemann R, Purschke G *et al*: **Phylogenomic analyses unravel annelid evolution**. *Nature* 2011, **471**:95–98.

4. Hausdorf B, Helmkampf M, Meyer A, Witek A, Herlyn H, Bruchhaus I, Hankeln T, Struck TH, Lieb B: **Spiralian phylogenomics supports the resurrection of Bryozoa comprising Ectoprocta and Entoprocta.** *Mol Biol Evol* 2007, **24**(12):2723-2729.
5. Galtier N, Nabholz B, Glemin S, Hurst GDD: **Mitochondrial DNA as a marker of molecular diversity: a reappraisal.** *Mol Ecol* 2009, **18**:4541-4550.
6. Janke A, Gemmell NJ, Feldmaier-Fuchs G, von Haeseler A, Pääbo S: **The mitochondrial genome of monotreme - The platypus (*Ornithorhynchus anatinus*).** *J Mol Evol* 1996, **42**:153-159.
7. Janke A, Xu X, Arnason U: **The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria.** *Proc Natl Acad Sci U S A* 1997, **94**:1276-1281.
8. Janke A, Magnell O, Wieczorek G, Arnason U: **Phylogenetic analysis of 18S rRNA and the mitochondrial genomes of the wombat, *Vombatus ursinus*, and the spiny anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis.** *J Mol Evol* 2002, **54**:71-80.
9. Kumazawa Y, Ota H, Nishida M, Ozawa T: **The complete nucleotide sequence of snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions.** *Genetics* 1998, **150**:313-329.
10. Penny D, Hasegawa M: **The platypus put in its place.** *Nature* 1997, **387**:549-550.
11. Zardoya R, Meyer A: **Complete mitochondrial genome suggests diapsid affinities of turtles.** *Proc Natl Acad Sci U S A* 1998, **95**:14226-14231.
12. Griffiths M: **The Biology of the Monotremes**. New York: Academic Press; 1978.
13. 13. Killian JK, Buckley TR, Stewart N, Munday BL, Jirtle RL: **Marsupials and eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution.** *Mamm Genome* 2001, **12**:513-517.
14. Lee M-H, Shroff R, Cooper SJB, Hope R: **Evolution and molecular characterization of a b-globin gene from the Australian echidna *Tachyglossus aculeatus* (Monotremata).** *Mol Phylogenet Evol* 1999, **12**:205-214.
15. Phillips MJ, Penny D: **The root of the mammalian tree inferred from whole mitochondrial genomes.** *Mol Phylogenet Evol* 2003, **28**(2):171-185.
16. Wägele JW: **Foundations of Phylogenetic Systematics**, 2nd Edition edn. München: Verlag Dr. Friedrich Pfeil; 2005.
17. Wägele JW, Mayer C: **Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects**. *BMC Evol Biol* 2007, **7**:147.
18. Rouse GW, Fauchald K: **Cladistics and polychaetes.** *Zool Scr* 1997, **26**:139-204.
19. Struck TH, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D, Halanych KM: **Annelida phylogeny and the status of Sipuncula and Echiura**. *BMC Evol Biol* 2007, **7**:57.
20. Rouse GW, Pleijel F: **Polychaetes**. Oxford: University Press; 2001.
21. Hessle C: **Zur Kenntnis der terebellomorphen Polychaeten**. *Zool Bidr Upps* 1917, **5**:39-258.
22. Holthe T: **Polychaeta Terebellomorpha**, vol. 7. Oslo: Norwegian University Press; 1986.
23. Zhong M, Struck TH, Halanych KM: **Phylogenetic information from three mitochondrial genomes of Terebelliformia (Annelida) worms and duplication of the methionine tRNA**. *Gene* 2008, **416**(1):11-21.
24. 24. Struck TH, Purschke G, Halanych KM: **Phylogeny of Eunicida (Annelida) and Exploring Data Congruence using a Partition Addition Bootstrap Alteration (PABA) approach.** *Syst Biol* 2006, **55**:1-20.
25. Burland TG: **DNASTAR's lasergene sequence analysis software.** *Methods Mol Biol* 2000, **132**:71-91.
26. Altschul SF, Gish W, Miller W, Myers EM, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
27. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25**:955–964.
28. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **24**:4876-4882.
29. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17**:540-552.
30. Maddison DR, Maddison WP: **MacClade4: Analysis of Phylogeny and Character Evolution, version 4.0.** In. Sunderland, MA: Sinauer Associates; 2002.
31. Rambaut A: **The Use of Temporally Sampled DNA Sequences in Phylogenetic Analysis.** Oxford, UK: Oxford University; 1996.
32. Swofford DL: **PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).** In., 4.0b edn. Sunderland, MA: Sinauer Associates; 2002.

33. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**:817-818.
34. Posada D, Crandall KA: **Selecting the best-fit model of nucleotide substitution.** *Syst Biol* 2001, **50**:580-601.
35. 35. Stamatakis A, Hoover P, Rougemont J: **A rapid bootstrap algorithm for the RAxML web-servers**. *Syst Biol* 2008, **75**(5):758-771.
36. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754-755.
37. Nylander JAA: **MrModeltest.** In., 1.1 edn: **http://morphbank.ebc.uu.se/MR.BAYES;** 2002.
38. Nylander JAA: **MrModeltest v2.** In. Evolutionary Biology Centre, Uppsala University: Program distributed by the author.; 2004.
39. **Tracer v1.4** [Available from **http://beast.bio.ed.ac.uk/Tracer**]
40. Gowri-Shankar V, Rattray M: **A Reversible Jump Method for Bayesian Phylogenetic Inference with a Nonhomogeneous Substitution Model**. *Mol Biol Evol* 2007, **24**(6):1286-1299.
41. Shimodaira H: **An approximately unbiased test of phylogenetic tree selection**. *Syst Biol* 2002, **51**(3):492-508.
42. Shimodaira H, Hasegawa M: **CONSEL: for assessing the confidence of phylogenetic tree selection**. *Bioinformatics* 2001, **17**(12):1246-1247.
43. Lockhart PJ, Howe C, Barbrook A, Larkum AWD, Penny D: **Spectral Analysis, Systematic Bias, and the Evolution of Chloroplasts**. *Mol Biol Evol* 1999, **16**(4):573-576.
44. Lockhart PJ, Penny D, Meyer A: **Testing the phylogeny of swordtail fishes using split decomposition and spectral analysis**. *J Mol Evol* 1995, **41**(5):666-674.
45. Cao Y, Fujiwara M, Nikaido M, Okada N, Hasegawa M: **Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data.** *Gene* 2000, **259**:149-158.
46. 46. Mouchaty SK, Gullberg A, Janke A, Arnason U: **Phylogenetic position of the Tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of *Echinops telfari*.** *Zool Scr* 2000, **29**:307-317.
47. Schmitz J, Ohme M, Zischler H: **The complete mitochondrial sequence of *Tarsius bancanus*: Evidence for an extensive nucleotide compositional plasticity of primate mitochondrial DNA.** *Mol Biol Evol* 2002, **19**.
48. Härlid A, Arnason U: **Analysis of mitochondrial DNA nest ratite birds within the Neognathae - supporting a neotenous origin of ratite morphological characters.** *Proc R Soc London, B* 1999, **266**:1-5.
49. Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Ast JC, Yuri T: **Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes.** *Syst Biol* 1999, **48**:138-152.
50. Foster PG, Hickey DA: **Compositional Bias May Affect Both DNA-Based and Protein-Based Phylogenetic Reconstructions**. *J Mol Evol* 1999, **48**(3):284-290.
51. Hassanin A, LéGer N, Deutsch J: **Evidence for Multiple Reversals of Asymmetric Mutational Constraints during the Evolution of the Mitochondrial Genome of Metazoa, and Consequences for Phylogenetic Inferences**. *Syst Biol* 2005, **54**(2):277 - 298.
52. Longhorn SJ, Foster PG, Vogler AP: **The nematode-arthropod clade revisited: phylogenomic analyses from ribosomal protein genes misled by shared evolutionary biases**. *Cladistics* 2007, **23**:130-144.
53. Stach T, Braband A, Podsiadlowski L: **Erosion of phylogenetic signal in tunicate mitochondrial genomes on different levels of analysis**. *Mol Phylogenet Evol* 2010, **55**(3):860-870.
54. 54. Perna NT, Kocher TD: **Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes**. *J Mol Evol* 1995, **41**(3):353-358.
55. Reyes A, Gissi C, Pesole G, Saccone C: **Asymmetrical directional mutation pressure in the mitochondrial genome of mammals.** *Mol Biol Evol* 1998, **15**:957-966.
56. Bergsten J: **A review of long-branch attraction**. *Cladistics* 2005, **21**(2):163-193.
57. Struck TH, Nesnidal MP, Purschke G, Halanych KM: **Detecting possibly saturated positions in 18S and 28S sequences and their influence on phylogenetic reconstruction of Annelida (Lophotrochozoa)**. *Mol Phylogenet Evol* 2008, **48**(2):628-645.
58. Lecointre G, Philippe H, Van Le HL, Le Guyader H: **Species sampling has a major impact on phylogenetic inference**. *Mol Phylogenet Evol* 1993, **2**(3):205-224.
59. Milinkovitch MC, LeDuc RG, Adachi J, Farnir F, Georges M, Hasegawa M: **Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on DNA sequence data in Cetaceans.** *Genetics* 1996, **144**:1817-1833.
60. Sheffield NC, Song H, Cameron SL, Whiting MF: **Nonstationary evolution and compostional heterogeneity in beetle mitochondrial phylogenomics**. *Syst Biol* 2009, **58**(4):381-394.

61. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM: **Chapter 11 – Phylogenetic Inference**. In: *Molecular Systematics*. Edited by Hillis DM, Moritz C, Mable BK, 2 edn. Sunderland, MA: Sinauer Associates; 1996: 407-514.
62. Rousset V, Pleijel F, Rouse GW, Erséus C, Siddall ME: **A molecular phylogeny of annelids**. *Cladistics* 2007, **23**(1):41-63.
63. 63. Phillips MJ, Lin Y-H, Harrison GL, Penny D: **Mitochondrial genomes of a bandicoot and a brushtail possum confirm the monophyly of australidelphian marsupials.** *Proc R Soc London, B* 2001, **268**:1533-1538.
64. Springer MS, Douzery EJP: **Secondary structure and patterns of evolution among mammalian 12S rRNA molecules.** *J Mol Evol* 1996, **43**:357-373.
65. Tanaka M, Ozawa T: **Strand asymmetry in human mitochondrial DNA mutations.** *Genomics* 1994, **22**:327-335.
66. Pérez-Losada M, Høeg JT, Kolbasov GA, Crandall KA: **Reanalysis of the relationship among the Cirripedia and the Ascothoracida and the phylogenetic position of the Facetotecta (Maxillopoda: Thecostraca) using 18S rDNAsequences.** *J Crust Biol* 2002, **22**:661-669.
67. Malmgren AJ: **Nordiska Hafs - Annulater.** *Öfv af K Sven Vet Akad Förhandl* 1866, **22**:355-410.
68. Rousset V, Rouse G, Féral J-P, Desbruyères D, Pleijel F: **Molecular and morphological evidence of Alvinellidae relationships (Terebelliformia, Polychaeta, Annelida)**. *Zool Scr* 2003, **32**:185–197.
69. Glasby CJ, Hutchings PA, Hall K: **Assessment of monophyly and taxon affinities within the polychaete clade Terebelliformia (Terebellida)**. *J Mar Biol Ass UK* 2004, **84**(05):961-971.
70. Mwinyi A, Meyer A, Bleidorn C, Lieb B, Bartolomaeus T, Podsiadlowski L: **Mitochondrial genome sequence and gene order of *Sipunculus nudus* give additional support for an inclusion of Sipuncula into Annelida**. *BMC Genomics* 2009, **10**:27.
71. 71. Boore JL, Staton JL: **The Mitochondrial Genome of the Sipunculid *Phascolopsis gouldii* Supports Its Association with Annelida Rather than Mollusca**. *Mol Biol Evol* 2002, **19**(2):127-137.
72. Shen X, Ma X, Ren J, Zhao F: **A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of *Phascolosoma esculenta***. *BMC Genomics* 2009, **10**:136.
73. Desbruyères D, Laubier L: ***Alvinella pompejana* gen. sp. nov., Ampharetidae abberant des sources hydrothermales de la ride Est-Pacifique.** *Oceanol Acta* 1980, **3**:267-274.
74. Spears T, Abele LG, Applegate MA: **A phylogenetic study of cirripeds and their relatives (Crustacea: Thecostraca).** *J Crust Biol* 1994, **14**:641-656.

5.6. Mitochondrial genomes to the rescue – Diurodrilidae in the myzostomid trap

Anja Golombek[1,†], Sarah Tobergte[1,2,†], Maximilian P. Nesnidal[2,3], Günter Purschke[2], Torsten H. Struck[1]

[1] Zoological Research Museum Alexander Koenig, Adenauerallee 160, D-53113 Bonn, Germany; torsten.struck.zfmk@uni-bonn.de, A.Golombek@gmx.de
[2] Universität Osnabrück; FB05 Biologie/Chemie; AG Zoologie; Barbarastr. 11; D-49076 Osnabrück; Germany; Purschke@biologie.uni-osnabrueck.de, sarah.tobergte@freenet.de
[3] Zoological Museum, University of Hamburg, Martin-Luther-King-Platz 3, D-20146 Hamburg, Germany; nesnidal@gmail.com

† Shared first authorship, both authors contributed equally to this work

Corresponding author:
Torsten H. Struck; Zoological Research Museum Alexander Koenig; Adenauerallee 160; D-53113 Bonn; Germany
Phone: +49-225-9122-401; Fax: +49-225-9122-212; E-Mail: torsten.struck.zfmk@unibonn.de

Abstract

Diurodrilidae is a taxon of Lophotrochozoa comprising about six, exclusively interstitial species, which are up to 500 µm long and dorsoventrally flattened. Traditionally, Diurodrilidae had been regarded as an annelid family. However, recently Diurodrilidae had been excluded from Annelida and been placed in closer relationship to platyzoan taxa based on both morphological and nuclear rRNA data. Since both, Diurodrilidae and platyzoan taxa, exhibit long branches in the molecular analyses, the close relationship might be due to a long branch attraction artifact. The annelid taxon Myzostomida had been trapped in a similar long branch attraction artifact with platyzoan taxa using nuclear rRNA data, but determination of the nearly complete mitochondrial genome of myzostomids revealed their annelid affinity. Therefore, we determined the nearly complete mitochondrial genome of *Diurodrilus subterraneus* as well as new nuclear rRNA data for *D. subterraneus* and some platyzoan taxa. All our analyses of nuclear rRNA and mitochondrial sequence and gene order data presented herein clearly place Diurodrilidae within Annelida and with strong nodal support values in some analyses. Therefore, the previously suggested exclusion of Diurodrilidae from Annelida and its close relationship with platyzoan taxa can be attributed to a long branch artifact. Morphological data do not unambiguously support a platyzoan affinity of Diurodrilidae, but instead would also be in line with a progenetic origin of Diurodrilidae within Annelida.

Key words: Diuroldrilidae, Annelida, Playtzoa, mitochondrial genomes, next generation sequencing, nuclear rDNA

1. Introduction

Diurodrilidae is a monogeneric taxon only comprising the genus *Diurodrilus*, which currently includes only six described species, but there are eventually four more species not being described yet (Paxton, 2000; Riser, 1984; Worsaae and Kristensen, 2005; Worsaae and Rouse, 2008). *Diurodrilus* is a meiofaunal interstitial worm occurring from intertidal sandy beaches to subtidal habitats with records from all oceans. Individuals of all species are minute ranging in size from 250 up to 500 □m. They are ventral-dorsally flattened and possess a body organization, which has been suggested to consist of a prostomium, a peristomium, five

vaguely-defined segments and a pygidium given an annelid affinity (Worsaae and Kristensen, 2005). A characteristic feature of the genus *Diurodrilus* is the forked, toe-like posterior end (pygidium) carrying duo-gland systems of adhesive glands, which allow a fast attachment and detachment to sand grains (Worsaae and Kristensen, 2005).

The first species, *Diurodrilus minimus,* was described by Remane (1925) and he placed *Diurodrilus* with the annelid taxon Dinophilidae as part of "Archiannelida". The in the meantime abolished group "Archiannelida" comprised interstitial polychaete families such as Polygordiidae, Protodrilidae and Nerillidae, which were considered to show the primitive annelid condition. However, later studies showed that there is no synapomorphy supporting the monophyly of "Archiannelida" and its taxa are most likely highly derived annelid taxa adapted to the interstitial habitat instead of being close to the annelid stem species (e.g., Fauchald, 1974; Purschke, 1985a, b; Purschke and Jouin, 1988; Rouse and Fauchald, 1997; Struck et al., 2005; Struck et al., 2008; Struck et al., 2002; Tzetlin and Purschke, 2005; Westheide, 1985; Westheide, 1990; Worsaae and Kristensen, 2005; Worsaae and Rouse, 2008). Some of them like Dinophilidae have supposedly evolved by progenesis, that is the retention of ancestral larval and/or juvenile traits in the adult stage (e.g., Struck, 2006; Westheide, 1985, 1987). In the course of these new considerations *Diurodrilus* was also removed from Dinophilidae as no morphological character supported the placement of *Diurodrilus* within Dinophilidae and a new family Diurodrilidae or even order Diurodrilida was erected within Annelida (Kristensen and Niilonen, 1982; Westheide, 1985). A recent comprehensive study of *Diurodrilus* based on morphological and molecular data could not find any close affinity of *Diurodrilus* to Annelida (Worsaae and Rouse, 2008). Typical annelid characters such as segmentation, nuchal organs, head appendages, parapodia and chaetae are lacking. Even analyses of the nervous system were not able to reveal traces of an ancestral segmentation (Worsaae and Rouse, 2008), as it had been possible for the nonsegmented annelid taxa Echiura and Sipuncula (Hessling, 2002, 2003; Hessling and Westheide, 2002; Kristof et al., 2008; Purschke et al., 2000). Moreover, the analyses of 18S and 28S rDNA data did not place *Diurodrilus* within Annelida, but rather recovered a closer relationship to platyzoan taxa such as Micrognathzoa, however, with moderate support by the 28S data (Worsaae and Rouse, 2008). This suited well with the findings that peripharyngeal and paired trunk ciliated cells, so called ciliophores (Worsaae and Rouse 2008) are present in Micrognathozoa and *Diurodrilus*, that the two pairs of protonephridia in Micrognathozoa and *Diurodrilus* are in an anterior position and that there are further resemblances between Gnathostomulida and *Diurodrilus* (Worsaae and Rouse, 2008). Therefore, Worsaae and Rouse (2008) suggested removing Diurodrilidae from Annelida and placing them as *incertae sedis* within Spiralia with affinities to playtzoan taxa.

However, in the molecular analyses of Worsaae and Rouse (2008) *Diurodrilus* as well as the platyzoan taxa exhibit strongly increased substitution rates in both genes. This is known to result in the artificial attraction of long-branched taxa (for review see Bergsten, 2005). Therefore, to address the uncertain phylogenetic affinity of Diurodrilidae we determined the nearly complete mitochondrial genome of *Diurodrilus subterraneus* Remane, 1934 using a next-generation sequencing approach as well as complete 18S and 28S sequence data. Additionally, we determined 18S and 28S sequence data of three gastrotrichs, one rotifer and one gnathostomulid and added published 18S and 28S data of several annelid taxa not included by Worsaae and Rouse (2008) in order to extend the taxon sampling and thus to achieve a better resolution in the analyses (see Bergsten, 2005). In our analyses, the mitochondrial gene order data strongly support an annelid affinity of *Diurodrilus*. Moreover, the phylogenetic analyses of both mitochondrial and nuclear rRNA data are also in agreement with such a placement. In the light of these findings the morphological characters have been reevaluated.

2. Material and Methods

2.1. Material

Specimens of *Diurodrilus subterraneus* (location: N 55°00.061'/E 008°23.302'), *Tetranchyroderma megastoma* (Remane, 1927) (Gastrotricha, Macrodasyida, location: N 55°01.506'/E 008°26.180'), *Dactylopodola baltica* (Remane, 1926) (Gastrotricha, Macrodasyida, location: N 55°01.506'/E 008°26.179'), and *Gnathostomula paradoxa* Ax, 1956 (Gnathostomulida, Bursovaginoidea, location: N 55°01.508'/E 008°26.180') were collected near List, North Sea Island of Sylt, Germany. Samples of *Lepidodermella squamata* (Dujardin, 1841) (Gastrotricha, Chaetonotida) were obtained from Carolina Biological Supply Company (Burlington, NC, USA). Specimens of *Brachionus calyciflorus* Pallas, 1766 (Rotifera, Monogononta) were a gift from the Hariston Lab, Ithaca, USA.
Animals were preserved in RNAlater (Sigma, Hamburg, Germany) or snap frozen in liquid nitrogen and then stored at -70°C. Genomic DNA was extracted using the DNeasy Tissue Kit (QIAGENSigman A, Hilden, Germany) according to manufactures instruction.

2.2. Determination of 18S and 28S genes of gastrotrichs, gnathostomulids and rotifers

18S and 28S rRNA genes were amplified from the genomic DNA as described in Struck et al. (2006). PCR fragments of the expected sizes were excised from TAE agarose gel and purified via QIAquick Spin columns (QIAGEN, Hilden, Germany). Purified fragments were sequenced on an ABI PRISM 377 DNA sequencer using BigDye sequencing chemistry (Applied Biosystems, Carlsbad, CA, USA). 18S and 28S sequences were assembled using SeqMan II (DNASTAR Inc., Madison, WI, USA).

2.3. Determination of the mitochondrial genome and 18S and 28S genes of Diurodrilus

To increase the amount of genomic DNA the whole genome of *Diurodrilus* was amplified using the illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare Life Science, Freiburg, Germany) according to manufacturers instructions. The amplified genomic DNA was send to Genterprise Genomics (Mainz, Germany) for preparation of a genomic DNA shot-gun library including fragmentation and tagging adaptors for multiplexing. The library was 100 bp paired-end sequenced on a 1/24th of a lane on an Illumina HiSeq2000 using TruSeq v3 chemistry. After quality filtering and adapter trimming using the Chastity filter a total of 11.70 million reads with an average read length of 96 bp were present. These were assembled into contigs with CLC Genomic workbench (CLC bio, Aarhus, Denmark) using default parameters and a fragment size window of 200 to 600 bp as the maximum peak of the fragmented library was at 400 bp. Using protein sequence information of the 13 proteincoding mitochondrial genes of *Platynereis dumerilii* (AF178678) as well as the sequence of the 18S (EF552204) and 28S (EF552205) of *Diurodrilus* sp. as query sequences in TBLASTN or BLASTN searches, respectively, we searched in the assembled contigs for fragments of the mitochondrial genome as well as the genomic rRNA gene cluster. As both the mitochondrial genome and the genomic 18S and 28S rRNA gene cluster of *Diurodrilus* was present in several fragments the gaps between the fragments were closed using speciesspecific primers (Table 1), which were designed with the aid of the Primer3Plus web-interface (Untergasser et al., 2007). The missing fragments were amplified from the same genomic DNA, which had also been used to for the whole genome amplification, using the QIAGEN® Multiplex PCR Kit (QIAGEN, Hilden, Germany) (20 µl reaction: 10 µl multiplex mix, 2 µl Q solution, 1.6 µl 10 pmol/µl forward primer, 1.6 µl 10 pmol/µl reverse primer, 2.5 µl genomic DNA and 2.3 µl water) and a touch down PCR (initial denaturation: 15' 95°C; 15 cycles: 35" 94°C, 90" 55°C or 60°C (decreasing

1°C at each cycle), 90" 72°C; 25 cycles: 35" 94°C, 90" 50°C or 55°C, 90" 72°C; final elongation: 10' 72°C). PCR fragments of the expected sizes were excised from TBE agarose gel and purified via NucleoSpin® Gel and PCR Clean-up (MACHEREY-NAGEL, Düren, Germany). Purified fragments were sent to Macrogen Europe (Amsterdam, Netherland) for sequencing. The sequences of the amplified fragments were assembled with present contigs using SeqMan II (DNASTAR Inc., Madison, WI, USA).
The mitochondrial genome was annotated using the MITOS webserver with the invertebrate mitochondrial code (NCBI code table 5) and manually corrected based on the raw data of the pipeline provided by the MITOS webserver (Bernt et al., in press). Specifically the positions of the mitochondrial rRNAs and tRNAs were checked. The rRNA cluster was annotated by BLASTN 2.2.27+ searches against the reference RNA sequences (refseq_rna) of NCBI.

Table 1: Sequence information of species-specific primers used to fill the gaps.

Primer	Sequence (5' -> 3')	Direction	Tm (°C)
Diuro28S_1F	GCACGTGAAACCGCTTAGAC	Forward	60.0
Diuro28S_2F	TACTGTTCGTAGCAGTGGGG	Forward	60.0
Diuro28S_3R	CTGAGGGAAACTTCGGAGGG	Reverse	63.0
Diuro28S_R3	AGCAAGCCGCTAACAACTCTC	Reverse	61.0
Diuro28S_F3	CTCATGAGATACCATAAAAGG	Forward	55.0
Diuro28S_R4	CGAAGAAAAGAGAACTCTTCC	Reverse	57.0
Diuro28S_4R	CGAAGCACTTTTCGCATCCG	Reverse	60.0
Diuro7422_F	ATATCCGCCCCACCAAGAAC	Forward	60.0
Diuro8483_R	TGGGCTCATGTGACAGTAGC	Reverse	60.0
Diuro784_F	AAGCTTGGCCCCTACTCATG	Forward	60.0
Diuro784in_R	AAGGTAGCTAGAAGGGGACG	Reverse	60.0
Diuro784in_F	AATCCCAGGCCGTCTTAACC	Forward	60.0
Diuro5151_R	AATAAGGAGGCGGCCATAGC	Reverse	60.0
Diuro8483_F	TATGCCCATATCGGACGAGC	Forward	60.0
Diuro5151in_R	GGAGGTTCGACTGGTCTTCC	Reverse	63.0
Diuro5151in_F	CGGTGTACTAGCTATGGCCG	Forward	63.0
Diuro8224_R	ACCCTAAACCATCTCAGCCC	Reverse	60.0
Diuro5151_F	TGTTCGAAGCCGCAATTTGC	Forward	58.0
Diuro7422_R	GGGTGATTAGGTTGGGGGAG	Reverse	63.0
Diuro8224_F	CCCTCAAACCAAACTAATCCGAC	Forward	63.0
Diuro784_R	TCCACATTCAAAGGGCGAGG	Reverse	60.0

2.4. Phylogenetic analyses

The newly determined nuclear rRNA data of a diurodrilid, three gastrotrichs, one rotifer and one gnathostomulid were complemented with the data of Worsaae and Rouse (2008) plus additionally one rotifer and the annelid taxa of Struck et al. (2008) (Table 2). 18S and 28S data were independently aligned using MAFFT with the L-INS-I option, ambiguously aligned positions were masked using AliScore (Katoh et al., 2005; Kück et al., 2010; Misof and Misof, 2009) with default settings except for gaps, which were treated as missing data, and concatenated into one 18S/28S dataset using FASconCAT (Kück and Meusemann, 2010). Similarly, we obtained complete or nearly complete mitochondrial (mt) genomes from GenBank (Table 3) and annotated each genome using the MITOS webserver with the invertebrate mitochondrial code (NCBI code table 5) (Bernt et al., in press) as we did for *Diurodrilus subterraneus* to ensure that the annotation procedure of mitochondrial genes was the same for all taxa. We used CREx (Bernt et al., 2007) to conduct pairwise comparisons of the mitochondrial gene orders of these genomes and *D. subterraneus* using two dissimilarity measurements, breakpoint and reversal distance, as well as a similarity measurement, common interval. Moreover, CREx also determined the most likely genome rearrangement scenario between the gene order of *D. subterraneus* and each annelid and platyzoan gene order including

transpositions, reverse transpositions, reversals, and tandem-duplicationrandom- loss (tdrl) events. From these genomes we compiled a dataset of all mitochondrial protein-coding genes except for NADH dehydrogenase subunit 4L (nad4L), ATP synthase subunit 6 (atp6) and 8 (atp8) due to their very high variability. The nucleotide (nuc) sequences were translated into amino acid (aa) sequences for each gene using the invertebrate mitochondrial code (NCBI code table 5) using GeneDoc (Nicholas and Nicholas, 1997). The aa sequences of each gene were aligned using MAFFT (Katoh et al., 2005). The nuc sequences were then aligned based on the aa alignment using the PAL2NAL webserver (Suyama et al., 2006), so that the nuc alignment directly corresponded to the aa alignment. Finally, for both the nuc and aa alignments of each gene ambiguously aligned position were masked using AliScore (Kück et al., 2010; Misof and Misof, 2009) and concatenated into one dataset using FASconCAT (Kück and Meusemann, 2010) resulting in two aligned datasets: mt-nuc and mt-aa.
For the three concatenated datasets (18S/28S, mt-nuc and mt-aa) partitioned Maximum Likelihood (ML) analyses were conducted with RAxML 7.3.1 (Stamatakis, 2006) applying individual substitution models and branch lengths for each gene and 100 replicates. For the nucleotide datasets the GTR+Γ+I substitution model was applied to each gene. For the amino acid data the best fitting substitution model was determined using the ProteinModelSelection script of RAxML 7.3.1. For all genes, the MTZOA+Γ+I substitution model was the best fitting model and in addition empirical frequencies (+F) had to be applied for the genes cytochrome oxidase subunit 2 (cox2), cytochrome b (cob), NADH dehydrogenase subunit 2 (nad2), 4 (nad4) and 5 (nad5). Confidence values for the edges of the ML tree were computed using the automatic bootstopping option (-# autoMRE) in RAxML to a maximum of 1,000 bootstrap (BP) replicates (Felsenstein, 1985). As Myzostomida showed long branches in the partitioned ML analyses of the mitochondrial data we rerun the analyses without Myzostomida to test if a long branch attraction of Myzostomida and Diurodrilidae placed Diurodrilidae within Annelida.

3. Results

3.1. Determination of the mitochondrial genome and the nuclear rRNA cluster of *Diurodrilus subterraneus*

The small genomic shot-gun library of *Diurodrilus subterraneus* assembled into 38335 contigs larger than 300 bp and 7243 were larger than 1000 bp. The N50 value of the library was 864 bp and the average contig length 875 bp. Moreover, the library contained five larger fragments of the mitochondrial genome. The largest fragment had a length of 6,885 bp with 17.3x average coverage and comprised nad2, nad3, cox1, cox2 and parts of nad1 and cox3. The other four fragments were 1419 bp, 1326 bp, 1419 bp and 1086 bp with an average coverage ranging from 6.9x to 8.4x and coded for cox3, nad6, cob, atp6, nad5 and nad4 or parts of these genes. Using species-specific primers all gaps between the fragments could be closed except for the one containing the unknown or control region, which is not unusual for mitochondrial genomes due to its secondary structure and the presence of long nucleotide homopolymers (Halanych and Janosik, 2006; Vallès and Boore, 2006). In total we were able to determine 13,025 bp of the mitochondrial genome of *Diurodrilus subterraneus* containing all protein-coding, rRNA and tRNA genes typical for Bilateria (Fig. 1) except for the small rRNA (rrnS), atp8 and the tRNAs for cysteine (C), methionine (M), asparagine (N), arginine (R), and serine (S1(AGA)).
In the shot-gun library four fragments of the nuclear rRNA cluster were also present. The largest fragment with a length of 4,256 bp and an average coverage of 32.0x contained the complete 18S and 5.8S rRNA as well as the 5' end of the 28S. The second largest fragment with a length of 2,341 bp and coverage of 22.4x contained the 3' end of 28S. The two smaller fragments of 643 bp and 462 bp length and coverage of 4.6x and 4.5x, respectively, contained intermediate

parts of the 28S. Using species-specific primers we were able to close the gaps between the four fragments so that the length of the fragment containing the nuclear rRNA cluster was 8,095 bp. The cluster exhibited the typical arrangement with 18S rRNA coming first, followed by 5.8S and 28S (Fig. 1). Between the rRNA were the internal transcribed spacer 1 and 2 (ITS1, ITS2). We were not able to determine the exact beginning of the external transcribed spacer 1 (ETS1) and the end of external transcribed spacer 2 (ETS2) due to lack of appropriate RNA sequences.

3.2. Phylogenetic analyses of nuclear rRNA

Phylogenetic analyses of the 18S and 28S rRNA data recovered neither the monophyly of Annelida nor of Platyzoa (Fig. 2). Four annelid taxa (i.e., Dinophilida, Protodrilidae, *Owenia fusiformis* and *Apistobranchus typicus*) are placed among the outgroup taxa. Platyhelminthes was sister to all other lophotrochozoan taxa except Protodrilidae. The remaining platyzoan taxa (i.e. Gastrotricha, Gnathostomulida, Micrognathozoa and Syndermata) grouped together and the clade was sister to the remaining lophotrochozoan taxa. However, bootstrap support values for these basal relationships were low (BS < 50, Fig. 2).
Diurodrilus subterraneus was sister to *Diurodrilus* sp. with maximal bootstrap support of 100. Diurodrilidae was placed with Terebelliformia as part of the large clade of annelid taxa albeit given low bootstrap support (Fig. 2). We also conducted individual analyses of 18S and 28S. Both analyses placed Diurodrilidae within a larger clade of Annelida albeit given low bootstrap support and never close to any platyzoan taxon, especially *Limnognathia maerski* (Micrognathozoa) (Supplementary Figs. 1 & 2).

3.3. Phylogenetic analyses of mitochondrial protein-coding genes

The phylogenetic analyses of mitochondrial protein-coding genes as amino acids recovered *Diurodrilus subterraneus* as part of monophyletic Annelida with moderate bootstrap support (BS = 73, Fig. 3A). Within this clade a sistergroup relationship of *Diurodrilus* and Myzostomida was strongly supported (BS = 96). On the other hand, monophyly of Platyzoa comprising Platyhelminthes and Syndermata was also substantiated (BS = 92). It should be noted that mitochondrial genome data are still lacking for the platyzoan taxa Gnathostomulida, Gastrotricha and Micrognathozoa. The sistergroup relationship of Acoela and Platyzoa was also strongly supported by the mitochondrial data (BS = 98, Fig. 3A).
To test for a potential long-branch attraction of *Diurodrilus* to Myzostomida, which exhibited long branches in the partitioned analyses, we excluded Myzostomida. Excluding Myzostomida resulted in a tree placing *Diurodrilus* within monophyletic Annelida with significant bootstrap support (BS = 98, Fig. 3B). Within this clade *Diurodrilus* is recovered as sister to Orbiniidae but with low support (BS = 61). Similarly, a sistergroup relationship of Platyzoa and Acoela is also significantly supported by the mitochondrial data (BS = 100) and monophyly of Platyzoa is recovered as well (BS = 84, Fig. 3B).
Analyses on the nucleotide level with and without Myzostomida obtained similar results (Supplementary Fig. 3). Monophyly of Annelida including Diurodrilidae was recovered but with low support (BS < 50). However, this was due to the terebelliform taxa *Auchenoplax* and *Eclysippe* and not *Diurodrilus*. It is known that these two terebelliform taxa exhibit strongly deviating nucleotide compositions (Zhong et al., 2011) and therefore were sister to all other Annelida. Within Annelida placement of *Diurodrilus* within a larger clade of annelids was moderately supported by BS values of 70 and 86 for the analyses with and without Myzostomida, respectively. In the analyses, which included Myzostomida, *Diurodrilus* was sister to Myzostomida with low support (BS = 67).

Table 2: Sequence accession numbers of the taxa used in the analyses with 18S and 28S. Sequences that were newly obtained for the present study are indicated in boldface.

Taxon	Species	18S	28S
Diurodrilidae	*Diurodrilus* sp.	EF552204	EF552205
	Diurodrilus subterraneus Remane, 1934	**XXNNNNNN**	**XXNNNNNN**
Annelida			
Dinophilidae	*Dinophilus* sp.	FJ200245	FJ200246
	Trilobodrilus axi Westheide, 1967	AF412806	AY732231
	Trilobodrilus heideri Remane, 1925	AF412807	AY894292
Nerillidae	*Leptonerilla prospera* (Sterrer & Iliffe, 1982)	AY834758	EU418871
Polygordiidae	*Polygordius appendicularis* (Fraipont, 1887)	AY525629	EU418872
Protodrilidae	*Protodrilus ciliatus* Jägersten, 1952	AY525631	EU418873
	Protodrilus purpureus (Schneider, 1868)	AY527057	EU418874
Protodriloidae	*Protodriloides chaetifer* (Remane, 1926)	AY527058	EU418875
	Protodriloides symbioticus (Giard, 1904)	AF508125	EU418876
Saccocirridae	*Saccocirrus* sp.	EU418861	EU418877
Oweniidae	*Owenia fusiformis* delle Chiaje, 1841	AF448160	DQ790049
Apistobranchidae	*Apistobranchus typicus* (Webster & Benedict, 1887)	AF448150	EU418870
Chaetopteridae	*Chaetopterus variopedatus* (Renier, 1804)	U67324	AY145399
Sipuncula	*Phascolopsis gouldi* (Pourtalès, 1851)	AF342796	AF342795
Amphinomidae	*Paramphinome jeffreysii* (Mcintosh, 1868)	AY838856	AY838865
	Eurythoë complanata (Pallas, 1766)	AY364851	AY364849
Aphroditidae	*Aphrodita* sp.	AY894295	DQ790024
Polynoidae	*Gattyana ciliata* Moore, 1902	AY894297	DQ790035
	Lepidonotus sublevis Verrill, 1873	AY894301	DQ790039
Sigalionidae	*Sigalion spinosus* (Hartman, 1939)	AY894304	DQ790062
	Sthenelanella uniformis Moore, 1910	AY894306	DQ790064
Hesionidae	*Ophiodromus pugettensis* (Johnson, 1901)	DQ790086	DQ790046
Pilargidae	*Ancistrosyllis groenlandica* McIntosh, 1879	DQ790075	DQ790023
Nereididae	*Nereis succinea* (Frey and Leuchart, 1847)	AY210447	AY210464
	Nereis vexillosa Grube, 1851	DQ790083	DQ790043
Syllidae	*Exogone naidina* Oersted, 1845	AF474290	
	Exogone verugera (Claparede, 1868)		DQ790033
	Proceraea cornuta (Agassiz, 1862)	AF212179	AF212165
	Typosyllis anoculata (Hartmann-Schröder, 1962)	DQ790098	DQ790071
Glyceridae	*Glycera dibranchiata* Ehlers, 1868	AY995208	AY995207
	Glycera americana Leidy, 1855	EU418856	EU418864
Goniadidae	*Goniada brunnea* Treadwell, 1906	DQ790080	DQ790037
	Glycinde armigera Moore, 1911	DQ790079	DQ790036
Nephtyidae	*Nephtys longosetosa* (Oersted, 1842)	DQ790082	DQ790042
	Nephtys incisa Malmgren, 1865	EU418857	EU418865
	Aglaophamus circinata (Verrill, 1874)	DQ790072	DQ790020
Paralacydoniidae	*Paralacydonia paradoxa* Fauvel, 1913	DQ790088	DQ790050
Alciopidae	*Alciopina* sp.	DQ790073	DQ790021
	Torrea sp.	DQ790096	DQ790068
Tomopteridae	*Tomopteris* sp.	DQ790095	DQ790067
Dorvilleidae	*Protodorvillea kefersteini* (McIntosh, 1869)	AF412799	AY732230
	Parougia eliasoni (Oug, 1978)	AF412798	DQ790053
Eunicidae	*Marphysa sanguinea* (Montagu, 1815)	AY525621	AY838861
	Eunice sp.	AF412791	AY732229
Lumbrineridae	*Lumbrineris latreilli* Audouin & Milne-Edwards, 1834	AY525623	AY366512
	Lumbrineris inflata (Moore, 1911)	AY525622	AY366518
	Ninoe nigripes Pettibone, 1982	AY838852	AY838862
Oenonidae	*Arabella semimaculata* (Moore, 1911)	AY838844	AY838857
	Drilonereis longa Webster, 1879	AY838847	AY838860

Table 2 (continued)

Taxon	Species	18S	28S
	Oenone fulgida Pettibone, 1982	AY838853	AY838863
Onuphidae	*Diopatra aciculata* Knox and Cameron, 1971	AY838845	AY838858
	Hyalinoecia tubicola O.F. Müller, 1776	AF412794	AY732228
Orbiniidae	*Orbinia swani* Pettibone, 1957	DQ790087	DQ790048
	Scoloplos fragilis (Verrill, 1873)	AY532360	EU418863
Parergodrilidae	*Parergodrilus heideri* Reisinger, 1925	AJ310504	AY366514
	Stygocapitella subterranea Knöllner, 1934	AF412810	AY366516
Sabellidae	*Schizobranchia insignis* Bush, 1905	AY732222	AY732225
Serpulidae	*Serpula vermicularis* Linnaeus, 1767	AY732224	AY732227
Sabellaridae	*Sabellaria cementarium* Moore, 1906	AY732223	AY732226
Spionidae	*Polydora ciliata* (Johnston, 1838)	U50971	
	Polydora sp.		DQ790059
	Prionospio dubia Maciolek, 1985	EU418859	EU418867
	Scolecolepis viridis Verrill, 1873	EU418860	EU418868
Trochochaetidae	*Trochochaeta* sp.	DQ790097	DQ790070
Poecilochaetidae	*Poecilochaetus serpens* Allen, 1904	AY569652	EU418869
Siboglinidae	*Riftia pachyptila* Jones, 1981	AF168745	AY210470
	Siboglinum fiordicum Webb, 1963	X79876	DQ790061
Cirratulidae	*Cirratulus spectabilis* (Kinberg, 1866)	AY708536	DQ790029
Ctenodrilidae	*Ctenodrilus serratus* (Schmidt, 1857)	AY364850	AY364864
Fauveliopsidae	*Fauveliopsis scabra* Hartman & Fauchald, 1971	AY708537	DQ790034
Flabelligeridae	*Diplocirrus glaucus* (Malmgren, 1867)	AY708534	DQ790031
	Pherusa plumosa (Müller, 1776)	AY708528	DQ790056
Poeobiidae	*Poeobius meseres* Heath, 1930	AY708526	DQ790058
Sternaspidae	*Sternaspis scutata* (Ranzani, 1817)	AY532329	DQ790063
Ampharetidae	*Auchenoplax crinita* Ehlers, 1887	DQ790077	DQ790026
Pectinariidae	*Pectinaria gouldi* (Verrill, 1873)	DQ790091	DQ790054
Trichobranchidae	*Terebellides stroemi* Sars, 1835	DQ790094	DQ790066
Terebellidae	*Amphitrite ornata* (Leidy, 1855)	DQ790074	DQ790022
	Pista cristata (O. F. Mueller, 1776)	AY611461	DQ790057
	Polycirrus sp.	EU418858	EU418866
Arenicolidae	*Arenicola brasiliensis* (Nonato, 1958)	DQ790076	
	Abarenicola affinis (Ashworth, 1903)		DQ790025
Maldanidae	*Clymenella torquata* (Leidy, 1855)		DQ790030
	Clymenura clypeata (Saint-Joseph, 1894)	AF448152	
	Axiothella rubrocincta (Johnson, 1901)	DQ790078	DQ790027
Echiura	*Arhynchite pugettensis* Fisher, 1949	AY210441	AY210455
	Urechis caupo Fisher & MacGinitie, 1928	AF119076	AF519268
Capitellidae	*Heteromastus filiformis* (Claparede, 1864)	DQ790081	DQ790038
	Notomastus tenuis Moore, 1909	DQ790084	DQ790044
Paranoidae	*Aricidea* sp.	EU418855	DQ790052
Scalibregmatidae	*Scalibregma inflatum* Rathke, 1843	DQ790093	DQ790060
	Travisia brevis Moore, 1923		DQ790069
	Travisia forbesii Johnston, 1840	AF508127	
Opheliidae	*Armandia brevis* (Moore, 1906)	EU418854	EU418862
	Ophelia rathkei McIntosh, 1908	AF448157	AY366513
	Ophelina acuminata Oersted, 1843	DQ790085	DQ790045
Clitellata	*Lumbricus terrestris* Linnaeus, 1758	AJ272183	
	Lumbricus sp.		DQ790041
	Lumbriculus variegatus (Mueller, 1774)	AY040693	
	Lumbriculus sp.		DQ790040
	Eisenia foetida Savigny, 1826	AB076887	
	Eisenia sp.		DQ790032
	Stylaria sp.	U95946	DQ790065
	Erpobdella octoculata (Linnaeus, 1758)	AF099949	AY364865
	Hirudo medicinalis Linnaeus, 1758	Z83752	AY364866

Table 2 (continued)

Taxon	Species	18S	28S
Aeolosomatidae	*Aeolosoma* sp.	Z83748	DQ790019
	Hrabeiella periglandulata Pizl & Chalupsky, 1984	AJ310501	AY364867
Mollusca			
Bivalvia	*Crassostrea virginica* (Gmelin, 1791)	AB064942	AY145400
	Solemya velum Say, 1822	AF120524	AY145421
	Yoldia limaluta (Say, 1831)	AF120528	AY145424
Gastropoda	*Ilyanassa obsoleta* (Say, 1822)	AY145379	AY145411
Polyplacophora	*Chaetopleura apiculata* (Say, 1834)	AY145370	AY145398
Brachiopoda	*Terebratalia transversa* (Sowerby, 1846)	AF025945	AF342802
	Glottidia pyramidata (Stimpson, 1860)	U12647	AY210459
Nemertea	*Cerebratulus lacteus* (Leidy, 1851)	AY145368	AY145396
Entoprocta	*Bartensia gracilis* (Sars, 1835)	AY210442	AY210456
Syndermata	*Brachionus calyciflorus* Pallas, 1766	**XXNNNNNN**	**XXNNNNNN**
	Brachionus urceolaris	DQ089734	DQ089740
	Philodina roseola Ehrenberg, 1832	AF154567	AY210469
Micrognathozoa	*Limnognathia maerski* Kristensen & Funch, 2000	AY218108	AY218135
Gnathostomulida	*Austrognathia christianae* Farris, 1977	DQ079919	DQ079946
	Gnathostomula paradoxa Ax, 1956	DQ079925	**XXNNNNNN**
	Haplognathia ruberrima (Sterrer, 1966)	DQ079930	DQ079954
Playthelminthes	*Dugesia tigrina* (Girard, 1850)	AF013157	U78718
	Schistosoma japonicum (Katsurada, 1904)	AY157226	AY157607
	Stylochus zebra (Verrill, 1892)	AF342801	AF342800
Gastrotricha	*Dactylopodola baltica* (Remane, 1926)	**XXNNNNNN**	**XXNNNNNN**
	Tetranchyroderma megastoma (Remane, 1927)	**XXNNNNNN**	**XXNNNNNN**
	Tetranchyroderma sp.	DQ079911	DQ079939
	Lepidodermella squamata (Dujardin, 1841)	**XXNNNNNN**	**XXNNNNNN**
Arthropoda	*Homarus americanus* H. Milne Edwards, 1837	AF235971	AY859581

3.4. Mitochondrial gene order

The order of the 13 mitochondrial protein-coding and two mitochondrial rRNA genes of *Diurodrilus subterraneus* is very similar to the gene order observed in the majority of Annelida to date (Fig. 4). The CREx analyses showed that the differences between *Diurodrilus* and the majority of annelids can be explained by just three transpositions of rrnL, atp8 and nad1. In contrast, the gene order of the platyzoan or acoel taxa substantially differs from the one observed in *Diurodrilus*. Conserved blocks, which can be found in a platyzoan or acoel species as well as in *Diurodrilus,* are usually blocks, which can also be found in Annelida (i.e., nad2 to cox1, cox1/cox2 or nad4L/nad4, Fig. 4). The most overlap in gene order can be found with Acanthocephala (i.e., *Oncicola* and *Leptorhynchoides*) (Fig. 4, Table 4). However, instead of three transpositions five transpositions would be necessary as indicated by the CREx analyses. For all other platyzoan taxa even more events would be required including excessive tandem duplication and random loss events.

Moreover, the pairwise comparisons of mitochondrial gene orders of the CREx analyses also revealed the closer affinity of *D. subterraneus* and Annelida (Table 4). Whereas both dissimilarity measurements, breakpoint and reversal distance, between *D. subterraneus* and platyzoan taxa is at least 9 the values in the comparisons to Annelida range from 5 to 7, which is similar for example to the values obtained for the comparisons of the two *Urechis* species, *Phascolopsis*, *Phascolosoma* and *Sipunculus* to the each other and the other annelids (Table 4).

Similarly, the common interval values as a similarity measurement are also substantially higher for comparisons of *D. subterraneus* and Annelida than they are for comparisons of *D. subterraneus* and platyzoan and acoel taxa (Table 4).

We also investigated whether the mitochondrial gene order including tRNAs would reveal a potential sistergroup of *D. subterraneus* among the annelid taxa using pairwise comparisons (Table 5). *D. subterraneus* showed lowest dissimilarity to *Endomyzostoma* sp. with values of 7, but the values for the other myzostomid *Myzostoma seymourcollegiorum* are much higher with values of 14 and 13. Similarly, for the low values observed in the sipunculid *Phascolopsis gouldi* are also not reflected in the other two sipunculids, *Phascolosoma esculenta* and *Sipunculus nudus*. Both *Endomyzostoma* sp. and *P. gouldi* are only represented by partial mitochondrial genomes, as is *D. subterraneus* in this study. Thus, the low dissimilarity values obtained in these comparisons are more likely due to shared lack of data.

The highest similarity scores are obtained for the comparisons of *D. subterraneus* with the mitochondrial gene order of the sedentary annelids *Terebellides stroemi*, *Pista cristata*, *Pectinaria gouldi*, *Paralvinella sulfincola* (all four Terebelliformia), *Galathealinum brachiosum*, *Riftia pachyptila* (both Siboglinidae), *Lumbricus terrestris*, *Helobdella robusta*, and *Perionyx excavatus* (all three Clitellata) (Table 5). The dissimilarity scores are also among the lowest values. However, the similarity scores for the errant annelids *Nephtys* sp. and *Platynereis dumerilii* are only slightly worse and the dissimilarity scores for *Nephtys* sp. are even slightly lower. Moreover, the pairwise comparison of *Nephtys* sp. to these nine sedentary annelids (i.e, *Terebellides stroemi* in Table 5) results in both much higher similarity and substantially lower dissimilarity scores. Thus, even using tRNAs *D. subterraneus* cannot be placed confidentially in either Sedentaria or Errantia (Struck, 2011; Struck et al., 2011) using annelid mitochondrial gene orders known to date.

Table 3: Sequence accession numbers of mitochondrial genomes used in this study. Sequences that were newly obtained are indicated in boldface.

Taxon	Species	Accession
Diurodrilidae	*Diurodrilus subterraneus* Remane, 1934	**XXNNNNNN**
Annelida	*Phascolopsis gouldii* (De Pourtalès, 1851)	AF374337
	Phascolosoma arcuatum (Gray, 1828)	NC_012618
	Sipunculus nudus Linnaeus, 1766	FJ422961
	Endomyzostoma sp.	FJ975144
	Myzostoma seymourcollegiorum Rouse & Grygier, 2005	EF506562
	Nephtys sp.	NC_010559
	Platynereis dumerilii Audouin &Milne-Edwards, 1833	NC_000931
	Orbinia latreillii Audouin &Milne-Edwards, 1833	AY961084
	Questa ersei Jamieson & Webb, 1984	FJ612452
	Scoloplos cf. *armiger*	DQ517436
	Galathealinum brachiosum Ivanov, 1961	AF178679
	Riftia pachyptila Jones, 1981	AY741662
	Pectinaria gouldii (Verrill, 1874)	FJ976040
	Pista cristata (Müller, 1776)	EU239688
	Auchenoplax crinita Ehlers, 1887	FJ976041
	Eclysippe vanelli (Fauvel, 1936)	EU239687
	Terebellides stroemi Sars, 1835	EU236701
	Paralvinella sulfincola Desbruyères & Laubier, 1993	FJ976042
	Clymenella torquata (Leidy, 1855)	AY741661
	Urechis caupo Fisher and MacGinitie, 1928	AY619711

Table 3 (continued)

Taxon	Species	Accession
	Urechis unicinctus (von Drasche, 1881)	EF656365
	Lumbricus terrestris Linnaeus, 1758	NC_001673
	Perionyx excavatus Perrier, 1872	NC_009631
	Helobdella robusta Shankland, et al. 1992	AF178680
	Ectoprocta *Bugula neritina* (Linnaeus, 1758)	NC_010197
	Flustra foliacea (Linnaeus, 1758)	NC_016722
	Flustrellidra hispida (Fabricius, 1780)	DQ157889
	Watersipora subtorquata (d'Orbigny, 1852)	NC_011820
Mollusca	*Albinaria caerulea* (Deshayes, 1835)	NC_001761
	Aplysia dactylomela Rang, 1828	NC_015088
	Aplysia vaccaria Winkler, 1955	DQ991928
	Cristaria plicata Leach, 1815	GU944476
	Ilyanassa obsoleta (Say, 1822)	NC_007781
	Graptacme eborea (Conrad, 1846)	AY484748
	Hyriopsis cumingii Lea, 1852	HM347668
	Hyriopsis schlegelii von Martens, 1861	NC_015110
	Katharina tunicata (W. Wood, 1815)	NC_001636
	Loligo opalescens Berry, 1911	GQ225110
	Mytilus galloprovincialis Lamarck, 1819	NC_006886
	Mytilus trossulus Gould, 1850	NC_007687
	Nassarius reticulatus (Linnaeus, 1758)	NC_013248
	Sepia officinalis Linnaeus, 1758	NC_007895
	Sthenoteuthis oualaniensis (Lesson, 1830 in 1830-1831)	NC_010636
Brachiopoda	*Terebratalia transversa* (Sowerby, 1846)	NC_003086
	Laqueus rubellus (Sowerby, 1846)	NC_002322
Phoronida	*Phoronis psammophila* Cori, 1889	AY368231
Nemertea	*Cephalothrix simula* Iwata, 1952	FJ594739
	Cephalothrix sp.	NC_014869
	Emplectonema gracile (Johnston, 1837)	NC_016952
	Lineus viridis (Müller, 1774)	FJ839919
	Nectonemertes cf. *mirabilis*	NC_017874
	Paranemertes cf. *peregrina*	NC_014865
	Zygeupolia rubens (Coe, 1895)	NC_017877
Entoprocta	*Loxocorone allax* Iseto, 2002	NC_010431
	Loxosomella aloxiata Iseto, 2001	AB264800
Syndermata	*Brachionus plicatilis* Mueller, 1786	NC_010472
	Leptorhynchoides thecatus (Linton, 1891)	NC_006892
	Oncicola luehei (Travassos, 1917)	NC_016754
	Rotaria rotatoria (Pallas, 1766)	NC_013568
Playthelminthes	*Dugesia ryukyuensis* Kawakatsu, 1976	AB618488
	Dugesia japonica Ichikawa & Kawakatsu, 1964	AB618487
	Echinococcus multilocularis Leuckart, 1863	NC_000928
	Microstomum lineare (Müller OF, 1773)	AY228756
	Opisthorchis viverrini (Poirier 1886)	JF739555
	Schistosoma mansoni Sambon, 1907	HE601612
	Schistosoma japonicum (Katsurada, 1904)	NC_002544
	Taenia crassiceps (Zeder, 1800)	NC_002547
	Taenia solium (Linnaeus, 1758)	NC_004022
Acoela	*Symsagittifera roscoffensis* (Graff, 1891)	NC_014578
Priapulida	*Priapulus caudatus* Lamarck, 1816	NC_008557
Arthropoda	*Bothropolys* sp.	NC_009458
	Onisimus nanseni G. O. Sars, 1900	FJ555185
	Epiperipatus biolleyi (Bouvier, 1902)	NC_009082

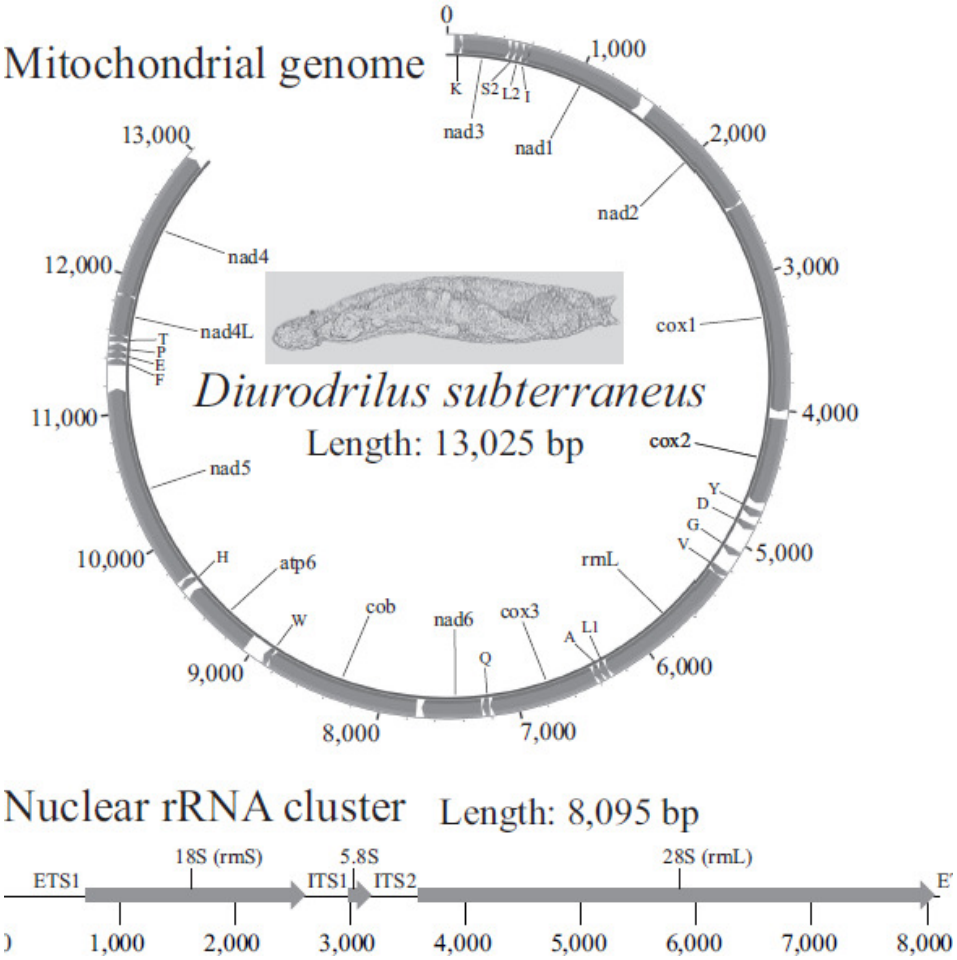


Fig. 1: Gene order of the mitochondrial genome and nuclear rRNA cluster of *Diurodrilus subterraneus*. The circular view of the genome was generated using the CGview webpage (Grant and Stothard, 2008). nad1-6, 4L = NADH dehydrogenase subunits 1-6 and 4L; cox1-3 = cytochrome oxidase subunits 1-3; cob = cytochrome b; atp6 = ATP synthase subunit; rrnS = small ribosomal RNA (also known as 12S); rrnL = large ribosomal RNA (also known as 16S); single letters = mitochondrial tRNAs coding for alanine (A), aspartic acid (D), glutamatic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L1 (CUA), L2(UUA)), proline (P), glutamine (Q), serine (S2(UCA)), threonine (T), valine (V), tryptophan (W), tyrosine (Y); 18S = 18S ribosomal RNA; 28S = 28S ribosomal rRNA; 5.8S = 5.8S ribosomal RNA; ITS1 & 2 = internal transcribed spacer 1 and 2; ETS1 & 2 = external transcribed spacer 1 and 2.

4. Discussion

All our analyses, even the nuclear rRNA analyses, place Diurodrilidae within Annelida and never with platyzoan taxa or as part of a basal radiation in Lophotrochozoa as it was the case in the analyses of Worsaae and Rouse (2008). Moreover, support from both, mitochondrial sequence and gene order data, for a placement of Diurodrilidae within Annelida is very strong and, hence, Diurodrilidae has to be included within Annelida as an annelid subtaxon. Thus, on the basis of these new data exclusion from Annelida and a basal placement within Lophotrochozoa can be rejected.

Presence of peripharyngeal and paired trunk ciliophores as well as of two pairs of protonephridia in an anterior position in Micrognathozoa and Diurodrilidae have been put forward as indication of an affiliation of Diurodrilidae with Micrognathozoa or more generally platyzoan taxa in contrast to annelid affinities of Diurodrilidae (Worsaae and Rouse, 2008).

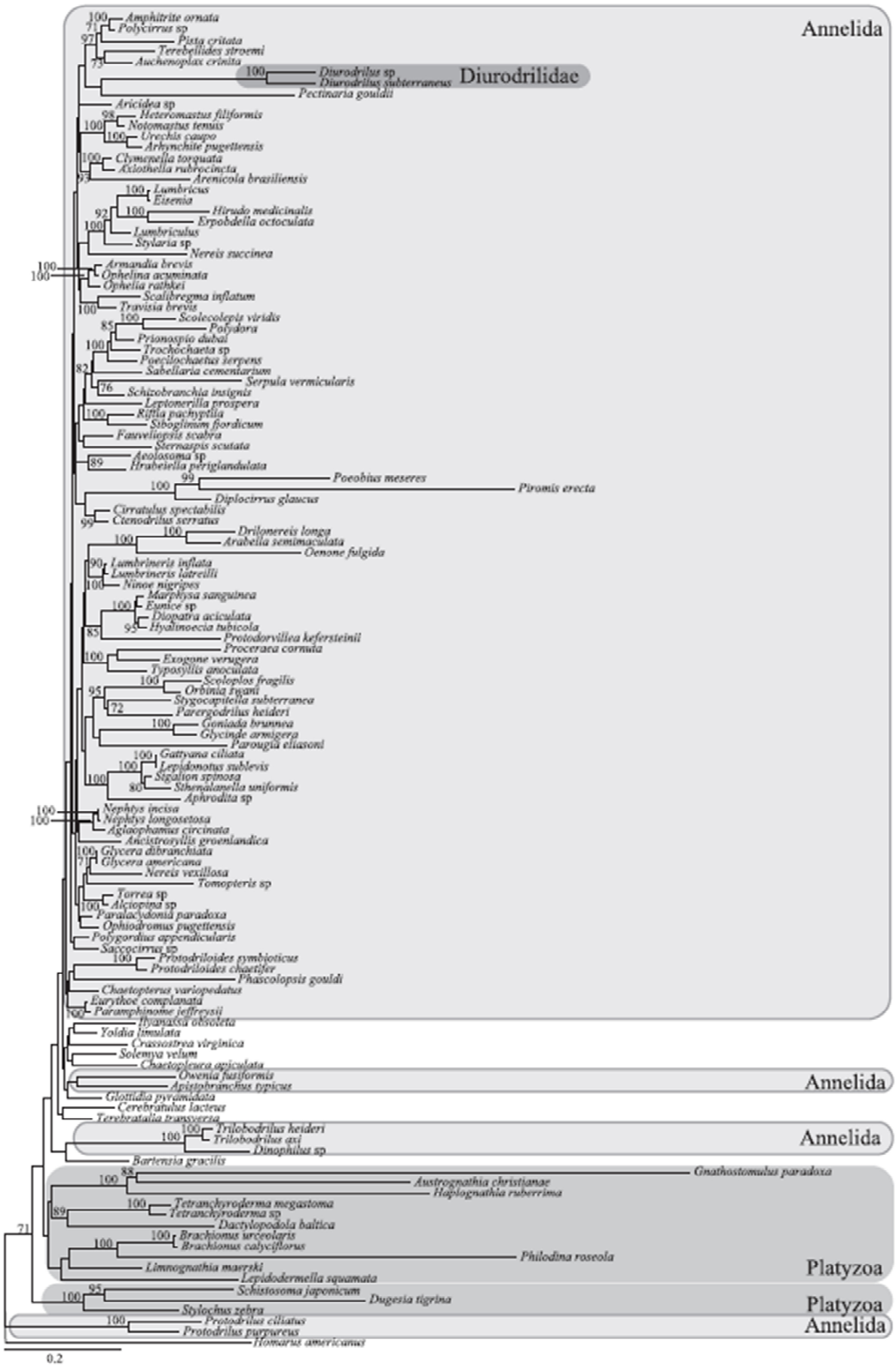

Annelida
Diurodrilidae
Diurodrilus sp
Diurodrilus subterraneus
Pectinaria gouldii
Amphitrite ornata
Polycirrus sp
Pista cristata
Terebellides stroemi
Auchenoplax crinita
Aricidea sp
Heteromastus filiformis
Notomastus tenuis
Urechis caupo
Arhynchite pugettensis
Clymenella torquata
Axiothella rubrocincta
Arenicola brasiliensis
Lumbricus
Eisenia
Hirudo medicinalis
Erpobdella octoculata
Lumbriculus
Stylaria sp
Nereis succinea
Armandia brevis
Ophelina acuminata
Ophelia rathkei
Scalibregma inflatum
Travisia brevis
Scolecolepis viridis
Polydora
Prionospio dubal
Trochochaeta sp
Poecilochaetus serpens
Sabellaria cementarium
Serpula vermicularis
Schizobranchia insignis
Leptonerilla prospera
Riftia pachyptila
Siboglinum fiordicum
Fauveliopsis scabra
Sternaspis scutata
Aeolosoma sp
Hrabeiella periglandulata
Poeobius meseres
Piromis eruca
Diplocirrus glaucus
Cirratulus spectabilis
Ctenodrilus serratus
Drilonereis longa
Arabella semimaculata
Oenone fulgida
Lumbrineris inflata
Lumbrineris latreilli
Ninoe nigripes
Marphysa sanguinea
Eunice sp
Diopatra aciculata
Hyalinoecia tubicola
Protodorvillea kefersteinii
Proceraea cornuta
Exogone verugera
Typosyllis anoculata
Scoloplos fragilis
Orbinia fwani
Stygocapitella subterranea
Parergodrilus heideri
Goniada brunnea
Glycinde armigera
Paroug1a eliasoni
Gattyana ciliata
Lepidonotus sublevis
Sigalion spinosa
Sthenolamella uniformis
Aphrodita sp
Nephtys incisa
Nephtys longosetosa
Aglaophamus circinata
Ancistrosyllis groenlandica
Glycera dibranchiata
Glycera americana
Nereis vexillosa
Tomopteris sp
Torrea sp
Alciopina sp
Paralacydonia paradoxa
Ophiodromus pugettensis
Polygordius appendiculatis
Saccocirrus sp
Protodriloides symbioticus
Protodriloides chaetifer
Phascolopsis gouldi
Chaetopterus variopedatus
Eurythoe complanata
Paramphinome jeffreysii
Hyalascia obsoleta
Yoldia limulata
Crassostrea virginica
Solemya velum
Chaetopleura apiculata
Owenia fusiformis
Apistobranchus typicus
Glottidia pyramidata
Cerebratulus lacteus
Terebratalia transversa
Trilobodrilus heideri
Trilobodrilus axi
Dinophilus sp
Bartensia gracilis
Gnathostomulus paradoxa
Austrognathia christianae
Haplognathia ruberrima
Tetranchyroderma megastoma
Tetranchyroderma sp
Dactylopodola baltica
Brachionus urceolaris
Brachionus calyciflorus
Philodina roseola
Limnognathia maerski
Lepidodermella squamata
Schistosoma japonicum
Dugesia tigrina
Stylochus zebra
Protodrilus ciliatus
Protodrilus purpureus
Homarus americanus
0.2

Fig. 2: Phylogram of the partitioned ML analyses using the nuclear rRNA dataset (4,314 positions, 125 taxa, -ln L = 92,556.01). Only bootstrap values 70 are shown. Taxa highlighted in light grey are Annelida, in medium gray Platyzoa and in dark grey Diurodrilidae.

However, these can also be attributed to convergent adaptations to the interstitial habitat and corresponding decreases in body size as well as similar diets and locomtory patterns. Densely ciliated epidermal cells with long rootlets are widespread in Metazoa and are used for generating currents for various purposes (Hausen, 2005). As in metazoan larvae such cilia are often used for locomotion in interstitial animals (Giere, 2009; Martin, 1978). According to Worsaae and Kristensen (2005) ciliophores differ from other multiciliated cells, which are frequently present in annelids, only by possessing more than ten rows of cilia and arrangements in transverse bands being separated by rows of unciliated cells.

The main morphological characters regarded to disclose inclusion of *Diurodrilus* into Annelida are the lack of segmentation and nuchal organs (Worsaae and Rouse, 2008). However, lack of characters is crucial in phylogenetic reasoning since there are several examples of character losses, which might be misleading in phylogenetic analyses (Bleidorn, 2007; Bleidorn et al., 2009a; Purschke et al., 2000; Struck, 2006). Several annelids lack nuchal organs or possess reduced or modified nuchal organs such as Psammodrilidae, Pisionidae, Siboglinidae, and *Parergodrilus heideri* (Purschke, 1997, 2005).

Presence of a cuticle with similar characteristics as found in other meiofauna annelids like Nerillidae (Hausen, 2005; Purschke, 1985b, 1997; Tzetlin et al., 1992), the organization of the body cavity, the basiepithelial position of the nervous system as well as the structure of the muscular system with the main longitudinal muscle bundles located ventrally (see Worsaae and Rouse, 2008) does not contradict inclusion into Annelida since these characters are present in certain other annelids as well (e.g., Purschke, 2002). Another example is the ventral pharyngeal organ, several types of which have been found in many annelid taxa (Tzetlin and Purschke, 2005). Although clearly differing from the pharynx and the muscle bulbus described for Dinophilidae, and to which *Diurodrilus* formerly was assigned (Kristensen and Niilonen, 1982; Worsaae and Rouse, 2008), the presence of transverse and investing muscle fibers as well as certain areas with an epidermis bearing specialized microvilli and cuticle is in line with pharyngeal organs described in Annelida albeit so far no clear and exclusive correspondences to any specific type of organ can be found (Tzetlin and Purschke, 2005; Worsaae and Rouse, 2008). In the midgut of *Diurodrilus* sp. certain cells probably form a gutter lined by a phalanx of rows of long microvilli (Fig. 9F in Worsaae and Rouse, 2008), the structure of which is similar to the ciliary gutter observed in *Nerillidium troglochaetoides* and *Dinophilus gyrociliatus* (see Oster, 1986; Tzetlin et al., 1992). On the other hand, no morphological characters have been found for clearly forcing or allowing an alternative placement of Diurodrilidae to any other metazoan taxon.

At the end of 19th and beginning of 20th the placement of other interstitial taxa such as Dinophilidae, Polygordiidae or Protodrilida within Annelida had been problematic as well due to lack of chaetae and parapodia, and a clearly visible segmentation. Accordingly, these taxa were regarded as showing the primitive conditions in Annelida and grouped together as "Archiannelida" apart from the other annelids (see Hermans, 1969 for review). Especially interesting in this aspect is Dinophilidae. As Diurodrilidae Dinophilidae seemed to lack nuchal organs and evident internal and external signs of segmentation (Westheide, 1990). The body organization is generally similar between these two taxa, so that Diurodrilidae were first placed within Dinophilidae (Remane, 1925). Moreover, in our nuclear rRNA analyses as well as the ones of Worsaae and Rouse (2008) Dinophilidae are not part of the annelid clade but exclusion of Dinophilidae from Annelida has never been considered. Instead Dinophilidae was regarded as an interstitial annelid taxon of progenetic origin since they generally resemble juvenile stages of some annelid taxa (e.g., Struck, 2006; Westheide, 1987; Westheide, 1990; Worsaae and

Kristensen, 2005). Support for this hypothesis was derived from the finding of nuchal organs in Dinophilidae (see Purschke, 1997) and especially from investigations of the serotonergic part of the central nervous system showing characteristics similar to juvenile stages of other annelid taxa (Müller and Westheide, 2002). Interestingly, the organization of serotonergic part of the ventral nerve cord of Diurodrilidae is not that dramatically different from the one exhibited in Dinophilidae. The ventral nerve cord of both Dinophilidae and Diurodrilidae consists of two main nerves and two paramedian nerves as well as pairs of serotonergic perikarya along the anterior part of the two main nerves, which are connected via thin commissures, and either the number of perikarya decreases in the posterior part in Dinophilidae or are lacking in Diurodrilidae (Figs. 6 & 7 in Müller and Westheide, 2002; Fig. 5A in Worsaae and Rouse, 2008). In contrast to Dinophilidae the perikarya of Diurodrilidae do not correspond to constrictions in the outer body wall. Similar structural correspondences have also been observed in the ventral nerve cord of the interstitial annelid *Parapodrilus psammophilus* (Müller and Westheide, 2002). This species lacks parapodia in the posterior segments and perikarya in the posterior part of the ventral nerve cord. A progenetic origin of *P. psammophilus* within "Dorvilleidae" (Annelida) has been supported by morphological and molecular data (Müller and Westheide, 2002; Struck et al., 2006; Struck et al., 2002).

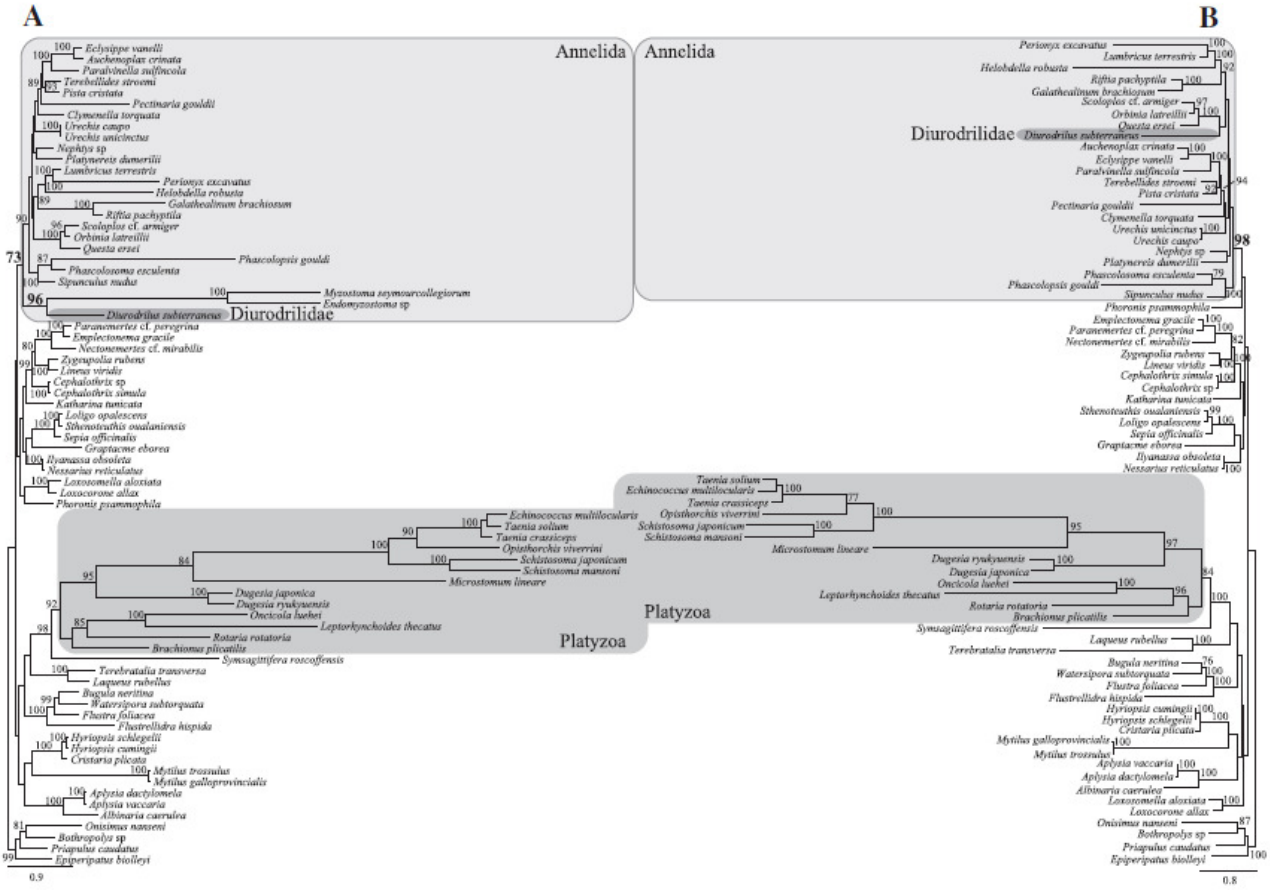


Fig. 3: Phylogram of the partitioned ML analyses using the mitochondrial protein-coding gene dataset coded as amino acids (A) with Myzostomida (2,388 positions, 74 taxa and -ln L = 143,186.24) and (B) without Myzostomida (2,388 positions, 72 taxa and -ln L = 139,608.92). Only bootstrap values 70 are shown. Taxa highlighted in light grey are Annelida, in medium gray Platyzoa and in dark grey Diurodrilidae.

There is also an apparent discrepancy between our molecular analyses and the ones of Worsaae and Rouse (2008) regarding the placement of Diurodrilidae. In the nuclear rRNA analyses of Worsaae and Rouse (2008) Diurodrilidae exhibit a long branch as do the platyzoan taxa.

Therefore, the close phylogenetic association of Platyzoa and Diurodrilidae might be due to long branch attraction. A similar case is known for the annelid taxon Myzostomida (Bleidorn et al., 2007). Myzostomida also shows strongly increased substitution rates and it had been suggested that Myzostomida are more closely related to platyzoan taxa rather than to Annelida (Eeckhaut et al., 2000; Zrzavy et al., 2001; Zrzavy et al., 1998). Sperm ultrastructure was also put forward to support this placement of Myzostomida (Zrzavy et al., 2001). However, several comprehensive morphological studies revealed that morphological data like the nectochaeta larvae, chaetae or presence of cirri strongly support a placement of Myzostomida within Annelida (Eeckhaut and Lanterbecq, 2005; Lanterbecq et al., 2008; Müller and Westheide, 2000). Moreover, analyses of genetic data not affected by increased substitution rates like the mitochondrial genomes, Myosin II heavy chain, hox genes or microRNAs of Myzostomida consistently recovered an annelid affinity of Myzostomida (Bleidorn, 2008; Bleidorn et al., 2007; Bleidorn et al., 2009b; Helm et al., 2012). In our analyses Myzostomida seemed to have increased substitution rates in the mitochondrial genomes as well (Fig. 3A). However, this is due to the fact that partitioned analyses were conducted with the branch length unlinked between each partition. Myzostomida completely lack sequence information for nad1-3 genes and hence missing sequence information in these partitions is compensated by increased branch length for these partitions. This increased branch length is also displayed in the combined information for the complete dataset and is not due to an increased substitution rate. Myzostomida do not exhibit long branches in nonpartitioned analyses of the mitochondrial data herein as well as in previous analyses (e.g., Bleidorn et al., 2007; Mwinyi et al., 2009). The same is true for other taxa such as *Phascolopsis gouldi*, *Pectinaria gouldii* or *Perionyx excavatus* also lacking sequence information for some genes, but necessarily the same ones. In summary, Myzostomida was trapped within Platyzoa in most molecular datasets due to a long-branch artifact, but can be rescued using more slowly evolving data in Myzostomida such as the mitochondrial genome (Bleidorn et al., 2007; Bleidorn et al., 2009c).

Similarly in our analyses Diurodrilidae exhibit long branches in the analyses of the nuclear rRNA data (Fig. 2), but not for the mitochondrial data (Fig. 3). Moreover, as is the case for Myzostomida (Bleidorn et al., 2007) the mitochondrial gene order supports clearly an annelid affinity of Diurodrilidae. Therefore, mitochondrial genomic data was also able to rescue Diurodrilidae from the long branch artifact placing them with Platyzoa instead of Annelida as it was for Myzostomida.

In contrast to Myzostomida our analyses were also able to show the annelid affinity of Diurodrilidae using nuclear rRNA data. This is mostly like due to two reasons. First, in our analyses the taxon sampling, especially of annelid taxa, was substantially increased (125 taxa including 99 annelid and 14 platyzoan taxa vs. 37 taxa including 31 annelid and 6 playtzoan taxa (Worsaae and Rouse, 2008)). It has been observed before that increased taxon sampling of potential sister-taxa can minimize the effects of a long branch attraction in phylogenetic reconstructions (Bergsten, 2005). For example, the phylogenomic study of Dunn et al. (2008) placed Myzostomida with Platyzoa and Acoelomorpha. Increasing the taxon sampling, especially of Acoelomorpha, however, resulted not only a placement of Acoelomorpha as sister to all other bilaterians, but also of Myzostomida with Annelida and not Platyzoa (Hejnol et al., 2009). Second, the close relationship of Diurodrilidae and Micrognathozoa stemmed mostly from the 28S dataset in the analyses of Worsaae and Rouse (2008). However, data coverage of 28S rRNA was low for many crucial taxa (e.g., 316 bp in *Limnognathia maerski* (Micrognathozoa), 1,097 bp in *Tetranchyroderma* sp. (Gastrotricha), 974 bp in *Austrognathia christianae* and 1,037 bp in *Haplognathia ruberrima* (both Gnathostomulida)). Herein, we were able to provide additional 28S data for Gastrotricha, Gnathostomulida and Syndermata. Moreover, of the 3,167 bp of the 28S sequence of *Diurodrilus* sp. 443 positions were not determined. We also presented herein the complete 28S rRNA of *Diurodrilus subterraneus*. Therefore, both the increased taxon representation of potential sister-taxa and increased data

coverage for 28S for many crucial taxa contributed to the result that Diurodrilidae grouped with Annelida even in our 28S rRNA analyses.

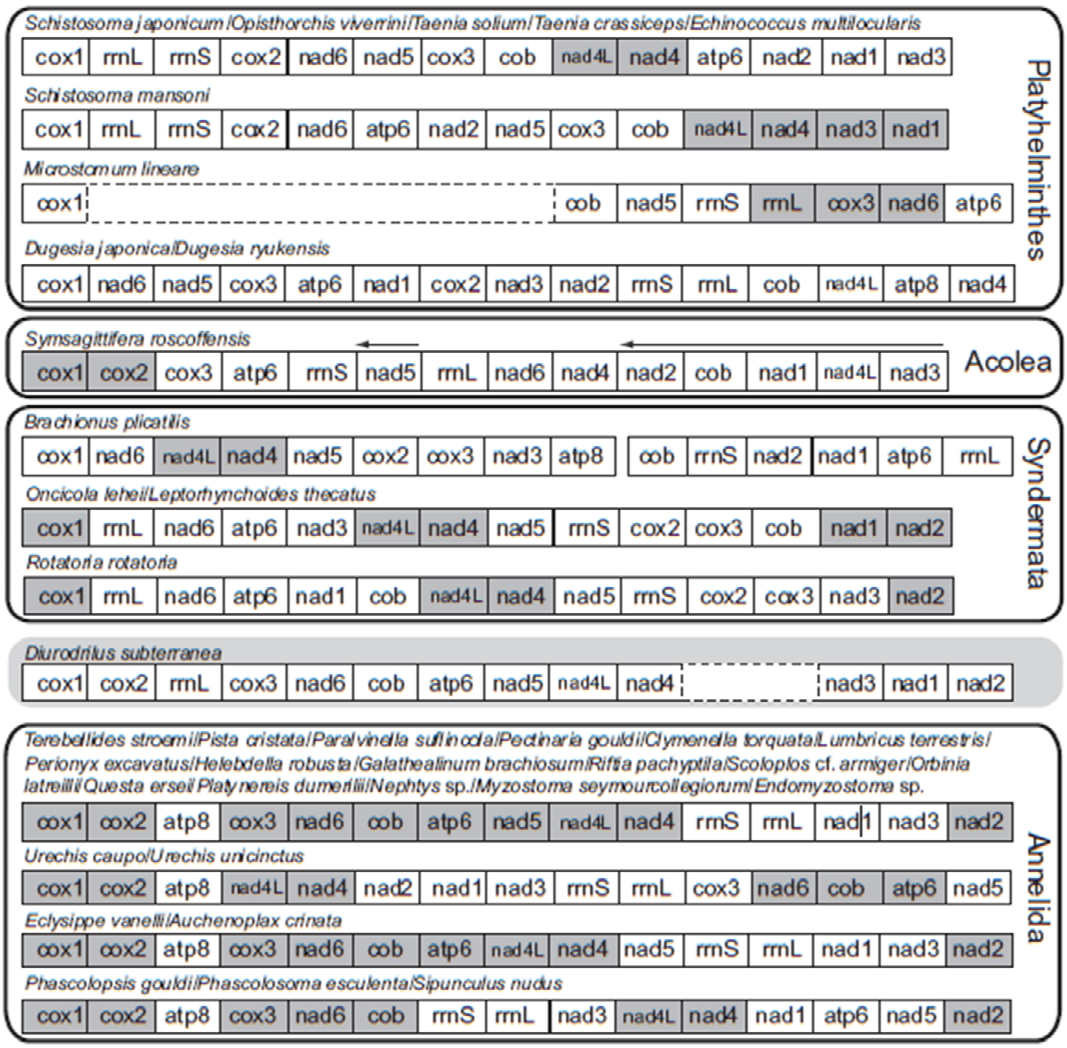


Fig. 4: Comparison of the mitochondrial protein-coding and rRNA gene order of Platyhelminthes, Acoela, Syndermata and Annelida with the gene order of *Diurodrilus subterraneus*. Blocks of identical gene order in *Diurodrilus subterraneus* and another of the gene orders are highlighted in grey. For abbreviations see Fig. 1.

Whereas many morphological characters substantiated the inclusion of Myzostomida within Annelida, Worsaae and Rouse (2008) concluded from their morphological studies that no morphological characters strongly supported the placement of Diurodrilidae within Annelida. This reasoning was mainly based on the lack of the annelid key characters mentioned above. However, this is similar to the situation found in certain other taxa, which were only recently being placed within Annelida. Traditionally, Sipuncula and Echiura had been regarded as phyla separate from Annelida mainly due to the lack of segmentation (Fauchald and Rouse, 1997; Halanych, 2004). However, recent evidence by molecular data strongly supported the inclusion of both within the annelid radiation (e.g., Bleidorn et al., 2006; Bleidorn et al., 2003; Dordel et al., 2010; Dunn et al., 2008; Hejnol et al., 2009; Struck et al., 2011; Struck et al., 2007) and hence it has to be concluded that segmentation had been lost in the stem lineages of these taxa. The segmental ancestry is only visible by very few and subtle traces in the development of the nervous system and not in the adult nervous system organization (Hessling, 2002, 2003; Hessling and Westheide, 2002; Kristof et al., 2008).

Table 4: Results of the CREx analyses of pairwise comparisons of mitochondrial gene orders of platyzoans, annelids and Diurodrilus subterranea. Both dissimilarity [breakpoint (upper value) and reversal distance (middle value)] and similarity analyses [common interval (lower value)] have conducted. Gene orders are in the same order as in Fig. 4 and are indicated by the first species name (abbreviated in the head row) of the corresponding gene order in Fig. 4.

	S. j.	*S. m.*	*D. j.*	*S. r.*	*B. p.*	*O. l.*	*R. r.*	*D. s.*	*T. s.*	*U. c.*	*E. v.*	*P. g.*
Schistosoma japonica		5	10	14	12	10	10	**12**	12	11	12	13
		5	9	13	11	7	7	**11**	11	9	11	11
		64	14	2	8	12	12	**20**	18	10	22	4
Schistosoma mansoni	5		11	14	13	9	9	**10**	13	13	13	13
	5		9	13	11	7	7	**9**	11	11	11	11
	64		10	0	4	14	14	**18**	6	6	6	12
Dugesia japonica	10	11		13	14	13	10	**12**	13	14	13	14
	9	9		10	13	13	9	**11**	11	13	13	13
	14	10		2	8	2	10	**2**	4	6	4	6
Symsagittifera roscoffensis	14	14	13		13	11	11	**12**	12	13	12	11
	13	13	10		12	9	9	**10**	11	12	11	9
	2	0	2		2	12	10	**6**	10	6	14	12
Brachionus plicatilis	12	13	14	13		11	10	**12**	14	13	13	12
	11	11	13	12		11	9	**11**	13	11	11	11
	8	4	8	2		18	24	**6**	4	4	6	6
Oncicula lehei	10	9	13	11	11		5	**10**	11	13	9	10
	7	7	13	9	11		5	**9**	9	11	7	9
	12	14	2	12	18		66	**18**	10	4	16	10
Rotatoria rotatoria	10	9	10	11	10	5		**11**	10	13	8	11
	7	7	9	9	9	5		**11**	9	11	7	7
	12	14	10	10	24	66		**6**	18	2	24	8
Diurodrilus subterranea	**12**	**10**	**12**	**12**	**12**	**10**	**11**		**5**	**6**	**7**	**7**
	11	**9**	**11**	**10**	**11**	**9**	**11**		**5**	**5**	**7**	**7**
	20	**18**	**2**	**6**	**6**	**18**	**6**		**60**	**46**	**42**	**32**
Terebellides stroemi	12	13	13	12	14	11	10	**5**		6	3	6
	11	11	11	11	11	9	9	**5**		5	3	5
	18	6	4	10	4	10	18	**60**		46	154	60
Urechis caupo	11	13	14	13	13	13	13	**6**	6		7	8
	9	11	13	12	11	11	11	**5**	5		7	7
	10	6	6	6	4	4	2	**46**	46		34	28
Eclysippe vanelli	12	13	13	12	13	9	8	**7**	3	7		7
	11	11	13	11	11	7	7	**7**	3	7		7
	22	6	4	14	6	16	24	**42**	154	34		56
Phascololepsis gouldi	13	13	14	11	12	10	11	**7**	6	8	7	
	11	11	13	9	11	9	7	**7**	5	7	7	
	4	12	6	12	6	10	8	**32**	60	28	56	

Table 5: Results of the CREx analyses of pairwise comparisons of mitochondrial gene orders including tRNAs of annelids and *Diurodrilus subterranea*. Both dissimilarity [breakpoint (upper value) and reversal distance (middle value)] and similarity analyses [common interval (lower value)] have conducted.

	T. s.	*E. v.*	*U. c.*	*C. t.*	*O. l.*	*Q. e.*	*D. s.*	*E.* sp.	*M. s.*	*P. d.*	*N.* sp.	*P. g.*	*P. e.*	*S. n.*
Terebellides stromi [1]		11 9 184	18 15 162	3 3 644	10 10 308	9 9 296	**10** **10** **246**	11 11 382	16 16 76	12 12 358	3 3 656	5 5 174	24 21 44	11 9 166
Eclysippe vanelli [2]	11 9 184		23 21 60	11 9 184	19 17 88	15 13 80	**15** **13** **62**	16 15 56	16 15 52	20 18 96	13 11 166	7 7 92	28 27 24	16 15 104
Urechis caupo [3]	18 15 162	23 21 60		21 17 90	23 23 74	18 18 82	**15** **13** **114**	17 14 84	16 15 56	22 22 148	18 17 162	15 15 20	26 25 38	23 21 54
Clymenella torquata	3 3 644	11 9 184	21 17 90		12 12 244	8 8 252	**12** **12** **162**	11 11 382	16 16 76	15 14 176	6 5 290	5 5 174	27 23 30	14 11 148
Orbinia latreilli [4]	10 10 308	19 17 88	23 23 74	12 12 244		3 3 866	**15** **13** **126**	11 11 424	15 14 160	15 13 146	9 9 294	7 7 92	28 25 28	18 17 86
Questa ersei	9 9 296	15 13 80	18 18 82	8 8 252	3 3 866		**14** **13** **124**	6 6 378	11 10 176	12 11 290	9 8 266	7 7 72	25 21 32	17 15 72
Diurodrilus subterranea	**10** **10** **246**	**15** **13** **62**	**15** **13** **114**	**12** **12** **162**	**15** **13** **126**	**14** **13** **124**		**7** **7** **196**	**14** **13** **48**	**11** **11** **214**	**9** **9** **242**	**9** **8** **52**	**21** **21** **32**	**16** **15** **62**
Endomyzostoma sp	11 11 382	16 15 56	17 14 84	11 11 382	11 11 424	6 6 378	**7** **7** **196**		7 7 226	11 11 346	11 11 382	8 8 134	17 13 32	17 16 64
Myzostoma seymourcollegiorum	16 16 76	16 15 52	16 15 56	16 16 76	15 14 160	11 10 176	**14** **13** **48**	7 7 226		15 15 74	15 14 160	8 8 70	20 17 20	17 16 52
Platynereis dumerilii	12 12 358	20 18 96	22 22 148	15 14 176	15 13 146	12 11 290	**11** **11** **214**	11 11 346	15 15 74		12 10 354	11 11 44	27 25 34	19 17 70
Nephtys sp	3 3 656	13 11 166	18 17 162	6 5 290	9 9 294	9 8 266	**9** **9** **242**	11 11 382	15 14 160	12 10 354		8 7 72	25 23 40	13 11 144
Phascolopsis gouldi	5 5 174	7 7 92	15 15 20	5 5 174	7 7 92	7 7 72	**9** **8** **52**	8 8 134	8 8 70	11 11 44	8 7 72		13 11 38	3 3 182
Phascolosoma esculenta	24 21 44	28 27 24	26 25 38	27 23 30	28 25 28	25 21 32	**21** **21** **32**	17 13 32	20 17 20	27 25 34	25 23 40	13 11 38		14 13 260
Sipunculus nudus	11 9 166	16 15 104	23 21 54	14 11 148	18 17 86	17 15 72	**16** **15** **62**	17 16 64	17 16 52	19 17 70	13 11 144	3 3 182	14 13 260	

[1] *Galathealinum brachiosum*, *Riftia pachyptila*, *Pista cristata*, *Pectinaria gouldi*, *Paralvinella sulfincola*, *Lumbricus terrestris*, *Helobdella robusta*, and *Perionyx excavatus* have an identical gene order.
[2] *Auchenoplax crinita* has an identical gene order.
[3] *Urechis unicinctus* has an identical gene order.
[4] *Scoloplos armiger* has an identical gene order.

Although Worsaae and Kristensen (2005) denied a progenetic origin of *Diurodrilus*, according to the clear placement in Annelida this possibility has to be reconsidered. Progenetic evolution allows explaining the lack of annelid key characters such as chaetae, segmentation and coelom together with the occurrence of a non-coelomic body cavity, protonephridia, and other characters discussed above. Thus, as for other interstitial annelids Diurodrilidae most likely originated by progenetic evolution from juvenile stages of a larger annelid ancestor. However, in accordance with Worsaae and Kristensen (2005) and Worsaae and Rouse (2008) there is no evidence for a dinophilid relationship.

5. Acknowledgments

We would like to thank L. Becks (Cornell University) and B. Rothe (University of Bielefeld) for providing and identifying specimens. We also thank to anonymous reviewers for helpful comments. This study was funded by DFG STR-683/5-2 as part of the priority program 1174 „Deep Metazoan Phylogeny“ as well as DFG STR-683/6-1, 6-2 and 8-1 of the Deutsche Forschungsgemeinschaft to THS.

6. References

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163-193.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., in press. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol.

Bernt, M., Merkle, D., Ramsch, K., Fritzsch, G., Perseke, M., Bernhard, D., Schlegel, M., Stadler, P., Middendorf, M., 2007. CREx: Inferring Genomic Rearrangements Based on Common Intervals. Bioinformatics 23, 2957-2958.

Bleidorn, C., 2007. The role of character loss in phylogenetic reconstruction as exemplified for the Annelida. J. Zool. Syst. Evol. Res. 45, 299–307.

Bleidorn, C., 2008. Lophotrochozoan relationships and parasites. A snap-shot. Parasite 15, 329-332.

Bleidorn, C., Eeckhaut, I., Podsiadlowski, L., Schult, N., McHugh, D., Halanych, K.M., Milinkovitch, M.C., Tiedemann, R., 2007. Mitochondrial genome and nuclear sequence data support Myzostomida as part of the annelid radiation. Mol. Biol. Evol. 24, 1690-1701.

Bleidorn, C., Hill, N., Ersèus, C., Tiedemann, R., 2009a. On the role of character loss in orbiniid phylogeny (Annelida): Molecules vs. morphology. Mol. Phylogenet. Evol. 52, 57-69.

Bleidorn, C., Lanterbecq, D., Eeckhaut, I., Tiedemann, R., 2009b. A PCR survey of Hox genes in the myzostomid *Myzostoma cirriferum*. Dev. Genes Evol. 219, 211-216.

Bleidorn, C., Podsiadlowski, L., Bartolomaeus, T., 2006. The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreilli* (Annelida, Orbiniidae) - a novel gene order for Annelida and implications for annelid phylogeny. Gene 370, 96-103.

Bleidorn, C., Podsiadlowski, L., Zhong, M., Eeckhaut, I., Hartmann, S., Halanych, K., Tiedemann, R., 2009c. On the phylogenetic position of Myzostomida: can 77 genes get it wrong? BMC Evol. Biol. 9, 150.

Bleidorn, C., Vogt, L., Bartolomaeus, T., 2003. New insights into polychaete phylogeny (Annelida) inferred from 18S rDNA sequences. Mol. Phylogenet. Evol. 29, 279-288.

Dordel, J., Fisse, F., Purschke, G., Struck, T.H., 2010. Phylogenetic position of Sipuncula derived from multi-gene and phylogenomic data and its implication for the evolution of segmentation. J. Zool. Syst. Evol. Res. 48, 197–207.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745-750.

Eeckhaut, I., Lanterbecq, D., 2005. Myzostomida: A review of the phylogeny and ultrastructure. Hydrobiologia V535-536, 253-275.

Eeckhaut, I., McHugh, D., Mardulyn, P., Tiedemann, R., Monteyne, D., Jangoux, M., Milinkovitch, M.C., 2000. Myzostomida: a link between trochozoans and flatworms? Proc. R. Soc. London, B 267, 1383 - 1392.
Fauchald, K., 1974. Polychaete Phylogeny: A Problem in Protostome Evolution. Syst. Zool. 23, 493-506.
Fauchald, K., Rouse, G., 1997. Polychaete systematics: Past and present. Zool. Scr. 26, 71-138.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783-791.
Giere, O., 2009. Meiobenthology - The microscopic motile fauna of aquatic sediments. Springer-Verlag, Berlin Heidelberg.
Grant, J.R., Stothard, P., 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36, W181-W184.
Halanych, K.M., 2004. The new view of animal phylogeny. Annu. Rev. Ecol. Evol. Syst. 35, 229-256.
Halanych, K.M., Janosik, A.M., 2006. A review of molecular markers used for Annelid phylogenetics. Integr. Comp. Biol. 46, 533-543.
Hausen, H., 2005. Comparative structure of the epidermis in polychaetes (Annelida). Hydrobiologia V535-536, 25-35.
Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E.G., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W., 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. R. Soc. B: Biol. Sci. 276, 4261-4270.
Helm, C., Bernhart, S.H., Siederdissen, C.H.n.z., Nickel, B., Bleidorn, C., 2012. Deep sequencing of small RNAs confirms an annelid affinity of Myzostomida. Mol. Phylogenet. Evol. 64, 198-203.
Hermans, C.O., 1969. The systematic position of the Archiannelida. Syst. Zool. 18, 85-102.
Hessling, R., 2002. Metameric organisation of the nervous system in developmental stages of *Urechis caupo* (Echiura) and its phylogenetic implications. Zoomorphology 121, 221-234.
Hessling, R., 2003. Novel aspects of the nervous system of *Bonellia viridis* (Echiura) revealed by the combination of immunohistochemistry, confocal laser-scanning microscopy and threedimensional reconstruction. Hydrobiologia 496, 225-239.
Hessling, R., Westheide, W., 2002. Are Echiura derived from a segmented ancestor? Immunohistochemical analysis of the nervous system in developmental stages of *Bonellia viridis*. J. Morphol. 252, 100-113.
Katoh, K., Kuma, K.-i., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518.
Kristensen, R.M., Niilonen, T., 1982. Structural evidence on *Diurodrilus* Remane (Diurodrilidae fam.n.), with description of *Diurodrilus westheidei* sp.n. from the artic interstitial meiobenthos, W. Greenland. Zool. Scr. 11, 1-12.
Kristof, A., Wollesen, T., Wanninger, A., 2008. Segmental mode of neural patterning in Sipuncula. Curr. Biol. 18, 1129-1132.
Kück, P., Meusemann, K., 2010. FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. 56, 1115-1118.
Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front. Zool. 7, 10.
Lanterbecq, D., Bleidorn, C., Michel, S., Eeckhaut, I., 2008. Locomotion and fine structure of parapodia in *Myzostoma cirriferum* (Myzostomida). Zoomorphology 127, 59-68.
Martin, G.G., 1978. Ciliary gliding in lower invertebrates. Zoomorphologie 91, 249-261.
Misof, B., Misof, K., 2009. A Monte Carlo Approach Successfully Identifies Randomness in Multiple Sequence Alignments : A More Objective Means of Data Exclusion. Syst. Biol. 58, 21-34.
Müller, M.C.M., Westheide, W., 2000. Structure of the nervous system of *Myzostoma cirriferum* (Annelida) as revealed by immunohistochemistry and cLSM analyses. J. Morphol. 245, 87-98.
Müller, M.C.M., Westheide, W., 2002. Comparative analysis of the nervous systems in presumptive progenetic dinophilid and dorvilleid polychaetes (Annelida) by immunohistochemistry and cLSM. Acta Zool. 83, 33 48.
Mwinyi, A., Meyer, A., Bleidorn, C., Lieb, B., Bartolomaeus, T., Podsiadlowski, L., 2009. Mitochondrial genome sequence and gene order of *Sipunculus nudus* give additional support for an inclusion of Sipuncula into Annelida. BMC Genomics 10, 27.
Nicholas, K.B., Nicholas, H.B.j., 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author.
Oster, U., 1986. Microfilament-supported macrovilli in the hindgut of the polychaete *Dinophilus gyrociliatus*. Z. Naturforsch., C: Biosci. 41, 1139-1143.
Paxton, H., 2000. Family Diurodrilidae. In: Beesly, P.L., Ross, G.J., Glasby, C.J. (Eds.), Polychaetes & Allies: The Southern Synthesis. Fauna of Australia. Vol. 4A Polychaeta, Myzostomida, Pogonophora, Echiura, Sipuncula. CSIRO Publishing, Melbourne, pp. 104-105.

Purschke, G., 1985a. Anatomy and ultrastructure of ventral pharyngeal organs and their phylogenetic importance in Polychaeta (Annelida) - I. The pharynx of the Dinophilidae. Zoomorphology 105, 223-239.
Purschke, G., 1985b. Anatomy and ultrastructure of ventral pharyngeal organs and their phylogenetic importance in Polychaeta (Annelida) - II. The pharynx of the Nerillidae. Mikrofauna Mar. 2, 23-60.
Purschke, G., 1997. Ultrastructure of nuchal organs in polychaetes (Annelida): new results and review. Acta Zool. 78, 123-143.
Purschke, G., 2002. On the ground pattern of Annelida. Org. Divers. Evol. 2, 181-196.
Purschke, G., 2005. Sense organs in polychaetes (Annelida). Hydrobiologia V535-536, 53-78.
Purschke, G., Hessling, R., Westheide, W., 2000. The phylogenetic position of the Clitellata and the Echiura – on the problematic assessment of absent characters. J. Zool. Syst. Evol. Res. 38, 165-173.
Purschke, G., Jouin, C., 1988. Anatomy and ultrastructure of the ventral pharyngeal organs of *Saccocirrus* (Saccocirridae) and *Protodriloides* (Protodriloidae fam. n.) with remarks on the phylogenetic relationships within Protodrilida (Annelida: Polychaeta). J. Zool. 215, 405-432.
Remane, A., 1925. Diagnosen neuer Archianneliden (zugleich 3. Beitrag zur Fauna der Kieler Bucht). Zool. Anz. 65, 15-17.
Riser, N.W., 1984. General observations on the intertidal interstitial fauna of New Zealand. Tane 30, 239-250.
Rouse, G.W., Fauchald, K., 1997. Cladistics and polychaetes. Zool. Scr. 26, 139-204.
Stamatakis, A., 2006. RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics 22, 2688–2690.
Struck, T.H., 2006. Progenetic species in polychaetes (Annelida) and problems assessing their phylogenetic affiliation. Integr. Comp. Biol. 46, 558–568.
Struck, T.H., 2011. Direction of evolution within Annelida and the definition of Pleistoannelida. J. Zool. Syst. Evol. Res. 49, 340-345.
Struck, T.H., Halanych, K.M., Purschke, G., 2005. Dinophilidae (Annelida) is most likely not a progenetic Eunicida: Evidence from 18S and 28S rDNA. Mol. Phylogenet. Evol. 37, 619–623.
Struck, T.H., Nesnidal, M.P., Purschke, G., Halanych, K.M., 2008. Detecting possibly saturated positions in 18S and 28S sequences and their influence on phylogenetic reconstruction of Annelida (Lophotrochozoa). Mol. Phylogenet. Evol. 48, 628-645.
Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., Bleidorn, C., 2011. Phylogenomic analyses unravel annelid evolution. Nature 471, 95–98.
Struck, T.H., Purschke, G., Halanych, K.M., 2006. Phylogeny of Eunicida (Annelida) and Exploring Data Congruence using a Partition Addition Bootstrap Alteration (PABA) approach. Syst. Biol. 55, 1-20.
Struck, T.H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., Halanych, K.M., 2007. Annelida phylogeny and the status of Sipuncula and Echiura. BMC Evol. Biol. 7, 57.
Struck, T.H., Westheide, W., Purschke, G., 2002. Progenesis in Eunicida ("Polychaeta," Annelida) – separate evolutionary events? Evidence from molecular data. Mol. Phylogenet. Evol. 25, 190-199.
Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, W609-W612.
Tzetlin, A., Purschke, G., 2005. Pharynx and intestine. Hydrobiologia 535-536, 199-225.
Tzetlin, A.B., Purschke, G., Westheide, W., Saphonov, M.V., 1992. Ultrastructure of th enteronephridia and general description of the alimentary canal in *Trochonerilla mobilis* and *Nerillidium troglochaetoides* (Polychaeta, Nerillidae). Acta Zool. 73, 163-176.
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 35, W71-W74.
Vallès, Y., Boore, J.L., 2006. Lophotrochozoan mitochondrial genomes. Integr. Comp. Biol. 46, 544-557.
Westheide, W., 1985. The systematic position of the Dinophildae and the archiannelid problem. In: Morris, S.C., George, J.D., Gibson, R., Platt, H.M. (Eds.), The origins and relationships of lower invertebrates. The Systematics Association, Oxford, pp. 310-326.
Westheide, W., 1987. Progenesis as a principle in meiofauna evolution. J. Nat. Hist. 21, 843-854.
Westheide, W., 1990. Polychaetes: Interstitial families. Universal Book Services, Oegstgeest.
Worsaae, K., Kristensen, R.M., 2005. Evolution of interstitial Polychaeta (Annelida). Hydrobiologia 535-536, 319-340.
Worsaae, K., Rouse, G.W., 2008. Is *Diurodrilus* an annelid? J. Morphol. 269, 1426-1455.
Zhong, M., Hansen, B., Nesnidal, M.P., Golombek, A., Halanych, K.M., Struck, T.H., 2011. Detecting the symplesiomorphy trap: A multigene phylogenetic analysis for terebelliform annelids. BMC Evol. Biol. 11, 369.
Zrzavy, J., Hypsa, V., Tietz, D.F., 2001. Myzostomida Are Not Annelids: Molecular and Morphological Support for a Clade of Animals with Anterior Sperm Flagella. Cladistics 17, 170-198.
Zrzavy, 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.

5.7. Evolution of mitochondrial gene order in Annelida

Anne Weigert[a,b*], Anja Golombek[c], Michael Gerth[a], Francine Schwarz[a], Torsten H. Struck[c,$] and Christoph Bleidorna[d,$]

Author affiliation:
[a] *Molecular Evolution and Animal Systematics, University of Leipzig, Talstr. 33, 04103 Leipzig, Germany*
[b] *Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany*
[c] *Zoological Research Museum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany*
[d] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*
[$] shared senior authors

[*] Corresponding author: Anne Weigert, Molecular Evolution and Animal Systematics, University of Leipzig, Talstr. 33, 04103 Leipzig, Germany; anne.weigert@unileipzig.de; Phone: +49-341-9736743; Fax: +49-341-9736789

Email addresses:
anne.weigert@uni-leipzig.de (Anne Weigert), A.Golombek@gmx.de (Anja Golombek), michael.gerth@uni-leipzig.de (Michael Gerth), francine.schwarz@yahoo.de (Francine Schwarz), torsten.struck.zfmk@uni-bonn.de (Torsten H. Struck), bleidorn@ uni-leipzig.de (Christoph Bleidorn)

Abstract

Annelida is a highly diverse animal group with over 21,000 described species. As part of Lophotrochozoa, the vast majority of annelids are currently classified into two groups: Errantia and Sedentaria, together forming Pleistoannelida. Besides these taxa, Sipuncula, Amphinomidae, Chaetopteridae, Oweniidae and Magelonidae can be found branching at the base of the tree. Comparisons of mitochondrial genomes have been used to investigate phylogenetic relationship within animal taxa. Complete annelid mitochondrial genomes are available for some Sedentaria and Errantia and in most cases exhibit a highly conserved gene order. Only two complete genomes have been published from the basal branching lineages and these are restricted to Sipuncula. We describe the first complete mitochondrial genome sequences for all other basal branching annelid families: *Owenia fusiformis* (Oweniidae), *Magelona mirabilis* (Magelonidae), *Eurythoe complanata* (Amphinomidae), *Chaetopterus variopedatus* and *Phyllochaetopterus* sp. (Chaetopteridae). The mitochondrial gene order of all these taxa is substantially different from the pattern found in Pleistoannelida. Additionally, we report the first mitochondrial genomes in Annelida that encode genes on both strands. Our findings demonstrate that the supposedly highly conserved mitochondrial gene order suggested for Annelida is restricted to Pleistoannelida, representing the ground pattern of this group. All investigated basal branching annelid taxa show a completely different arrangement of genes than observed in Pleistoannelida. The gene order of protein coding and ribosomal genes in *Magelona mirabilis* differs only in two transposition events from a putative lophotrochozoan ground pattern and might be the closest to an ancestral annelid pattern. The mitochondrial genomes of Myzostomida show the conserved pattern of Pleistoannelida, thereby supporting their inclusion in this taxon.

1. Introduction

Annelida is a major phylum within Lophotrochozoa, whose members occupy a broad range of habitats and are especially abundant in marine environments. This group shows

a high diversity in life modes, feeding and reproductive strategies, body forms and developmental patterns (Rouse and Pleijel, 2001). Until recently, relationships among annelid groups were poorly understood, but previous phylogenomic analyses resolved a robust annelid backbone and recovered two major groups comprising the major diversity of Annelida: Errantia and Sedentaria (Andrade et al., 2015; Struck et al., 2015; Struck et al., 2011; Weigert et al., 2014). Additionally, five groups, which are morphological extremely diverse from each other could be found outside of Pleistoannelida: Sipuncula, Amphinomidae, Chaetopteridae, Magelonidae and Oweniidae (Weigert et al., 2014).

To extend our knowledge on annelid evolution and phylogenetic relationships among them, investigation and comparison of mitochondrial gene arrangements is a powerful tool, since rearrangements rarely occur independently in different lineages and closely related species often share identical unchanged gene orders (Boore, 1999; Boore and Brown, 1994). In animals, mitochondrial genomes are usually circular molecules (except in e.g. cnidarians (Bridge et al., 1992) and sponges (Lavrov et al., 2013)), generally around 16 kb in size, possess only limited intergenic sequences apart from one large non-coding region which is correlated with the origin of replication, and encode for 13 protein-coding genes (PCG), 2 ribosomal RNAs and 22 transfer RNAs (Boore, 1999; Clary and Wolstenholme, 1984; Shadel and Clayton, 1997). The 37 genes can be transcribed either on both strands of the genome or on only one strand.

Lophotrochozoa show a high variability in mitochondrial genomes, including gene number and gene arrangements, strand usage for transcription, repetitive and intergenic regions and unusual modes of inheritance (Boore, 1999; Valles and Boore, 2006; Valles et al., 2008). The mitochondrial gene order in annelids is, unlike in other lophotrochozoan groups, fairly conserved for the families for which gene order has been described so far, especially when not taking tRNA translocations into account (e.g. members of Clitellata, Terebelliformia, Orbiniidae and Phyllodocidae). Exceptions are Sipuncula, Echiura, Ampharetidae, Diurodrilidae and Eunicidae, even though they differ only in a few rearranged genes (or blocks of genes) from the putative annelid ground pattern (Bleidorn et al., 2006; Boore, 2004; Golombek et al., 2013; Jennings and Halanych, 2005; Li et al., 2014; Mwinyi et al., 2009; Shen et al., 2009; Valles and Boore, 2006; Zhong et al., 2008). The only taxon so far completely deviating from this pattern are Syllidae (Errantia), which show completely rearranged mitochondrial genomes (Aguado et al., 2015). Nevertheless, up to now gene rearrangements within Annelida have occurred less often than in other Lophotrochozoa (Boore, 2004; Jennings and Halanych, 2005; Noguchi et al., 2000; Osca et al., 2014; Stechmann and Schlegel, 1999). Additionally, for all annelids from which data is available, genes are described only on one strand of the genome.

To further investigate the putatively conserved mitochondrial gene order evolution in Annelida and to draw a comparison to other lophotrochozoan gene orders, it is crucial to cover mitochondrial genomes from all major annelid groups. So far, ~40 complete mitochondrial genomes are available for annelids covering mainly species in Sedentaria (with 12 of them from clitellates and 10 from Siboglinidae) and species in Errantia. Representing the basal branching lineages, only two complete mitochondrial genomes of Sipuncula species are published (Mwinyi et al., 2009; Shen et al., 2009). In summary, while the majority of annelid taxa in both Sedentaria and Errantia are still not represented, the coverage is much better than for the basal branching lineages, for which there is actually almost no data. Additional information from those basal lineages would provide more insights into mitochondrial genome rearrangements within annelids and help to determine the mitochondrial gene order ground pattern of Annelida.

In this study five new mitochondrial genomes from basal branching annelid families were generated using Illumina-based whole genome shotgun sequencing. Together with the already available mitochondrial genomes of Sipuncula, we covered the complete base of the annelid tree with the taxa *Owenia fusiformis* (Oweniidae), *Magelona mirabilis* (Magelonidae), *Chaetopterus variopedatus* and *Phyllochaetopterus* sp. (Chaetopteridae), and *Eurythoe*

complanata (Amphinomidae). Using these data, we investigated the evolution of gene order arrangements in annelids. Moreover, we performed phylogenetic analyses to compare relationships within Annelida inferred by mitochondrial data with the current phylogeny based on transcriptomic data (Andrade et al., 2015; Struck et al., 2015; Struck et al., 2011; Weigert et al., 2014). Our data clearly show a higher variability in mitochondrial gene arrangements in the basal branching lineages in comparison to other annelids and provide additional insights into a putative ancestral mitochondrial gene order pattern for Annelida and Pleistoannelida. Interestingly, *Owenia fusiformis* (Oweniidae) and *Magelona mirabilis* (Magelonidae), representing the lineages which together form the sister taxon of all other annelids, are the only annelids described so far with genes transcribed on both strands of the mitochondrial genome. Especially *Magelona mirabilis* shares a gene order pattern with lophotrochozoans outgroups, which we regard as plesiomorphic. The hitherto reported conserved pattern of Annelida is supported as the ancestral condition for Pleistoannelida (Sedentaria + Errantia).

2. Material and Methods

2.1. Taxonomic sampling

Representatives from all basal branching annelid groups were selected according to Weigert et al. (2014). Specimens of *Magelona mirabilis* and *Chaetopterus variopedatus* were collected in Morgat (France), *Owenia fusiformis* in Helgoland (Germany), *Phyllochaetopterus* sp. in Southern New England (USA) and *Eurythoe complanata* was obtained from bought live rock of the Indian ocean kept in the aquarium in Leipzig (Germany). Data for additional annelid families and lophotrochozoan groups were extracted from public resources. Species sampling and accession numbers of all sequences are given in Table 1.

2.2. Library construction, sequencing and raw data processing

Genomic DNA was extracted from a single individual by proteinase K digestion followed by the standard phenol-chloroform extraction (Gustincich et al., 1991). For *Magelona mirabilis*, *Chaetopterus variopedatus* and *Eurythoe complanata*, doubleindexed libraries with an average insert size of 350 bp were prepared as described in Meyer and Kircher (2010) and sequenced at the Max Planck Institute for evolutionary Anthropology in Leipzig on the Illumina Hi-Seq 2000 as a 96-bp paired-end run. Base calling was conducted with freeIbis (Renaud et al., 2013), adaptor and primer sequences were removed, reads with low complexity as well as false paired indices were discarded. All three libraries were trimmed by applying a filter of 15, i.e., reads with more than five bases below a phred quality score of 15 were removed. For *Owenia fusiformis* and *Phyllochaetopterus* sp. library construction and sequencing as a 100-bp paired-end run on the Illumina HiSeq as well as quality filtering and adapter trimming with the Chastity filter, was performed by Genterprise Genomics in Mainz (Germany). The quality of all sequences was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and de novo assembly was conducted with CLC Genomics Workbench 7.5 (CLCbio, Arhus, Denmark) with the following settings: mismatch cost 3; insertion cost 3; deletion cost 3; length fraction 0.5; similarity fraction 0.8; minimum contig length 200; automatic word size; automatic bubble size; and contig adjustment by mapped reads. Assemblies were screened for possible (cross) contamination by investigating 18S rRNA gene sequences using local Blast. More information on raw data, including number of reads and contigs are given in Supplementary Table S1.

2.3. Mitochondrial genome annotation and comparison

We annotated the five newly sequenced mitochondrial genomes using MITOS under the mitochondrial code for invertebrate mitochondria (Bernt et al., 2013b) and subsequented

curated the annotation manually. To detect and analyse the secondary structure and duplication events of all tRNAs, the program ARWEN (Laslett and Canback, 2008) was applied. For pairwise comparison of the mitochondrial gene order of all basal branching annelids to the most likely ground pattern of Pleistoannelida and Lophotrochozoa, we used the program CREx (Bernt et al., 2007), which reconstructs events for reversals, transpositions, reverse transpositions and tandem duplication random loss. The analysis was performed by applying the common intervals parameter for distance measurement and only taxa with thecomplete mitochondrial gene order of protein-coding and ribosomal RNA genes were included (the more variable tRNAs were excluded). For Annelida we included the basal branching annelids, the most likely ground pattern of Pleistoannelida, which is realized in many representatives of Errantia and Sedentaria, as well as orders of Pleistoannelida which differ from that pattern (i.e., Echiura, Eunicidae, Ampharetidae). For Lophotrochozoa we included the most likely ground pattern according to Bernt et al. (2013a) and members of several phyla, as representatives of these phyla differing from that pattern. Moreover, the most likely genome rearrangement scenarios between the gene order of each basal branching annelid and the gene orders of either its sister group, Pleistoannelida or Lophotrochozoa were determined.

Table 1: Source of mitochondrial genomes used for phylogenetic analyses. Asterisks indicate incomplete mitochondrial data, bold taxa represent new mitochondrial data generated in this study.

Phylum	Family	Species	Accession
Annelida	Alvinellidae	*Paralvinella sulfincola**	FJ976042
	Ampharetidae	*Eclysippe vanelli*	EU239687
		Auchenoplax crinita	FJ976041
	Amphinomidae	***Eurythoe complanata***	**KT726962**
	Chaetopteridae	***Chaetopterus variopedatus***	**KT726958**
		Phyllochaetopterus sp.	**KT726961**
	Diurodrilidae	*Diurodrilus subterraneus**	KC790350
	Eunicidae	*Marphysa sanguinea*	KF733802
	Glossiphoniidae	*Helobdella robusta**	AF178680
	Hirudinidae	*Hirudo nipponia*	KC667144
	Lumbricidae	*Lumbricus terrestris*	LTU24570
	Magelonidae	***Magelona mirabilis***	**KT726959**
	Maldanidae	*Clymenella torquata*	AY741661
	Megascolecidae	*Perionyx excavates*	EF494507
	Myzostomida	*Myzostoma seymourcollegiorum**	EF506562
		*Endomyzostoma sp.**	FJ975144
	Nephtyidae	*Nephtys sp.*	EU293739
	Nereididae	*Platynereis dumerilii*	AF178678
		Tylorrhynchus heterochaetus	KM111507.1
		Perinereis aibuhitensis	KF611806
		Perinereis nuntia	JX644015
	Orbiniidae	*Orbinia latreillii*	AY961084
		*Questa ersei**	FJ612452
		*Scoloplos cf. armiger**	DQ517436
	Oweniidae	***Owenia fusiformis***	**KT726960**
	Pectinariidae	*Pectinaria gouldii**	FJ976040
	Phascolosomatidae	*Phascolosoma esculenta*	EF583817
	Siboglinidae	*Galathealinum brachiosum**	AF178679
		*Riftia pachyptila**	AY741662
	Sipunculidae	*Phascolopsis gouldii*	AF374337
		Sipunculus nudus	FJ422961
	Terebellidae	*Pista cristata*	EU239688
	Trichobranchidae	*Terebellides stroemii*	EU236701
	Urechidae	*Urechis caupo*	AY619711
		Urechis unicinctus	EF656365
Mollusca	Haliotoidae	*Haliotis tuberculata*	FJ599667
	Mophaliidae	*Katharina tunicata*	NC_001636

Table 1 (continued)

Phylum	Family	Species	Accession
	Sepiidae	*Sepia officinalis*	NC_007895
	Solemyidae	*Solemya velum*	NC_017612
Brachiopoda	Cancellothyrididae	*Terebratulina retusa*	NC_000941
	Laqueidae	*Laqueus rubellus*	NC_002322
Bryozoa	Bugulidae	*Bugula neritina*	NC_010197
	Tubuliporidae	*Tubulipora flabellaris*	NC_015646
	Watersiporidae	*Watersipora subtorquata*	NC_011820
Phoronida	Phoronidae	*Phoronis psammophila*	AY368231
Nemertea	Cephalothricidae	*Cephalothrix simula*	FJ594739
	Emplectonematidae	*Emplectonema gracile*	NC_016952
	Lineidae	*Lineus viridis*	FJ839919
	Nectonemertidae	*Nectonemertes cf. mirabilis*	NC_017874

2.4. Phylogenetic analyses

For phylogenetic analyses, we generated a data set including all annelid taxa of which all 13 protein-coding genes were available (data set 1) and a data set which also included annelid taxa with partial mitochondrial genomes (data set 2). Data set 1 comprises 40 taxa, including 4 nemerteans, 1 phoronid, 3 bryozoans, 2 brachiopods, 4 molluscs and 26 annelids, data set 2 comprises 48 taxa, including the same out group taxa and 34 annelids (Table 1). For Clitellates, we only included 3 representatives out of the 12 available genomes in all analyses. Sequences for all 13 mitochondrial protein-coding genes were first translated into amino acid sequences from the nucleotide sequences using the mitochondrial code for invertebrates and then independently aligned using MAFFT version 7 (Katoh et al., 2002). For each gene alignment columns containing highly diverse amino acids and many gaps were masked with REAP (Hartmann and Vision, 2008) and single alignments were concatenated into one data set using FASconCAT version 1.0 (Kück and Meusemann, 2010). Data set 1 covers 3654 amino acid positions and data set 2 3630.

For both datasets, we employed IQ-TREE version 1.3.4 (Nguyen et al., 2015) to determine the best fitting partitioning schemes as well as amino acid substitution models for each of the partitions (Supplementary Table S8). We then performed 10 independent maximum likelihood estimations of both partitioned datasets with RAxML version 8.1.3 (Stamatakis, 2014). Bootstrap support was estimated from 1000 pseudoreplicates. Maximum likelihood phylogenetic estimation was further conducted with IQ-TREE, which, with a different stochastic algorithm, produced topologies congruent to those inferred with RAxML.

In addition, we performed Bayesian phylogenetic analysis with PhyloBayes version 3.3. (Lartillot et al., 2009). First, we tested if for our data sets the siteheterogeneous CAT model implemented in PhyloBayes has a better statistical fit then the single-matrix models used in maximum likelihood analyses. To this end, we performed a comparison of the best-fitting single matrix model implemented in PhyloBayes (mtART) and the CAT models (CAT-GTR and CAT-Poisson) by crossvalidation. From both of our datasets, we created 10 learning sets and sampled for 1,000 generations for all models under a fixed topology (best maximum likelihood tree). For both datasets, CAT-GTR was supported as model that best fits the data and therefore used in subsequent analyses (Supplementary Table S9). MCMC sampling with PhyloBayes was performed by running two independent chains each for >16,000 cycles, discarding the first 6,000 as burnin. All summary variables of all runs were plotted to check for stationarity and convergence. Furthermore, the 'tracecomp' function implemented in PhyloBayes was used to ensure convergence of runs (maximal discrepancy of all variables: 0.3, minimal effective sampling size: 50). Convergence of bipartition frequencies was ensured by using the 'bpcomp'

function (maxdif < 0.1). Finally, a consensus tree was constructed from all trees of the posterior sample. Posterior probabilities were inferred from clade frequencies of postburnin-trees.

3. Results

3.1. Genomic features

Annotations, length and strand position of all genes and RNAs are given in Supplementary Tables S4-S8 and circular genomes are illustrated in Figures 1-2 and Supplementary Figures S1-S3. Mitochondrial genome size varies from 15239 bp (*Magelona mirabilis*) to 16204 bp (*Owenia fusiformis*). For each of the five mitochondrial genomes sequenced in this study, all 13 protein-coding genes, two rRNAs and 22 tRNAs could be detected as typical in most other metazoans. In all five genomes, two tRNAs encoding for serine (tRNA-S1 and -S2) and leucine (tRNA-L1 and -L2) were found. As usual for all other annelids investigated to date, all genes and RNAs are organized on a single strand, the „+" strand, with the only exception of rRNA-T and tRNA-P in *Owenia fusiformis* and *Magelona mirabilis*, which encode both tRNAs on the „-" strand. The mitochondrial genomes of *Chaetopterus variopedatus* and *Eurythoe complanata* both contain 2 copies of the tRNA encoding for methionine, which was verified by visual inspection of the secondary structure. All mitochondrial genomes are deposited in Genbank and accession numbers can be found in Table 1.

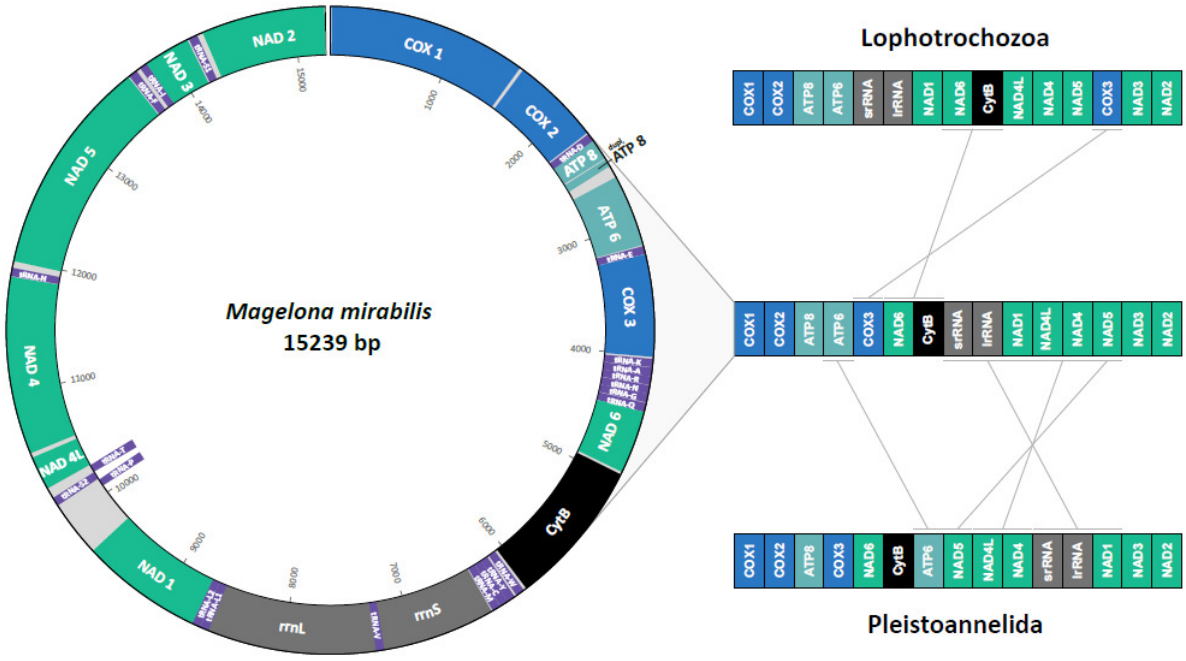


Fig. 1: Gene order of the mitochondrial genome of *Magelona mirabilis* and comparison of mitochondrial genes to the putative ground pattern of Pleistoannelida and Lophotrochozoa. All genes are transcribed on the „+" strand except for tRNA-T and tRNA-P. Abbreviations: NAD1-6, 4L – NADH dehydrogenase subunits 1-6 and 4L; COX1-3 – cytochrome oxidase subunits 1-3; CYTB – cytochrome b; ATP6/8 – ATP synthase subunit 6/8; lrRNA – large ribosomal RNA; srRNA – small ribosomal RNA; single letters – tRNAs encoding for amino acids: alanine (A), cysteine (C), aspartic acid (D), glutamatic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L1 and L2), methionine (M), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S1 and S2), threonine (T), valine (V), tryptophan (W), tyrosine (Y).

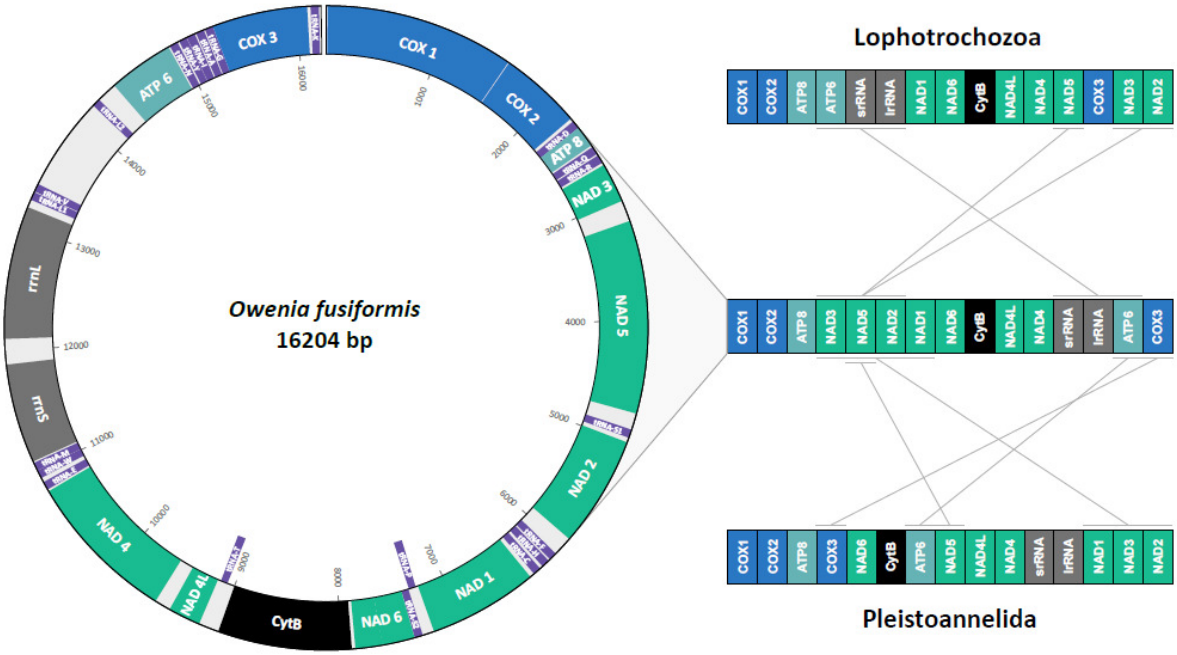


Fig. 2: Gene order of the mitochondrial genome of *Owenia fusiformis* and comparison of mitochondrial genes to the putative ground pattern of Pleistoannelida and Lophotrochozoa. All genes are transcribed on the „+" strand except for tRNA-T and tRNA-P. Abbreviations: NAD1-6, 4L – NADH dehydrogenase subunits 1-6 and 4L; COX1-3 – cytochrome oxidase subunits 1-3; CYTB – cytochrome b; ATP6/8 – ATP synthase subunit 6/8; lrRNA – large ribosomal RNA; srRNA – small ribosomal RNA; single letters – tRNAs encoding for amino acids: alanine (A), cysteine (C), aspartic acid (D), glutamatic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L1 and L2), methionine (M), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S1 and S2), threonine (T), valine (V), tryptophan (W), tyrosine (Y).

3.2. Unassigned, non-coding regions and duplication events

Characteristics of each of the five mitochondrial genomes investigated in this study are the larger intergenic regions which can be found besides the numerous smaller non-coding regions below 100 bp. In addition, copies of the tRNA encoding for methionine can be observed in two species, as well as full or partial gene duplications of COX3, NAD3 and ATP8.

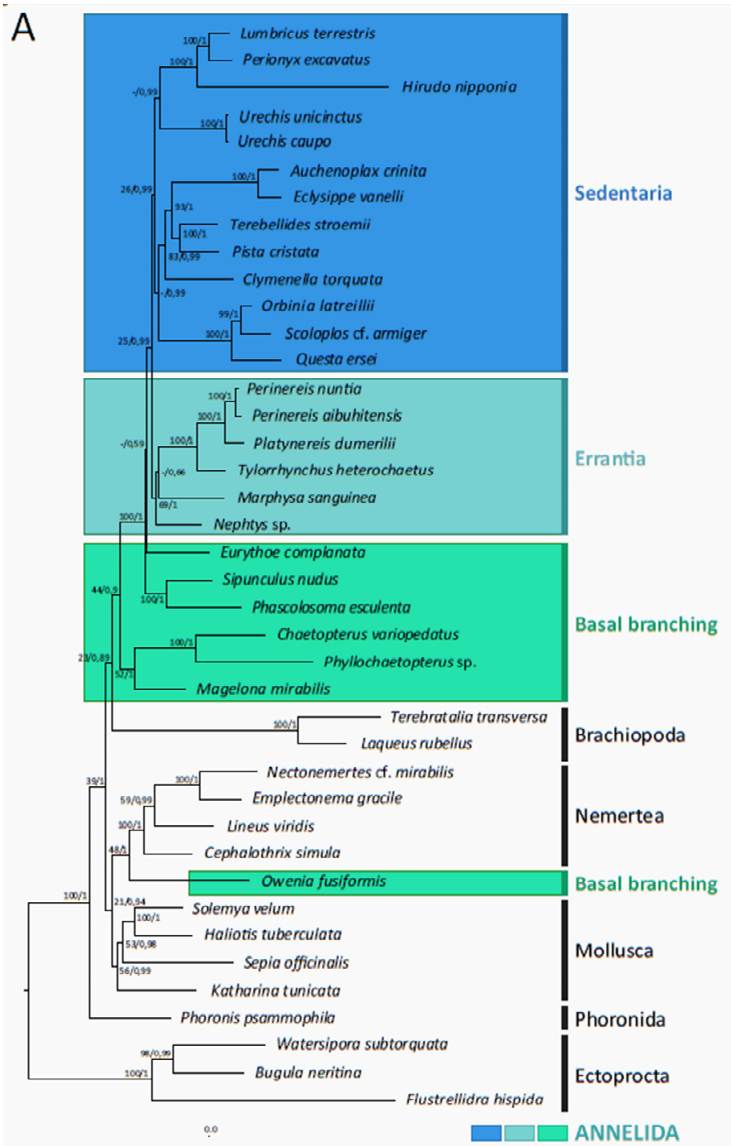
In the mitochondrial genome of *Eurythoe complanata* one large non-coding region of 1121 bp between NAD4 and COX1 could be assigned, as well as two copies of the tRNA encoding for methionine that are directly adjacent to each other (Supplementary Figure S1). We called the 5‟ upstream copy of the "+" strand tRNAM1 and the 3′ downstream copy tRNA-M2. tRNA-M1 has a shorter TΨC stem with only 3 bases, a shorter TΨC loop with 3 bases and a larger DHU loop with 8 bases, rather than tRNA-M2 with 4 matching bases in the TΨC stem, 4 bases in the TΨC loop and 7 bases in the DHU loop (Supplementary Figure S4).

For *Chaetopterus variopedatus* two large unassigned regions of 639 bp and 314 bp were found between tRNA-P and CytB (containing a 110 bp long fragment of a presumed duplication of NAD3) and within COX1, respectively (Supplementary Figure S2). Additionally, two copies of the tRNA encoding for methionine as found in the mitochondrial genome of *Eurythoe complanata* could be assigned. The upstream copy of the "+" strand was named tRNA-M1 and the 3′ downstream copy tRNA-M2. In both copies, the number of matching bases in the TΨC and DHU stem is the same. However, the TΨC loop in tRNA-M2 is shorter than in tRNA-M1 with only 3 bases instead of 5 and the DHU loop of tRNA-M1 is shorter than in tRNA-M2 with only 6 bases instead of 8 (Supplementary Figure S4).

The annotated mitochondrial genome of *Phyllochaetopterus* sp. contains one region of 208 bp between tRNA-G and ATP6; one region of 350 bp between tRNA-F and COX3 containing a 74 bp long fragment of a COX3 duplication (Supplementary Figure S3).
For *Magelona mirabilis* two larger intergenic regions can be observed. One of 163 bp between the genes ATP8 and ATP6, containing a 30 bp fragment of a ATP8 duplication, and one of 322 bp between NAD1 and tRNA-P are found (Figure 1).
Of all mitochondrial genomes analysed in this study the mitochondrial genome of *Owenia fusiformis* contains the most non-coding regions larger than 100 bp. In total, seven unassigned regions could be found: a 173 bp region between NAD3 and NAD5, a 135 bp region between NAD5 and tRNA-S1, a 161 bp region between NAD2 and tRNA-F, a 132 bp region between NAD4L and NAD4, a 209 bp region between the small and large ribosomal RNA, a 167 bp region between tRNA-L2 and ATP6 and the largest non-coding region of 773 bp between tRNA-V and tRNA-L2 (Figure 2).

3.3. Phylogenetic analyses

In all analyses of the two data sets monophyletic Annelida could not be recovered since *Owenia fusiformis* groups with Nemertea (Figure 3, Supplementary Figure S5 and S6). Additionally, in the ML analysis including complete and partial annelid mitochondrial genomes (data set 2), *Magelona mirabilis* groups together with Oweniidae as sister group to Nemertea and Chaetopteridae are sister group to Brachiopoda (Supplementary Figure S6).



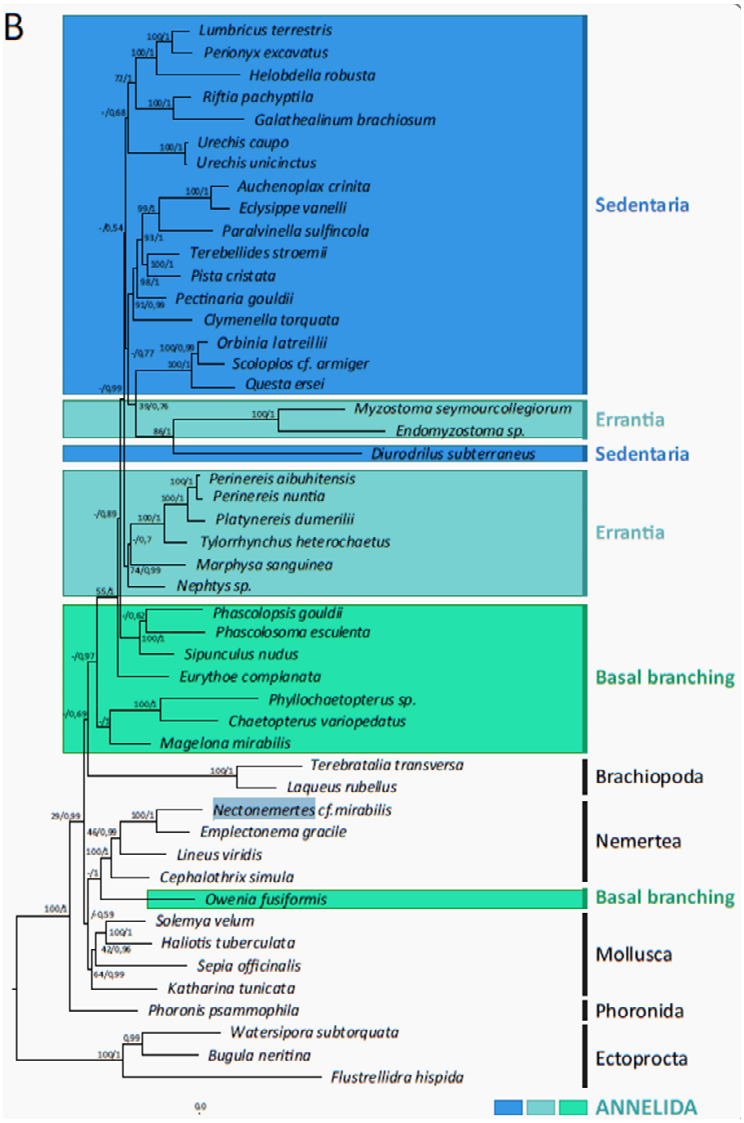


Fig. 3: Phylogenetic relationships of Annelida based on mitochondrial genome data. **A)** Consensus tree of the Bayesian analysis using the CAT-GTR model of data set 1 comprising only complete mitochondrial genomes (40 taxa, 3654 amino acid positions). **B)** Consensus tree of the Bayesian analysis using the CAT-GTR model of data set 2 comprising complete and partial annelid mitochondrial genomes (48 taxa, 3630 amino acid positions). Bootstrap support values (BS) where the topology of the Maximum Likelihood and Bayesian analyses agree are depicted before the posterior probabilities (PP) or indicated with a hyphen if not so.

Despite the position of Oweniidae, in the ML and BI analyses including only complete mitochondrial genomes (data set 1) the phylogenetic relationships within Annelida are in agreement with previous molecular analyses based on transcriptomes (Andrade et al., 2015; Struck et al., 2015; Weigert et al., 2014) by recovering monophyletic Errantia (BS=69, PP=1), Sedentaria (BS=26, PP=0.99) and Pleistoannelida (BS=25, PP=0.99), as well as the basal branching annelids (Figure 3A, Supplementary Figure S5). However, for the deeper nodes the support values are very low, but the topologies of both analyses are very similar, except for the position of Amphinomida and sister group relationships within Sedentaria.

In the ML and BI analyses including additional partial mitochondrial genomes (data set 2) annelid relationships as described above could not be resolved, except for monophyletic Errantia (without Myzostomida; BS=74, PP=1) and monophyletic Pleistoannelida (only BI analysis, PP=0.89). The trees differ significantly in the positions of Errantia, Sipuncula,

Amphinomidae, Myzostomida, Diurodrilidae, Chaetopteridae, and Magelonidae and the support for deeper nodes is also very low (Figure 3B, Supplementary Figure S6). In both analyses *Diurodrilus* sp. and the two myzostomids group within Annelida as sister group to Orbiniidae (Figure 3B, Supplementary Figure S6). Amphinomida and Sipuncula branch off early only in the BI analysis, whereas in the ML analysis they group with Errantia rendering Pleistoannelida paraphyletic (Figure 3B, Supplementary Figure S6).

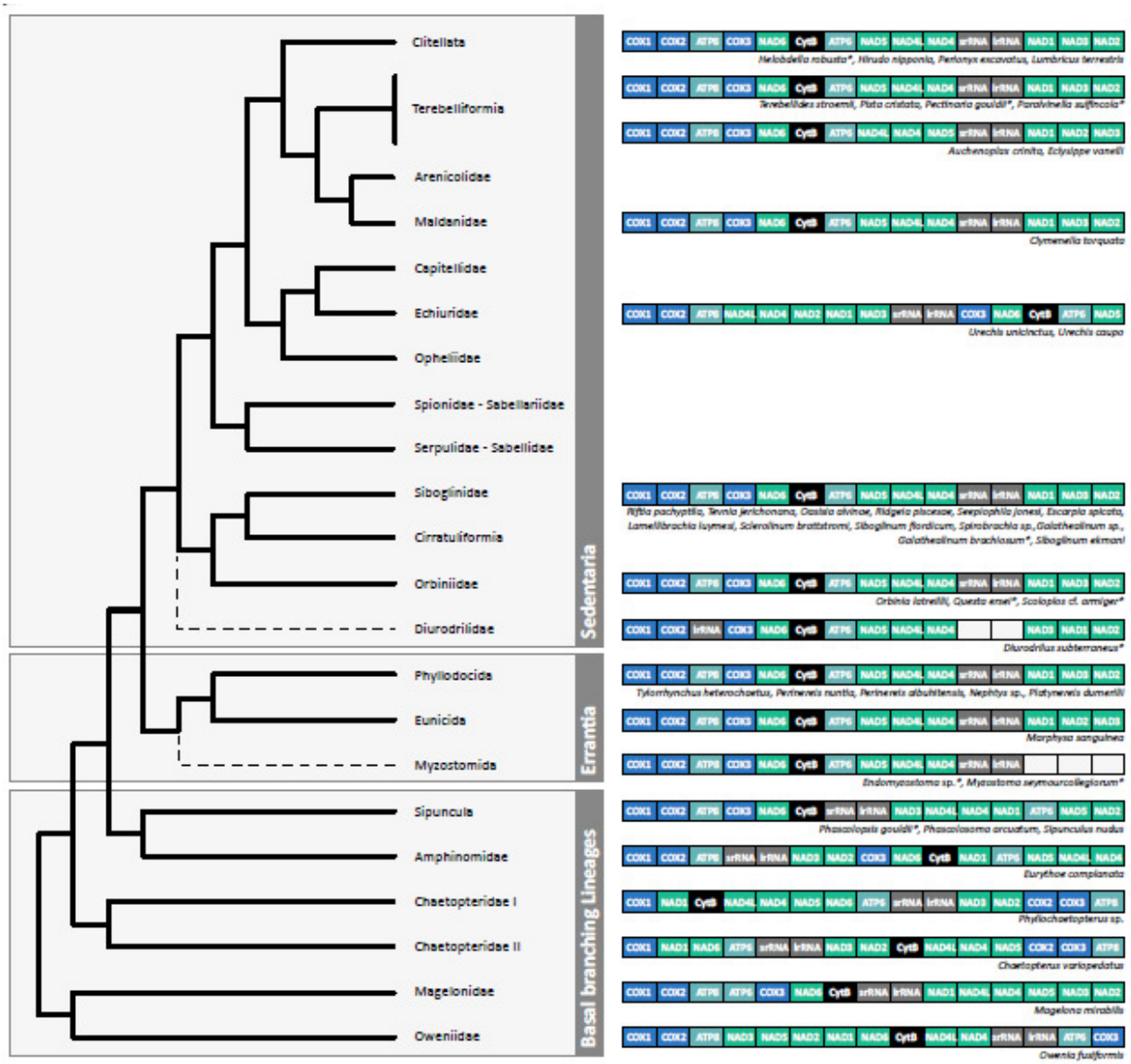


Fig. 4: Relationships within Annelida and different mitochondrial gene order of each taxon. Annelid phylogeny is depicted based on Weigert et al. (2014), sister group relationships of families not represented in Weigert et al. (2014) are obtained from Struck et al. (2007) (Maldanidae and Ampharetidae) and Golombek et al. (2013) (Diurodrilidae). Dashed lines indicate an uncertain phylogenetic position. Only protein-coding genes of available mitochondrial genomes and ones, which were generated in this study, are included. Taxa with partial mitochondrial genomes are marked with an asterisk and missing genes within the mitochondrial gene order are indicated with grey boxes. Missing genes are: ATP8 and srRNA (Diurodilidae); NAD1, NAD2 and NAD3 (Myzostomida). Genes are not scaled to real length and are indicated by standard abbreviations.

3.4. Mitochondrial gene order and rearrangements

The gene order of each of the five mitochondrial genomes of the basal branching annelids differ significantly from each other and show a high variability in contrast to gene orders from annelids belonging to either Errantia or Sedentaria (Figure 4). The latter groups generally share

a conserved mitochondrial ground pattern (Figure 4 and 5). In each scenario reconstructed with CREx several transpositions, reverse transpositions, reversals or tandem-duplication-random-loss (tdrl) events are necessary to rearrange genes of the basal branching annelids in comparison to the ground pattern of Lophotrochozoa or Pleistoannelida (Figure 5, Supplementary Table S9). For all rearrangement scenarios, tRNAs were not compared due to their higher variability in location. Nevertheless, the location of tRNA-P and tRNA-T in *Owenia fusiformis* and *Magelona mirabilis* has to be highlighted, since they are the only known genes/RNAs within Annelida which are encoded on the „-" strand. In general, the highest number of similarities in mitochondrial genome organization within annelids can be found between the putative ground pattern of Pleistoannelida and members of Pleistoannelida which differ in that pattern (Eunicidae and Ampharetidae), except for Echiura (Figure 5).

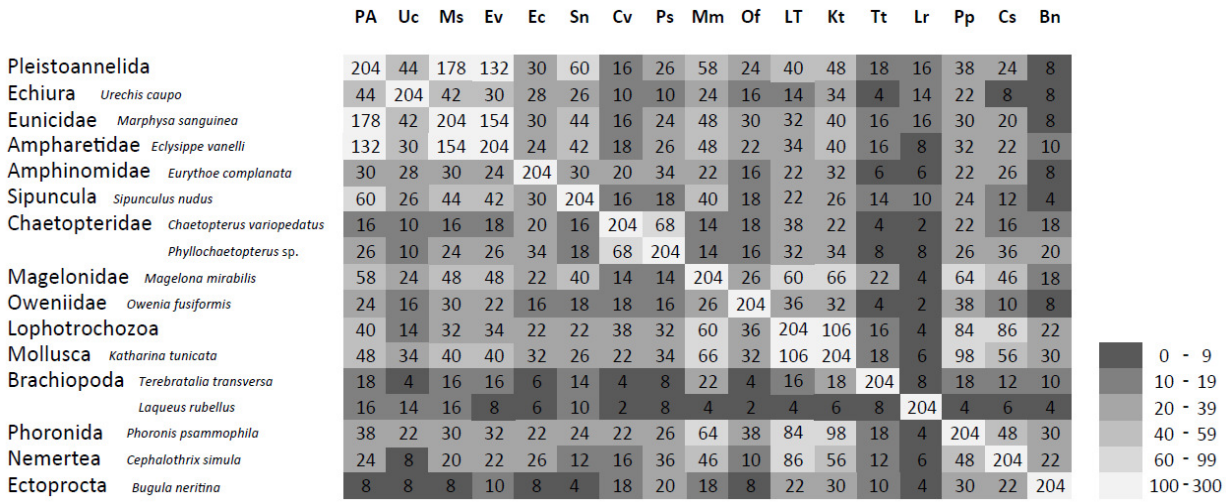


Fig. 5: Results of the pairwise comparisons of mitochondrial gene orders of basal branching annelids with the putative ground pattern of Pleistoannelida and Lophotrochozoa as well as with differing annelid and lophotrochozoan members. Scores of the CREx analysis of each pairwise comparison indicate the similarities of the compared mitochondrial gene orders, where 204 is the highest score and represents identical gene order. Only taxa with the complete mitochondrial gene order of protein-coding genes were included in the analysis. The gene order of the putative ground pattern of Pleistoannelida is identical with the one found in Clitellata (*Helobdella robusta, Hirudo nipponia, Perionyx excavates* and *Lumbricus terrestris)*, Terebelliformia (*Terebellides stroemii, Pista cristata, Pectinaria gouldii* and *Paralvinella sulfincola,* except Ampharetidae*),* Maldanidae *(Clymenella torquata),* Siboglinidae *(Riftia pachyptila),* Orbiniidae *(Orbinia latreillii, Questa ersei* and *Scoloplos* cf. *armiger),* Phyllodocida *(Tylorrhynchus heterochaetus, Perinereis nuntia, Perinereis aibuhitensis, Nephtys* sp. and *Platynereis dumerilii)* and Myzostomida *(Endomyzostoma* sp. and *Myzostoma seymourcollegiorum).* Abbreviated taxa: Bn – Bugula neritina, Cs - *Cephalothrix simula*, Cv - *Chaetopterus variopedatus*, Ev - *Eclysippe vanelli*, Ec - *Eurythoe complanata*, Kt - *Katharina tunicata*, Lr – *Laqueus rubellus*, LT - Lophotrochozoa, Ms - *Marphysa sanguinea*, Mm - *Magelona mirabilis*, Of - *Owenia fusiformis*, PA - Pleistoannelida, Pp - *Phoronis psammophila*, Ps - *Phyllochaetopterus* sp., Sn - *Sipunculus nudus*, Tt - *Terebratalia transversa*, Uc - *Urechis caupo*.

The gene order of *Eurythoe complanata* is more similar to the one found in Pleistoannelida and closely related families (Sipuncula and Chaetopteridae) than to Lophotrochozoan taxa and the most basal branching annelids *Owenia fusiformis* and *Magelona mirabilis* (except for the pattern found in *Katharina tunicata* and *Chaetopterus variopedatus*). Most similarities can be found in the pattern of Sipuncula, *Phyllochaetopterus* sp., Pleistoannelida (ground pattern and *Marphysa sanguinea*) and *Katharina tunicata* (Figure 5). The genome differences to its sister group Sipuncula can be reconstructed with three transposition events, to Pleistoannelida with one tdrl and one transposition event, and to *Phyllochaetopterus* sp. with two transposition, one reverse transposition, one reversal and one tdrl event (Figure 5, Supplementary Table S9).
The fewest events in gene order rearrangement can be observed between *Chaetopterus variopedatus* and *Phyllochaetopterus* sp. with the only difference being a transposition of the

CytB-NAD4L-NAD4-NAD5 cluster (Supplementary Table S9). The gene order of *Chaetopterus variopedatus* is more similar to the hypothetical ground pattern of Lophotrochozoa and other lophotrochozoans than to any other annelid pattern (except for *Phyllochaetopterus* sp.) differing in only three transposition events (Figure 5, Supplementary Table S9). For the gene order of *Phyllochaetopterus* sp. the similarity to other Lophotrochozoa is even higher and gene rearrangements can be reconstructed with four transposition events (Supplementary Table S9). From all annelids the two chaetopterids share the most similarities with the amphinomid *Eurythoe complanata*, which is also part of the basal radiation.
Of all basal branching taxa, the mitochondrial pattern of *Magelona mirabilis* and Sipuncula show the most similarities in mitochondrial gene order with the putative ground pattern of Pleistoannelida (Figure 5). Interestingly, *Magelona mirabilis* shares as many similarities with gene arrangements found in members of Mollusca, Phoronida and the ground pattern of Lophotrochozoa, differing in only two rearrangement events (Supplementary Table S9, Figure 5). The differences of the mitochondrial gene order to the one of its sister group *Owenia fusiformis* are much higher than to most other Pleistoannelida (Figure 5, Supplementary Table S9).
The mitochondrial gene order of *Owenia fusiformis* shares the most similarities with the one found in the ground pattern of Lophotrochozoa and *Phoronis psammophila* (Figure 5). The organization of the mitochondrial genome of *Owenia fusiformis* differs in one transposition, one reverse transposition, one reversal and two tdrl events to Pleistoannelida, and four reversals and two transposition events to Lophotrochozoa (Supplementary Table S9).

4. Discussion

4.1. Genome organization and structural features

Mitochondrial genes in Annelida are generally transcribed only from one strand. However, *Owenia fusiformis* and *Magelona mirabilis,* are the first known annelids where not all of the 37 genes are transcribed from one single strand, since the two tRNAs encoding for proline (tRNA-P) and threonine (tRNA-T) are located on the opposite strand (Figure 1 and 2). Boore (1999) suggested a “ratchet effect” for the scenario in which by chance all genes were placed on one strand and that transcription therefore would sooner or later be lost for one of the two strands, hindering further inversion without additional transcription elements. This scenario was proposed for the last common ancestor (LCA) of annelids and might be the most parsimonious hypotheses, if basal branching lineages would also show the same transcription direction for all genes (Boore, 1999; Valles and Boore, 2006). However, *Owenia fusiformis* and *Magelona mirabilis* are the exception to the typical annelid pattern. Both share the same tRNAs, which were placed on the opposite strand, presumably in the LCA of both families (which together form the sister taxon of all other annelids based on Weigert et al. (2014)). There are two hypotheses to interpret this pattern: 1) The LCA of Annelida had all genes on one strand and already lost the transcription signal on the other strand. With the inversion event of the two tRNAs in *Owenia fusiformis* and *Magelona mirabilis* necessary elements for transcription were also transposed. 2) The LCA of Annelida still had transcription signals on both strands and it was lost on one strand in the lineage leading to the rest of the Annelida, which forms the sister taxon of the clade comprising Oweniidae and Magelonidae. In this case, the LCA of Annelida likely possessed the same mitochondrial pattern as observed in Oweniidae and Magelonidae (all genes except for tRNA-P and tRNA-T are located on one strand), and both tRNAs were inverted on the opposite strand after the split of Oweniidae/Magelonidae from the rest of the annelid lineages. Nevertheless, if the strand usage and inversion of both tRNAs is a plesiomorphic condition in annelids rather than a synapomorphy for Oweniidae and Magelonidae, there should be traces in other lophotrochozoan groups. The mitochondrial genes

in brachiopods are generally encoded on one strand (Helfenbein et al., 2001) and in molluscs there is no general pattern. Different families of molluscs show a high diversity in strand usage and gene order (Osca et al., 2014). Interestingly, all but one mitochondrial genome of Nemertea, which are either sister group to Annelida or very closely related to them (Dunn et al., 2008; Laumer et al., 2015; Weigert et al., 2014), show a similar pattern to that found in *Owenia fusiformis* and *Magelona mirabilis,* where all genes are transcribed from one strand except for the two tRNAs encoding for threonine and proline (e.g. (Chen et al., 2012; Chen et al., 2011)). This shared strand usage and identical pattern in the two tRNAs between Nemerteans and the basal branching annelid families Oweniidae and Magelonidae favours hypothesis 2 suggesting that this pattern is the ancestral condition for annelids.

Another feature which could not be observed in annelids to date is the conjecture of the ATP8 and ATP6 gene, usually a common order in nearly all animal mitochondrial genomes with exception in members of certain lophotrochozoan phyla. In some species of Mollusca, Brachiopoda, Nemertea and Phoronida this gene boundary is disrupted (Boore, 2006; Chen et al., 2012; Helfenbein and Boore, 2004; Helfenbein et al., 2001; Noguchi et al., 2000). In Platyhelminthes, Acoelomorpha and Acanthocephala, the ATP8 gene is missing (Mwinyi et al., 2010; Steinauer et al., 2005; Valles and Boore, 2006). In Chaetognatha both ATP8 and ATP6 are missing (Papillon et al., 2004). *Magelona mirabilis* is the only annelid described so far which retained ATP8 adjacent to ATP6. This increases the likelihood of the ATP8-ATP6 conjecture as part of the annelid mitochondrial ground pattern, where the loss of this gene boundary might have occurred independently in Oweniidae and the LCA of Pleistoannelida + Chaetopteridae + Sipuncula + Amphinomidae, as observed among various other phyla.

4.2. Phylogenetic relationships based on mitochondrial sequence data

Reconstructing robust annelid relationships with morphological or few molecular markers failed in the past, recovering trees with paraphyletic Annelida or general low support values and lack of resolution (for review see (Struck, 2012). With the advent of new sequencing techniques including sequencing cost reduction, amplification of the relatively small mitochondrial genome became easily feasible and mitochondrial data an increasingly useful tool for investigating phylogenies. Nevertheless, robust relationships of higher ranked annelid groups and deeper splits could not be resolved by incorporating those data, whereas their application in affiliating uncertain taxa to Annelida (e.g. Diurodrilidae, Echiura, Siboglinidae, and Sipuncula) provided additional support and more stable results (Boore and Brown, 2000; Boore and Staton, 2002; Golombek et al., 2013; Jennings and Halanych, 2005; Mwinyi et al., 2009; Shen et al., 2009; Wu et al., 2009).

Our analyses yielded similarly unstable results. By analysing only complete mitochondrial genomes (data set 1), we found similar annelid relationships as proposed in recent molecular analyses (e.g. (Weigert et al., 2014), with Pleistoannelida comprising the reciprocal monophyletic Sedentaria and Errantia. The basal branching lineages group as well in the basal part of the tree, except of *Owenia fusiformis*, rendering Annelida not monophyletic (Figure 3A, Supplementary Figure S5). However, with the inclusion of additional partial genomes (data set 2), these relationships cannot be reconstructed (Figure 3B, Supplementary Figure S6). Pleistoannelida and Sedentaria form non-monophyletic groups, which can likely be explained due to the influence of long branch attraction (LBA) introduced by faster evolving taxa like myzostomids and diurodrilids. Similar problems with mitochondrial genomes of these two taxa have been already reported before (Golombek et al., 2013). Recent studies based on transcriptomes tend to group Myzostomida within Pleistoannelida, either as part or sister group to Errantia or within Sedentaria (Andrade et al., 2015; Weigert et al., 2014) and *Diurodrilus* within Sedentaria

(Andrade et al., 2015; Laumer et al., 2015; Struck et al., 2015), which is also supported in our analyses (Figure 3B, Supplementary Figure S6). However, our findings are in congruence with a proposed basal branching position of Sipuncula, Amphinomidae, Chaetopteridae, Oweniidae, and Magelonidae. As such it comes without surprise that these taxa are drawn to outgroup taxa in analyses based on only few molecular markers, e.g. mitochondrial genomes, as the presumed divergences date back into the Cambrian (Weigert et al., 2014). The lack of resolution of such datasets for deep metazoan phylogeny has been already demonstrated by Bernt et al. (2013a). Further investigations on phylogenetic relationships in major animal groups as old as annelids should be based on more suitable and a higher amount of molecular and morphological data.

4.3. Mitochondrial genome rearrangements in Annelida

Whereas the reconstruction of ancient relationships with mitochondrial sequences has its limits, comparison of mitochondrial gene order seems promising, since gene rearrangements seldom occur convergently and closely related species often share the same gene order (Boore, 1999). Furthermore, questionable assignments of taxa to certain groups can be investigated by comparing mitochondrial gene orders which thereby serve as an additional marker if morphology or molecular data is highly controversial or lacking, as previously demonstrated for Sipuncula (Mwinyi et al., 2009) and Diurodrilidae (Golombek et al., 2013). In contrast to closely related phyla like molluscs, brachiopods and nemerteans, Annelida were believed to possess a highly conserved mitochondrial gene order, given the available data (Jennings and Halanych, 2005; Valles and Boore, 2006). Our results demonstrate that this hypothesis is clearly restricted to members belonging to Pleistoannelida (Errantia and Sedentaria) and it is parsimonious to assume that this conserved order (even true for most tRNAs) represents the ground pattern for this clade. When comparing the mitochondrial gene order of Myzostomida with basal branching annelids or Pleistoannelida, the position of myzostomids is in congruence with recent molecular analyses as part of Pleistoannelida (Andrade et al., 2015; Weigert et al., 2014) instead of being part of the base of the annelid tree (Struck et al., 2011), since they exhibit the same conserved arrangement of genes as seen in members of Errantia and Sedentaria. In contrast, all represented taxa branching from the base of the annelid tree show a completely different arrangement of genes than observed in Pleistoannelida, but the reconstruction of a putative annelid mitochondrial ground pattern still remains difficult. However, as already mentioned, it seems very likely that all genes were encoded on one strand except for tRNA-P and tRNA-T, as found in the two annelids *Owenia fusiformis* and *Magelona mirabilis* and in the ground pattern of Nemertea (Chen et al., 2012; Chen et al., 2011). Additionally, there are a few conserved blocks of genes in annelids that, with a few exceptions, can also be found in the putative ground pattern of Lophotrochozoa and Bilateria (Bernt et al., 2013a; Lavrov and Lang, 2005) and might be represented in the annelid ground pattern: COX1-COX-2-ATP8, NAD6-CYTB, SrRNA-LrRNA and NAD4LNAD4. From all basal branching annelids, the gene order of *Magelona mirabilis* is most similar to the putative ground pattern of Lophotrochozoa (Bernt et al., 2013a), differing only in 2 transposition events and exhibiting a high number of identical gene blocks including tRNAs with conserved bilaterian gene blocks (Bernt et al., 2013a): NAD6-CytB-S2 (without tRNA-S2), SrRNA-V-LrRNA-L1-L2-NAD1, NAD4L-NAD4-HNAD5, NAD2-COX1-COX2-K-ATP8-ATP6-COX3 (tRNA-D instead of tRNA-K, tRNA21-E between ATP6-COX3). It is tempting to assume that the pattern of *Magelona mirabilis* is similar to the ancestral pattern for Annelida and this might be also close to the lophotrochozoan ground pattern. The fact that lophotrochozoan taxa have convergently lost and rearranged mitochondrial genes in numerous ways resulted in a blurry picture of the mitochondrial ground pattern of this group. However, our mitochondrial data on early branching annelids is an important contribution to understand evolutionary relationships within Annelida and perhaps of putative sister groups and demonstrates that annelids fall in line with other lophotrochozoan

animal groups regarding the high variability in mitochondrial gene rearrangements at least in their basal radiation.

Abbreviations:
NAD1-6, 4L – NADH dehydrogenase subunits 1-6 and 4L; COX1-3 – cytochrome oxidase subunits 1-3; CYTB – cytochrome b; ATP6/8 – ATP synthase subunit 6/8; lrRNA – large ribosomal RNA; srRNA – small ribosomal RNA; mt – mitochondrial; single letters – tRNAs encoding for amino acids: alanine (A), cysteine (C), aspartic acid (D), glutamatic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L1 and L2), methionine (M), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S1 and S2), threonine (T), valine (V), tryptophan (W), tyrosine (Y)

Acknowledgements
We are thankful to Conrad Helm (University of Leipzig, Germany) and the staff of the Station Biologique de Roscoff (France) for supplying the specimens and for providing laboratory space. We additionally thank Martin Schlegel (University of Leipzig, Germany) for providing materials and facilities and Sarah Tobergte (University of Osnabrück, Germany) for assistance in the laboratory. This study was supported by the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, by the DFG grant BL787/5-1 to CB, DFG-STR-683/6-1, 6-2, 8-1 & 8-2 to THS and the EU due to ASSEMBLE grant agreement no. 227799 to CB (http://www.assemblemarine.org).

References
Aguado, M.T., Glasby, C.J., Schroeder, P.C., Weigert, A., Bleidorn, C., 2015. The making of a branching annelid: an analysis of complete mitochondrial genome and ribosomal data of Ramisyllis multicaudata. Scientific reports 5, 12072.

Andrade, S.C.S., Novo, M., Kawauchi, G.Y., Worsaae, K., Pleijel, F., Giribet, G., Rouse, G.W., 2015. Articulating "archiannelids": Phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. Mol Biol Evol.

Bernt, M., Bleidorn, C., Braband, A., Dambach, J., Donath, A., Fritzsch, G., Golombek, A., Hadrys, H., Juhling, F., Meusemann, K., Middendorf, M., Misof, B., Perseke, M., Podsiadlowski, L., von Reumont, B., Schierwater, B., Schlegel, M., Schrodl, M., Simon, S., Stadler, P.F., Stoger, I., Struck, T.H., 2013a. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. Mol Phylogenet Evol 69, 352-364.

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., Putz, J., Middendorf, M., Stadler, P.F., 2013b. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69, 313-319.

Bernt, M., Merkle, D., Ramsch, K., Fritzsch, G., Perseke, M., Bernhard, D., Schlegel, M., Stadler, P.F., Middendorf, M., 2007. CREx: inferring genomic rearrangements based on common intervals. Bioinformatics 23, 2957-2958.

Bleidorn, C., Podsiadlowski, L., Bartolomaeus, T., 2006. The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (Annelida, Orbiniidae) - A novel gene order for Annelida and implications for annelid phylogeny. Gene 370, 96-103.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res 27, 1767-1780.

Boore, J.L., 2004. Complete mitochondrial genome sequence of *Urechis caupo*, a representative of the phylum Echiura. BMC Genomics 5: 67.

Boore, J.L., 2006. The complete sequence of the mitochondrial genome of *Nautilus macromphalus* (Mollusca : Cephalopoda). BMC Genomics 7: 182.

Boore, J.L., Brown, W.M., 1994. Mitochondrial genomes and the phylogeny of mollusks. Nautilus 108, 61-78.

Boore, J.L., Brown, W.M., 2000. Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: Sequence and gene arrangement comparisons indicate the Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol Biol Evol 17, 988-988.

Boore, J.L., Staton, J.L., 2002. The mitochondrial genome of the sipunculid *Phascolopsis gouldii* supports its association with Annelida rather than Mollusca. Mol Biol Evol 19, 127-137.

Bridge, D., Cunningham, C.W., Schierwater, B., Desalle, R., Buss, L.W., 1992. Class-level relationships in the phylum Cnidaria - Evidence from mitochondrial genome structure. P Natl Acad Sci USA 89, 8750-8753.

Chen, H.X., Sun, S.C., Sundberg, P., Ren, W.C., Norenburg, J.L., 2012. A comparative study of nemertean complete mitochondrial genomes, including two new ones for *Nectonemertes* cf. *mirabilis* and *Zygeupolia rubens*, may elucidate the fundamental pattern for the phylum Nemertea. BMC Genomics 13: 139.

Chen, H.X., Sundberg, P., Wu, H.Y., Sun, S.C., 2011. The mitochondrial genomes of two nemerteans, *Cephalothrix* sp. (Nemertea: Palaeonemertea) and *Paranemertes* cf. *peregrina* (Nemertea: Hoplonemertea). Mol Biol Rep 38, 4509-4525.

Clary, D.O., Wolstenholme, D.R., 1984. The *Drosophila* mitochondrial genome. Oxford surveys on eukaryotic genes 1, 1-35.

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

Golombek, A., Tobergte, S., Nesnidal, M.P., Purschke, G., Struck, T.H., 2013. Mitochondrial genomes to the rescue - Diurodrilidae in the myzostomid trap. Mol Phylogenet Evol 68, 312-326.

Gustincich, S., Manfioletti, G., Delsal, G., Schneider, C., Carninci, P., 1991. A fast method for high-quality genomic DNA extraction from whole human blood. Biotechniques 11, 298-&.

Hartmann, S., Vision, T.J., 2008. Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment? BMC Evol Biol 8: 95.

Helfenbein, K.G., Boore, J.L., 2004. The mitochondrial genome of *Phoronis architecta* – Comparisons demonstrate that phoronids are Lophotrochozoan protostomes. Mol Biol Evol 21, 153-157.

Helfenbein, K.G., Brown, W.M., Boore, J.L., 2001. The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. Mol Biol Evol 18, 1734-1744.

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

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059-3066.

Kück, P., Meusemann, K., 2010. FASconCAT: Convenient handling of data matrices. Mol Phylogenet Evol 56, 1115-1118.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286-2288.

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

Laumer, C.E., Bekkouche, N., Kerbl, A., Goetz, F., Neves, R.C., Sorensen, M.V., Kristensen, R.M., Hejnol, A., Dunn, C.W., Giribet, G., Worsaae, K., 2015. Spiralian Phylogeny Informs the Evolution of Microscopic Lineages. Curr Biol.

Lavrov, D.V., Lang, B.F., 2005. Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. Syst Biol 54, 651-659.

Lavrov, D.V., Pett, W., Voigt, O., Worheide, G., Forget, L., Lang, B.F., Kayal, E., 2013. Mitochondrial DNA of Clathrina clathrus (Calcarea, Calcinea): Six linear chromosomes, fragmented rRNAs, tRNA editing, and a novel genetic code. Mol Biol Evol 30, 865-880.

Li, S., Chen, Y., Zhang, M., Bao, X., Li, Y., Teng, W., Liu, Z., Fu, C., Wang, Q., Liu, W., 2014. Complete mitochondrial genome of the marine polychaete, *Marphysa sanguinea* (Polychaeta, Eunicida). Mitochondr DNA.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010.

Mwinyi, A., Bailly, X., Bourlat, S.J., Jondelius, U., Littlewood, D.T.J., Podsiadlowski, L., 2010. The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of *Symsagittifera roscoffensis*. BMC Evol Biol 10:309.

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

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. Mol Biol Evol 32, 268-274.

Noguchi, Y., Endo, K., Tajima, F., Ueshima, R., 2000. The mitochondrial genome of the brachiopod *Laqueus rubellus*. Genetics 155, 245-259.

Osca, D., Irisarri, I., Todt, C., Grande, C., Zardoya, R., 2014. The complete mitochondrial genome of *Scutopus ventrolineatus* (Mollusca: Chaetodermomorpha) supports the Aculifera hypothesis. BMC Evol Biol 14: 197.

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.

Renaud, G., Kircher, M., Stenzel, U., Kelso, J., 2013. freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. Bioinformatics 29, 1208-1209.

Rouse, G.W., Pleijel, F., 2001. Polychaetes. Oxford University Press.

Shadel, G.S., Clayton, D.A., 1997. Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66, 409 435.

Shen, X., Ma, X.Y., Ren, J.F., Zhao, F.Q., 2009. A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of Phascolosoma esculenta. BMC Genomics 10:136.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Stechmann, A., Schlegel, M., 1999. Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. P Roy Soc B-Biol Sci 266, 2043-2052.

Steinauer, M.L., Nickol, B.B., Broughton, R., Orti, G., 2005. First sequenced mitochondrial genome from the phylum Acanthocephala (*Leptorhynchoides thecatus*) and its phylogenetic position within metazoa. J Mol Evol 60, 706-715.

Struck, T.H., 2012. Phylogeny of Annelida. In: Schmidt-Rhaesa, A. (Ed.), Handbook of Zoology Online. De Gruyter, Berlin, Boston.

Struck, T.H., Golombek, A., Weigert, A., Franke, F.A., Westheide, W., Purschke, G., Bleidorn, C., Halanych, K.M., 2015. The evolution of annelids reveals two adaptive routes to the interstitial realm. Curr Biol in press.

Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hosel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., Bleidorn, C., 2011. Phylogenomic analyses unravel annelid evolution. Nature 471, 95-98.

Struck, T.H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., Halanych, K.M., 2007. Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol Biol 7: 57.

Valles, Y., Boore, J.L., 2006. Lophotrochozoan mitochondrial genomes. Integr Comp Biol 46, 544-557.

Valles, Y., Halanych, K.M., Boore, J.L., 2008. Group II introns break new boundaries: Presence in a bilaterian's genome. Plos One 3, e1488.

Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C., Struck, T.H., 2014. Illuminating the base of the annelid tree using transcriptomics. Mol Biol Evol 31, 1391-1401.

Wu, Z.G., Shen, X., Sun, M.A., Ren, J.F., Wang, Y.J., Huang, Y.L., Liu, B., 2009. Phylogenetic analyses of complete mitochondrial genome of *Urechis unicinctus* (Echiura) support that echiurans are derived annelids. Mol Phylogenet Evol 52, 558-562.

Zhong, M., Struck, T.H., Halanych, K.M., 2008. Phylogenetic information from three mitochondrial genomes of Terebelliformia (Annelida) worms and duplication of the methionine tRNA. Gene 416, 11-21.

5.8. Syllidae mitochondrial gene order is unusually variable for Annelida

M. Teresa Aguado[1], Sandy Richter[2,6], Rebekka Sontowski[2,6], Anja Golombek[3], Torsten H. Struck[4] and Christoph Bleidorn[5,6]

[1] Departamento de Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain.
[2] Molecular Evolution and Systematics of Animals, Institute of Biology, University of Leipzig, Talstraße 33, D-04103 Leipzig, Germany.
[3] Zoological Research Museum Alexander Koenig, Adenauerallee 160, D-53113 Bonn, Germany
[4] National Centre for Biosystematics, Natural History Museum, University of Oslo, P.O. Box 1172, Blindern, NO-0318 Oslo, Norway
[5] Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, Spanish National Research Council (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain
[6] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

Corresponding author: M. Teresa Aguado, e-mail: maite.aguado@uam.es

Abstract
Complete mitochondrial genomes of five syllids (*Streptosyllis* sp., *Eusyllis blomstrandi*, *Myrianida brachycephala*, *Typosyllis antoni* and *Typosyllis* sp.) have been obtained using Illumina sequencing. Together with two previous studied taxa (*Ramisyllis multicaudata* and *Trypanobia cryptica*), the analysed sequences represent most of the main lineages within the family Syllidae (Anoplosyllinae, Eusyllinae, Autolytinae and Syllinae). The genomic features, gene order and phylogenetic relationships are examined. Unusual for annelids, syllid mitochondrial genomes are highly variable in their gene order. Considering genomic features, such as length, skewness, gene content, and codon bias, most similar to the rest of annelids are the genomes of *E. blomstrandi* and *M. brachycephala*, while *Streptosyllis* sp. and the analysed sylline taxa (*R. multicaudata*, *T. cryptica*, *T. antoni* and *Typosyllis* sp.) are the most dissimilar. Two methionine tRNA's (*trnM*) have been found in *T. antoni* and *Typosyllis* sp. The mt genomes of these latter taxa are the longest with numerous non-coding regions. The 13 protein coding genes, as well as the rRNA's are used to perform phylogenetic analyses that recovered the relationships within the family explored before by previous authors. The gene order in Syllidae shows very different patterns. *E. blomstrandi* and *M. prolifera* show a similar pattern to the one found in Pleistoannelida; however this might have changed at least twice within Syllidae: in *Streptosyllis* sp. and within Syllinae. All analysed Syllinae show different gene order, thereby illustrating more variability as all other pleistoannelids analysed so far. The information provided herein allows a more accurate reconstruction of the possible evolutionary scenarios in Syllidae.

Key words: Anoplosyllinae, Autolytinae, Eusyllinae, mitochondrial genome, phylogeny, Syllinae

1. Introduction
Animal mitochondrial (mt) genomes are circular duplex molecules of DNA that usually contain 13 protein coding genes, 22 tRNA's and two rRNA's (Boore 1999). The analysis of the sequences and the order in which these genes are organized provide valuable information that has been widely used to resolve phylogenetic relationships of different taxa at different levels (Vallès and Boore 2006). The mt genomes have been shown to be quite suitable for younger divergences, though less informative for deep phylogenies (Bernt et al. 2013a). Within

Annelida, comparatively few analyses dealing with complete mt genomes were published during the first decade of the 2000s (e.g., Boore 2004; Jennings and Halanych 2005; Bleidorn et al. 2006; Vallès and Boore 2006; Zhong et al. 2008; Mwinyi et al. 2009; Shen et al. 2009). By using next-generation sequencing technology the generation of mt genomes became more feasible and less time intensive. Including the here newly published sequences, 89 complete annelid mt genomes were available in April 2016 (Supplementary Table 1). However, whereas there are a high number of genomes available for Glyceridae, Clitellata and Siboglinidae (Li et al. 2015; Richter et al. 2015; Zhang et al. 2016), most other annelid taxa are less well represented. Based on the available data, the arrangement of genes in Annelida was considered as highly conserved. Moreover, most of the available mt genomes belong to one of the two major groups within annelids: Errantia or Sedentaria, that together form the large clade Pleistoannelida (Struck 2011). Recently, Weigert et al. (2016) included several basal branching annelids and found that the gene order was more variable than expected. The common pattern previously assigned to Annelida could be considered as the ground pattern in Pleistoannelida, while the basal branching groups showed a completely different arrangement of genes. Within Pleistoannelida, a taxon which origin dates back to the late Cambrian – Early Ordovician (Hints and Eriksson 2007; Weigert and Bleidorn 2016), the order of protein coding and ribosomal genes is highly conserved, with the Echiura, Ampharetidae and *Diurodrilus* as the only exceptions (Boore 2004; Wu et al. 2009; Zhong et al. 2011; Golombek et al. 2013). All these three taxa belong to Sedentaria.

Aguado et al. (2015a) analysed the first two mt genomes of Syllidae, one of the largest groups not only in Errantia, but in Annelida in general. The Syllidae currently comprises more than 700 spp. that inhabit practically all marine benthic realms (Aguado et al. 2015b). Surprisingly, Aguado et al. (2015a) found that the gene order of the two investigated syllids (*Ramisyllis multicaudata* and *Trypanobia cryptica*) was completely different from the ground pattern proposed for Pleistoannelida. However, since these were the first two mt genomes from syllids, the authors could not assess if Syllidae in general or only members of the more recent clade including *Ramisyllis* and *Trypanobia* show these different gene orders.

The phylogenetic relationships within Syllidae were initially studied including few taxa (Licher 1999; Nygren 1999; Nygren and Sundberg 2003) or focusing on sub-taxa (i.e. *Typosyllis* by Licher (1999) and Autolytinae by Nygren (2004)). Some years later Aguado et al. (2007) performed a more comprehensive study including molecular information (the nuclear gene *18S* and the mitochondrial *rrnL* and *cox1*) from a larger amount of taxa (88 terminals). Aguado et al. (2012) included molecules (same genes) and morphology of 213 terminals. More recently, Aguado et al. (2015a; 2015b) performed new molecular analyses, including more terminals when compared with previous studies (genera *Alcyonosyllis*, *Trypanobia* and *Ramisyllis*). All analyses agreed in recovering the family Syllidae as monophyletic, early divided into two major clades: one corresponding to the subfamily Anoplosyllinae and the other one comprising the rest of syllids. Within the second one, the traditional subfamilies were found to be monophyletic: Autolytinae, Exogoninae, Syllinae, and Eusyllinae. The latter has been reorganized by Aguado et al. (2012). Some genera could not be assigned to any of these subfamilies and are currently considered as independent groups (such as *Anguillosyllis*, *Amblyosyllis* and *Perkinsyllis*) (Aguado et al. 2012). However, the available molecular information of most of syllids corresponds only to three genes, *18S*, *rrnL* and *cox1*.

In order to investigate mitochondrial gene order evolution and possible divergence scenarios within Syllidae, we have analysed 5 new complete mt genomes of syllids using next generation sequencing techniques. These new mt genomes, together with *R. multicaudata* and *T. cryptica*, represent most of the main groups within the family: Anoplosyllinae, Eusyllinae, Autolytinae and Syllinae. Additionally, the data provided herein allow to perform phylogenetic analyses including less taxa than previous analyses (Aguado et al., 2007, 2012, 2015 a, b), though much

larger amount of data (15 genes and 12,100-13,200 bp). Results provide more evidence about phylogenetic relationships within Syllidae.

2. Materials and methods

2.1. Taxa included

Specimens of *Eusyllis blomstrandi* and *Myrianida brachycephala* were collected near Helgoland (Germany); *Streptosyllis* sp. was collected in Sylt (Germany); and *Typosyllis antoni* and *Typosyllis* sp. were obtained from bought live rock of the Indian Ocean kept in the aquarium in Leipzig (Germany). The two other syllids, *Ramisyllis multicaudata* and *Trypanobia cryptica*, as well as outgroup mt genomes, were downloaded from Genbank (Table 1).

Table 1: Terminals used in the analyses and GenBank accession numbers

Taxon	Family	Acc code
Nephtys sp.	Nephtyidae	EU293739
Glycera capittata	Glyceridae	KT989320
Platynereis dumerilii	Nereididae	AF178678
Streptosyllis sp.	Syllidae; Anoplosyllinae	xxx
Eusyllis blomstrandi	Syllidae; Eusyllinae	xxx
Myrianida brachycephala	Syllidae; Autolytinae	xxx
Ramisyllis multicaudata	Syllidae; Syllinae	KR534502
Trypanobia cryptica	Syllidae; Syllinae	KR534503
Typosyllis antoni	Syllidae; Syllinae	xxx
Typosyllis sp.	Syllidae; Syllinae	xxx

2.2. Genome sequencing

Specimens were partly fixed in 95% ethanol for DNA sequencing and DNA was extracted from single individuals by using commercial kits. For *M. brachycephala*, *E. blomstrandi* and the two *Typosyllis* species Illumina sequencing, double index sequencing libraries with average insert sizes of around 300 bp were prepared as previously described (Meyer and Kircher 2010; Kircher et al. 2012). The libraries were sequenced as either 100 or 140 bp paired-end runs on an Illumina Hi-Seq 2000. Base calling was performed with freeIbis (Renaud et al. 2013), adaptor and primer sequences were removed using leeHom (Renaud et al. 2014), and reads with low complexity and false paired indices were discarded. Quality control was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw data of all libraries were trimmed by removing low quality reads. The quality of all sequences after trimming was checked using Trimmomatic v0.32 (Bolger et al. 2014). DNA extraction, genome amplification, library preparation and sequencing for *Streptosyllis* sp. followed the procedure described in Golombek et al. (2013). *De novo* assemblies were conducted with CLC Genomics Workbench 5.1 (CLC bio, Århus, Denmark) using default settings (*Streptosyllis* sp.) and with IDBA-UD (Peng et al. 2012), using an initial *k*-mer size of 21, an iteration size of 10 and a maximum *k*-mer size of 91 (all other syllids). All sequence data were submitted to the National Centre for Biotechnology Information (NCBI) (Table 1). The coverage of mt genomes was estimated by mapping sequence reads back to the contig comprising the mt genome using BWA under default options (Li and Durbin 2009). Visual exploration of the coverage of all assemblies was performed with the program Tablet (Milne et al. 2013).

2.3. Mitochondrial genome annotation and gene order analyses

The mt genomes were annotated using the MITOS webserver (Bernt et al. 2013b) with the "Invertebrates" NCBI code for translation. This server also provided the secondary structure of tRNA's and rRNA's. All automatic annotations were finally manually curated. AT and GC skew were determined for the complete mt genomes according to the formula defined by Perna and Kocher (1995), AT skew = (A - T)/(A + T) and GC skew = (G - C)/(G + C), where the letters stand for the number of the corresponding nucleotides in the sequences. Characterization of codon usage bias was calculated with the program DAMBE5 (Xia 2013).
We used CREx (Bernt et al. 2007) to conduct pairwise comparisons of the mitochondrial gene order. CREx determines the most parsimonious genome rearrangement scenario between the gene order of each pair of taxa including transpositions, reverse transpositions, reversals, and tandem-duplication-random-loss (tdrl) events. The analysis was performed applying the common intervals parameter for distance measurement. Two different analyses were performed, the first one including all the genes (protein coding genes, rRNA's and tRNA's), while in the second tRNA's were excluded, since they are usually more variable than other genes.

2.4. Phylogenetic analyses

Datasets of all mitochondrial protein-coding genes, together with *rrnL* and *rrnS* from the syllids and outgroups (Table 1) were constructed to perform phylogenetic analyses. Alignments of orthologous partitions were performed using the program Mafft (Katoh et al. 2002) with the default parameters for the protein coding genes and the iterative refinement method E-INS-i, and default gap open and extension values for *rrnL* and *rrnS*. Nucleotide alignments of protein coding genes were generated by back-translation for each gene using the software Translator X (Abascal et al. 2010). Ambiguously aligned and variable regions were recognized and excluded using the program Gblocks (Castresana 2000) with relaxed parameters (smaller final blocks, gap positions within the final blocks, and less strict flanking positions allowed). Concatenation of partitions for the combined data set was conducted with FASconCAT-G (Kück and Longo 2014). Several data sets were examined: 1. Alignments of nucleotides and *rrnL+rrnS*; 2. Alignments of amino acids; 3. Alignments of nucleotides and *rrnL+rrnS* without ambiguous regions; 4. Alignments of amino acids without ambiguous regions.
Maximum Likelihood (ML) analyses were conducted using RAxML version 8.1.3 (Stamatakis 2014), with GTR+I+G model, as GTR is the only available nucleotide model in this program. + For amino acids we chose the MtZoa +G +F model of sequence evolution, which has been specifically designed for lophotrochozan mt genomes (Rota-Stabelli et al. 2009). Bootstrap support values were generated with a rapid bootstrapping algorithm (Stamatakis et al. 2008) for 1000 replicates in RAxML. Analysis of the alignments of nucleotides and *rrnL+rrnS* without ambiguous regions was also performed through bayesian inference methods (BI). The program MRBAYES v. 3.2.6 (Huelsenbeck and Ronquist, 2001) was used to combine two independent runs of 1,000,000 generations each with three heated and one cold Monte Carlo Markov chain (MCMC), starting from default prior values and random trees and applying the GTR+I+G model. Trees were sampled every 1,000 generations each. All parameters were unlinked, and rates were allowed to vary freely over partitions. After discarding 250 first trees as burn-in, trees from the stationary phase were combined to obtain a majority rule consensus and posterior node probabilities (Huelsenbeck and Ronquist, 2001).

3. Results and discussion

3.1. Mitochondrial genomes

BLAST-searches identified the complete mitochondrial genomes of all investigated syllids as a single contig. All mt genomes are highly covered by sequence reads: *Myrianida brachycephala* 164x*, Eusyllis blomstrandi* 158x*, Typosyllis antoni* 206x*, Typosyllis* sp. 222x, *Streptosyllis* sp. 119x. All mt genomes are deposited in Genbank and accession numbers can be found in Table 1. The complete mt genomes of *Streptosyllis* sp., *E. blomstrandi* and *M. brachycephala* are around 15,000 bp long (Table 2). The mt genomes of *T. antoni* and *Typosyllis* sp. are longer, with 16,902 and 16,241 bp, respectively. The mt genome size of other annelids usually varies between 15,000 and 16,000 bp (see Supplementary Table 1). The longest mt genomes are those of sylline taxa, including the two previously investigated syllids, *Ramisyllis multicaudata* and *Trypanobia cryptica* (15,748 bp and 16,630 bp, respectively) (Aguado et al. 2015a) (Table 2). The gene orders of the five new mt genomes are illustrated in Fig. 1.

Table 2: Mitochondrial genome length, nucleotide contents, and nucleotide skew. Ramisyllis multicaudata and T. cryptica after Aguado et al. (2015a).

	Mt genome length (bp)	*A*	*C*	*G*	*T*	*GC*	*AT*	*GC skew*	*AT skew*
***Streptosyllis* sp.**	14983	31%	12%	14%	43%	26%	74%	0.006	-0.18
Eusyllis blomstrandi	14712	32%	15%	14%	39%	29%	71%	-0.006	-0.09
Myrianida brachycephala	15032	31%	19%	12%	38%	31%	69%	-0.021	-0.01
Typosyllis antoni	16902	35%	18%	11%	36%	29%	71%	-0.023	-0.001
***Typosyllis* sp.**	16241	34%	20%	13%	33%	33%	67%	-0.024	0.004
Ramisyllis multicaudata	15748	34%	22%	11%	33%	33%	67%	-0.031	0.02
Trypanobia cryptica	16630	36%	20%	11%	33%	31%	69%	-0.029	0.04

The syllid genomes are AT-rich (around 70%) (Table 2). In *Streptosyllis* sp., *E. blomstrandi* and *M. brachycephala*, the most common base is T (values around 40%), while in *T. antoni* and *Typosyllis* sp. A and T are approximately equally frequent (34-35%) and more common than G and C (Table 2). In *R. multicaudata* and *T. cryptica*, A is the most common base (34- 36%). In the mt genomes of *Streptosyllis*, *E. blomstrandi*, *M. brachycephala* and *T. antoni*, the AT skew is negative (-0.18, -0.09, -0.1, -0.01, respectively) (Table 2). It is positive though very low in *Typosyllis* sp. (0.004). The GC skew is negative in *E. blomstrandi*, *M. brachycephala, T. antoni* and *Typosyllis* sp. (-0.06, -0.21, -0.23, -0.24, respectively), while it is positive in *Streptosyllis* sp. (0,06) (Table 2). In *R. multicaudata* and *T. cryptica*, the AT skews are positive, while GC skews negative (Aguado et al. 2015a). In most studied annelids, the GC skew is usually more negative than the AT skew (Bleidorn et al. 2006; Zhong et al. 2008; Richter et al. 2015; Li et al. 2016; Patra et al. 2016).

For each of the five mt genomes sequenced in this study, all 13 protein coding genes, two rRNA's and the typical 22 tRNA's could be detected (Fig. 1, Supplementary Tables 2-6). Two genes in *Streptosyllis* sp. showed a reading frame shift due to the insertion of single bases (*nad3*, *nad4L*). These shifts were also supported by sequence reads mapped back to the mitochondrial contig. However, as during library preparation whole genome amplification was conducted for this species, this is likely a PCR artefact. Consequently, we manually curated this shifts using

alignments from the other syllids as a reference. In *T. antoni* and *Typosyllis* sp. one additional gene was found, as the methionine tRNA (*trnM*) is duplicated (Supplementary Tables 5, 6). The two *trnMs* are located one after another (Fig. 1). The *trnM* has previously been also found duplicated in the mt genomes of *Pista cristata Terebellides stroemi,* and *Pectinaria gouldi* (Zhong et al. 2008; Zhong et al. 2011), *Phascolosoma esculenta* (Shen et al. 2009), and also in *Chaetopterus variopedatus* and *Eurythoe complanata* (Weigert et al. 2016). As in the case for most of annelids so far studied, except *Owenia fusiformis* and *Magelona mirabilis* (Weigert et al. 2016), all genes are transcribed from the same strand (referred to as plus-strand).
Start codons in the protein-coding genes are highly biased towards ATG (Supplementary Tables 2-6). Other start codons, such as ATA, ATT, AGA, TTG, and GTG, have been found in *Streptosyllis* sp., *T. antoni* and *Typosyllis* sp. The stop codons are mostly TAA and, in less amount TAG. Occasionally, some genes of *Streptosyllis* sp., *T. antoni* and *Typosyllis* sp. end in TA or simply T. These variations from the typical stop and start codons have been also found in *R. multicaudata* and *T. cryptica* (Aguado et al. 2015a), and incomplete stop codons are also present in other annelids (Bleidorn et al. 2006; Li et al. 2016). Interestingly, alternative start and stop codons only appear in those syllids with highly rearranged genomes (i.e. *Streptosyllis* sp. and the sylline species). There is also a codon usage bias in the investigated genomes (Supplementary Table 7). In general, NNA and NNT codons are the most common codon types, while NNG codons are the least used. In general, the syllids share with other annelids a similar pattern of codon usage bias (Jennings and Halanych 2005; Bleidorn et al. 2006; Zhong et al. 2008; Zhong et al. 2011; Patra et al. 2016).

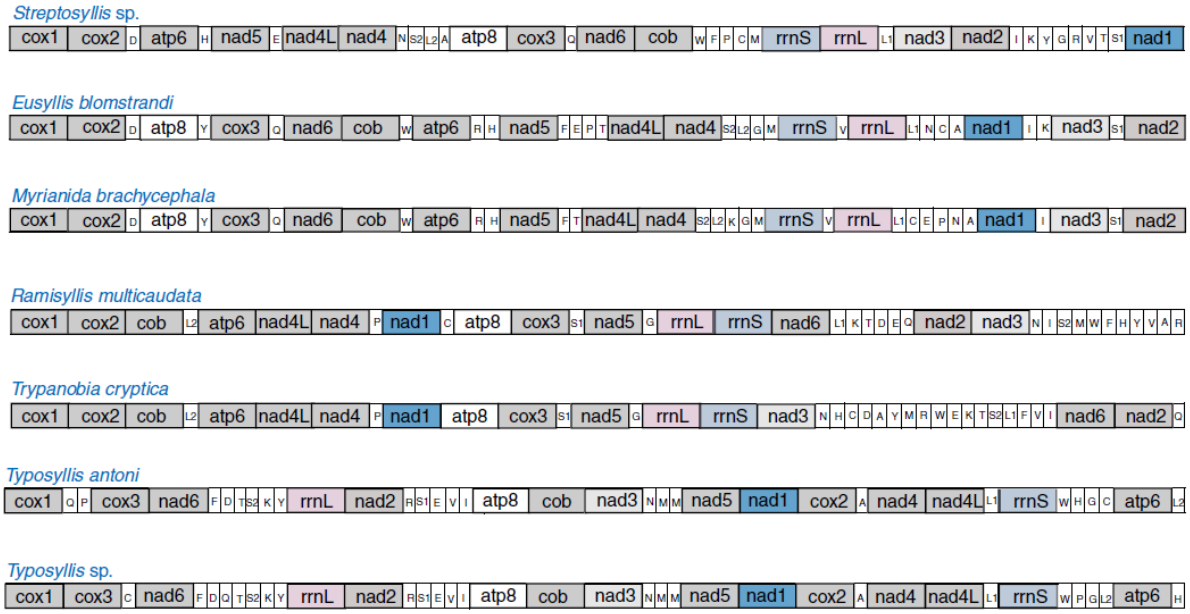


Fig. 1: Mitochondrial genomes from Syllidae. Form up to down: *Streptosyllis* sp., *Eusyllis blomstrandi*, *Myrianida brachycephala*, *Ramisyllis multicaudata*, *Trypanobia cryptica*; *Typosyllis antoni* and *Typosyllis* sp. *Ramisyllis multicaudata* and *T. cryptica* after Aguado et al. (2015a).

The tRNA's mostly possess the common cloverleaf structure, with an acceptor arm, anticodon arm, TΨC arm, DHU arm, and associated loop regions. The DHV stem is missing in the *trnR* of the five investigated mt genomes. The DHV stem is also missing in the *trnS2* of *Streptosyllis* sp., *T. antoni* and *Typosyllis* sp., in the *trnS1* of *Streptosyllis* sp. and *Typosyllis* sp., and in the *trnC* of *Typosyllis* sp.; while it is reduced in the *trnS1* and *trnS2* of *E. blomstrandi* and *trnS1* of *T. antoni*. The TΨC arm is missing in the *trnH* of *Streptosyllis* sp. In *M. brachycephala*, the *trnS1* has the DHV arm modified as a large loop and it has one short additional arm between the anticodon and the TΨC arm. Other syllids, such as *R. multicaudata* and *T. cryptica* also

show *trnS1* and *trnS2* with a shortened DHV stem. The DHU arm, and less often the TΨC arm, is also missing or reduced in the tRNA's of other annelids (Jennings and Halanych 2005; Bleidorn et al. 2006; Zhong et al. 2008; Mwinyi et al. 2009; Richter et al. 2015). The lengths of the ribosomal genes in the syllids are similar to each other, excepting in *Streptosyllis* sp., where they are shorter (*rrnL*: 587 bp, *rrnS*: 532 bp), especially comparing with the sylline. In *E. blomstrandi* the sizes of the rRNA's are: *rrnL*: 1000 bp, *rrnS*: 742 bp; in *M. brachycephala*, *rrnL*: 1011 bp, *rrnS*: 756 bp; in *T. antoni*, *rrnL*: 1000 bp, *rrnS*: 877 bp; and finally, in *Typosyllis* sp. the largest, *rrnL*: 1001 bp, *rrnS*: 904 bp (Supplementary Tables 2-6). The latter values are similar to *R. multicaudata*, *rrnL*: 1008 bp, *rrnS*: 787 bp; and *T. cryptica*, *rrnL*: 1007 bp, *rrnS*: 789 bp (Aguado et al. 2015a).

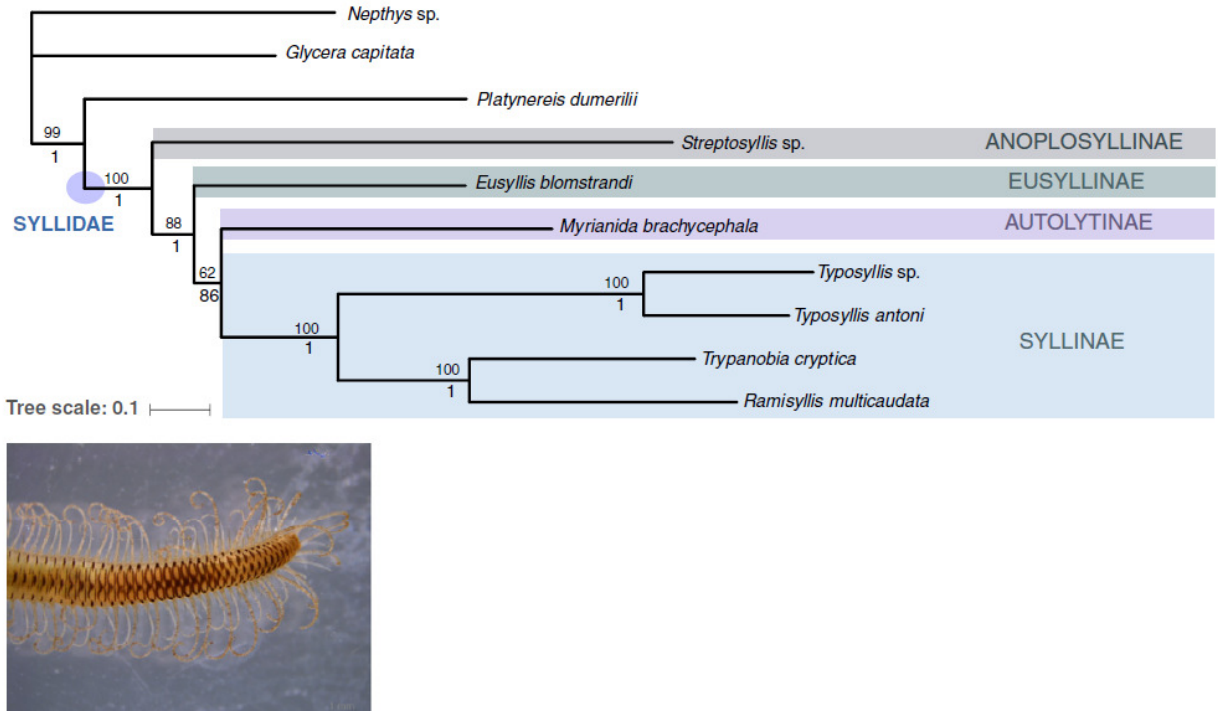


Fig. 2: Maximum likelihood tree obtained when analyzing nucleotides from protein coding genes and *rrnL* and *rrnS*, ambiguous regions excluded from all the genes. Bootstrap support values are below nodes. Picture of *Typosyllis* sp.

The non-coding regions present in the mt genomes vary considerably. In *E. blomstrandi* and *M. brachycephala*, most of genes are overlapping or separated by very short non-coding regions of few bp (Supplementary Tables 3, 4), except the non-coding region between the *trnV* and *rrnL* that are 136 bp and 135 bp, respectively. In contrast, *Streptosyllis* sp., *T. Antoni* and *Typosyllis* sp., show several longer non-coding regions with sizes between 50 and 445 bp (Supplementary Tables 2, 5, 6). Interestingly, BLAST searches did not recover any hints that these regions stem from tandem duplication random loss (tdrl) events (Moritz et al. 1987). In the five mt genomes, the longest non coding regions are AT rich and are suggested to be the putative control regions. The presence of non-coding regions of variable sizes through the mt genome has also been documented for other annelids (Jennings and Halanych 2005; Bleidorn et al. 2006; Weigert et al. 2016).

3.2. Phylogenetic analyses

The results obtained from nucleotide data sets (protein-coding genes and *rrnL+rrnS*), as well as those obtained when analyzing the amino acid data sets, widely agree in the topologies,

independently of the exclusion of the ambiguous regions in the alignments. Results obtained through ML and BI are also congruent. Syllidae is always recovered as a monophyletic group, though the support values depend on the analysis. The bootstrap (BS) and posterior probability (PP) values are high (100% and 1 respectively) in the analysis obtained from nucleotides excluding the ambiguous regions (Fig. 2), while low (75%) or very low (39%) in the analyses obtained from amino acids including the ambiguous regions or excluding them, respectively (Supplementary Figs. 1A-C).

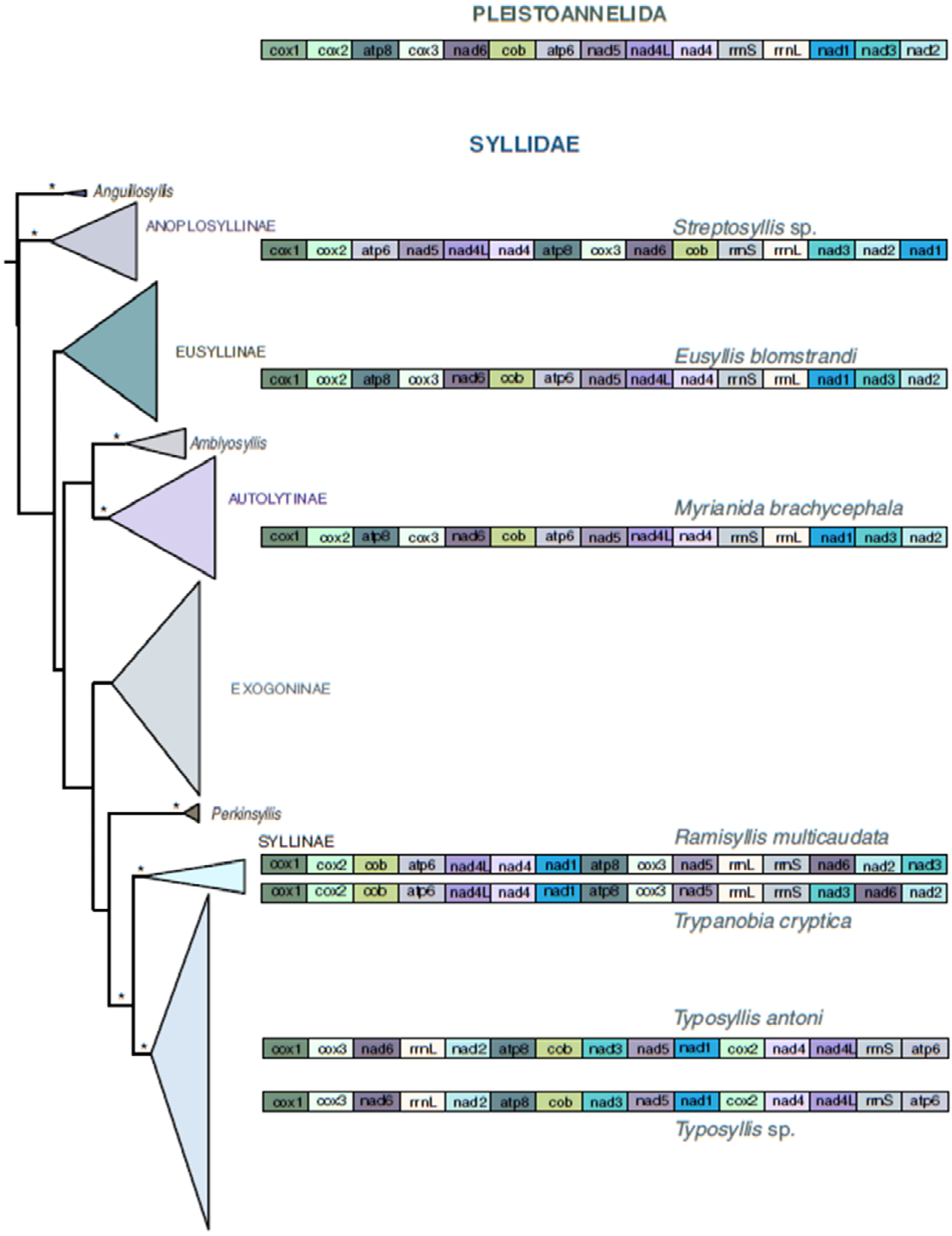


Fig. 3: Gene order in Syllidae. Mitochondrial genomes represented without *tRNA's*. The phylogenetic relationships are after Aguado et al. (2015a). * indicate well supported clades. Gene order in Pleistoannelida after Weigert et al. (2016).

Streptosyllis sp., a member of Anoplosyllinae, is part of Syllidae in all analyses. Usually Anoplosyllinae is the sister group to the rest of syllids or part of a basal polytomy (Fig. 2, Supplementary Figs. 1A, B). Only in the topology obtained when analyzing amino acids without ambiguous regions (Supplementary Fig. 1C), *Streptosyllis* sp. is sister to *M. brachycephala,* but the support value of this clade is considerably low (48%). *Eusyllis blomstrandi* and *M. brachycephala*, members of Eusyllinae and Autolytinae, respectively, vary their positions as well depending on the analyses, but all the Syllinae are in a wellsupported monophyletic group (Fig. 2, Supplementary Figs. 1A-C). Within Syllinae, *T. Antoni* and *Typosyllis* sp. always represent well-supported sister taxa, and *R. multicaudata* and *T. cryptica* are in a well-supported clade as well, except in the results from amino acids with ambiguous regions, where the relationships within Syllinae were not fully resolved (Supplementary Fig. 1B). In general, the topologies obtained when analyzing nucleotides are better resolved than those obtained from amino acids.

The results are in agreement with previous phylogenetic analyses (Aguado et al. 2007, 2012, 2015a, 2015b) which have been performed based on only three genes, one nuclear (*18S*) and two mitochondrial (*rrnL* and *cox1*), which all together were approximately 3000 bp. The herein performed phylogenetic analyses include 13 mitochondrial coding genes and two rRNA genes (*rrnL* and *rrnS*), which sum approximately 12,100-13,200 bp, depending on each taxon. The obtained results also support the conclusions of Aguado and Bleidorn (2010), who explained that a different location of Anoplosyllinae, outside syllids, was an artifact due to the effect in the alignment of a hypervariable region in the nuclear gene *18S*. In summary, all results considered suggest that the phylogenetic relationships of major clades within Syllidae are well established.

3.3. Gene order

The gene order within annelids had been considered as relatively conserved (e.g., Jennings and Halanych 2005; Vallès and Boore 2006; Zhong et al. 2008). However, Weigert et al. (2016) examined several basal branching terminals and demonstrated that the supposedly highly conserved mitochondrial gene order in Annelida is restricted to Pleistoannelida (which includes Errantia and Sedentaria), and likely represents the ground pattern of this group.

Within Syllidae, we found the Pleistoannelida ground pattern in two of the seven analyzed genomes: *E. blomstrandi* and *M. brachycephala* (members of Eusyllinae and Autolytinae, respectively) (Fig. 3). However, the gene order changes considerably in *Streptosyllis* sp. (Anoplosyllinae) and especially in the Syllinae. In Syllinae, the four terminals included, which represent the two main clades within this subfamily (*R. multicaudata* and *T. cryptica* on one hand, and *T. antoni* and *Typosyllis* sp. on the other) show two completely different gene arrangements (Figs. 1, 3).

The CREx analyses reveal possible scenarios for these two divergences from the common pattern. When compared with *P. dumerilii* (representing the pleistoannelid ground pattern), the gene order of *Streptosyllis* sp. might be the result of 2 transpositions and 3 tdrl (including tRNA's), or only 2 transpositions (excluding *tRNA's*). The second deviation occurs in the Syllinae where the number of tdrl increases considerably, 4-5 tdrl when including the *tRNA's*, accompanied by several transpositions in most cases; or 1-3 tdrl and 2-3 transpositions when excluding the tRNA's. The high number of non-coding regions contributing to longer mt genomes in Syllinae when compared to other syllids (see Supplementary Tables 2-6 and Aguado et al. (2015a)) and many other annelids could be hypothesized as possible remnants of several tdrl events. However, BLAST searches identified no similarities with other regions within the respective mt genomes.

Table 3: Matrix of gene order distance measure obtained from CREx analysis. tRNA's included. The highest the numbers the more similar are the compared gene orders.

	N	*G*	*P*	*S*	*E*	*M*	*R*	*Try*	*Tan*	*Typo*
Nephtys sp.	1326	1256	354	38	452	264	2	2	4	2
Glycera capittata	1256	1326	376	38	450	262	2	4	4	2
Platynereis dumerilii	354	376	1326	34	254	120	2	4	2	2
Streptosyllis sp.	38	38	34	1326	40	38	12	8	4	6
Eusyllis blomstrandi	452	450	254	40	1326	396	4	4	2	4
M. brachycephala	264	262	120	38	396	1326	4	4	2	2
R. multicaudata	2	2	2	12	4	4	1326	134	8	6
Trypanobia cryptica	2	4	4	8	4	4	134	1326	14	12
Typosyllis antoni	4	4	2	4	2	2	8	14	1326	674
Typosyllis sp.	2	2	2	6	4	2	6	12	674	1326

Tables 3 and 4 summarize these results using similarity scores, where the higher the numbers the more similar are the compared gene orders. The highest values, independently of including or excluding the tRNA's from the gene order, correspond to comparisons between outgroups with *E. blomstrandi* and *M. brachycephala* and among themselves. These values are followed by *Streptosyllis* sp. with any of them, while the lowest values correspond to the relationships of the syllinae with the rest. Within Syllinae, two clear groups are also recovered: *R. multicaudata* and *T. cryptica*; and *T. antoni* and *Typosyllis* sp. The most dissimilar order to the ground pattern would be the one in *T. antoni* and *Typosyllis* sp.

Table 4: Matrix of gene order distance measure obtained from CREx analysis tRNA's excluded. The highest the numbers the more similar are the compared gene orders.

	N	*G*	*P*	*S*	*E*	*M*	*R*	*Try*	*Tan*	*Typo*
Nephtys sp.	204	204	204	54	204	204	18	16	6	6
Glycera capittata	204	204	204	54	204	204	18	16	6	6
Platynereis dumerilii	204	204	204	54	204	204	18	16	6	6
Streptosyllis sp.	54	54	54	204	54	54	12	12	4	4
Eusyllis blomstrandi	204	204	204	54	204	204	18	16	6	6
M. brachycephala	204	204	204	54	204	204	18	16	6	6
R. multicaudata	18	18	18	12	18	18	204	154	2	2
Trypanobia cryptica	16	16	16	12	16	16	154	204	2	2
Typosyllis antoni	6	6	6	4	6	6	2	2	204	204
Typosyllis sp.	6	6	6	4	6	6	2	2	204	204

Mitochondrial gene orders of Pleistoannelida could be more diverse than expected. This would be not surprising considering that only few terminals have been investigated and they have been used to represent highly diverse taxa (such as Phyllodocida and Eunicida, for instance) (Supplementary Table 1). Aguado et al. (2015a) noted that there seem to be constraints that maintain the conserved pattern; however, when these constraints are violated, many changes might be possible. The results shown herein are fitting with this notion and provide more information about the scenario of gene order evolution in syllids. *Myrianida brachycephala* (Autolytinae) and *E. blomstrandi* (Eusyllinae) show the putative ground pattern of the pleistoannelid mitochondrial gene order. This order independently changed considerably in *Streptosyllis* sp. (Anoplosyllinae) and all analyzed members of the Syllinae. The Syllinae is the most complex group within Syllidae, with still many difficult systematic issues to be resolved and biological features to be investigated. They are animals with complex life cycles involving the development of independent sexual units called stolons and branching body patterns as in *R. multicaudata*. The phylogenetic relationships reveal two clear lineages: the ribbon clade (as

named by Aguado et al. (2015a)), including *Ramisyllis* and flattened syllids, and a large clade including mostly *Typosyllis* and *Syllis* spp. The four investigated sylline mt genomes do not show a clear common pattern and even closely related pairs of species (*Typosyllis* sp. and *T. antoni; R. multicaudata* and *T. cryptica*, respectively) show more evidence of gene re-arrangements as usually found across pleistoannelids at all. Obviously, mitochondrial gene order evolution is labile in Syllidae, underlying fewer constraints than in other annelids. This suggests that gene order analyses might be phylogenetically informative in this clade. However, we cannot provide any biological reason why the evolution of mitochondrial genomes is highly dynamic within this group. Diverse mitochondrial gene orders have also been found within tunicates, molluscs and brachiopods (Iannelli et al. 2007; Rawlings et al. 2010; Stach et al. 2010; Yuan et al. 2012; Luo et al. 2015). In the latter taxon, mitochondrial genomes of the (putatively) same species (*Lingula anatina*) from different locations showed considerable differences (Luo et al. 2015). The authors speculate that maybe other mechanisms besides tdrl are responsible for the shuffling of gene orders. Instead, a transposition mechanism involving recombination with double-stranded break repair is hypothesized. This mechanism would be an example of intramolecular recombination, which among others has been already demonstrated for cnidarians (Brockman and McFadden 2012) and bird mitochondrial genomes (Sammler et al. 2011). In some studies the presence of so-called minicircles containing fragments of the mitochondrial genomes coincide with a largely rearranged gene order of the mitochondrium. The generation of these mini-circles is hypothesized as a step necessary for intramolecular recombination. By homologous recombination these minicircles are integrated back into the mt genome, thereby generating gene rearrangements. Future studies are necessary to investigate why mitochondrial gene order in some taxa is less constraint than in other and syllids seem to be a putative test case for this question.

Acknowledgments

This study is a contribution of the project "Macroevolutionary transitions in Syllidae" CGL2015-63593-P supported by the **MINECO/FEDER**, UE funds. MTA was supported by the "José Castillejo" fellow by MINECO (Ministerio de Economía y Competitividad, Spanish Government) for a research stay at the University of Leipzig (Germany). CB is supported by a "Ramón y Cajal" fellow by the MINECO, project nº RYC-2014-15615. THS acknowledges support by the German Science Foundation (DFG STR-683/6-1, 6-2 and 8-1). This is a NHM Evolutionary Genomics Lab contribution. We are grateful to the crew from the Biological Station in Helgoland (Germany) for supporting the fieldtrip.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at xxx

References

Abascal, F., R. Zardoya & M. J. Telford, 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Research 38: W7-W13.

Aguado, M. T. & C. Bleidorn, 2010. Conflicting signal within a single gene confounds syllid phylogeny (Syllidae, Annelida). Molecular Phylogenetics and Evolution 55: 1128-1138.

Aguado, M. T., A. Nygren & M. E. Siddall, 2007. Phylogeny of Syllidae (Polychaeta) based on combined molecular analysis of nuclear and mitochondrial genes. Cladistics 23: 552-564.

Aguado, M. T., G. San Martin & M. E. Siddall, 2012. Systematics and evolution of syllids (Annelida, Syllidae). Cladistics 28: 234-250.

Aguado, M. T., C. J. Glasby, P. Schroeder, A. Weigert & C. Bleidorn, 2015a. The making of a branching annelid: an analysis of complete mitochondrial genome and ribosomal data of *Ramisyllis multicaudata*. Scientific Reports srep12072.

Aguado, M. T., C. Helm, M. Weidhase & C. Bleidorn, 2015b. Description of a new syllid species as a model for evolutionary research of reproduction and regeneration in annelids. Organisms Diversity & Evolution 15: 1-21.

Bernt, M., D. Merkle, K. Ramsch, G. Fritzsch, M. Perseke, D. Bernhard, M. Schlegel, P. F. Stadler & M. Middendorf, 2007. CREx: inferring genomic rearrangements based on common intervals. Bioinformatics 23: 2957-2958.

Bernt, M., C. Bleidorn, A. Braband, J. Dambach, A. Donath, G. Fritzsch, A. Golombek, H. Hadrys, F. Juhling, K. Meusemann, M. Middendorf, B. Misof, M. Perseke, L. Podsiadlowski, B. von Reumont, B. Schierwater, M. Schlegel, M. Schrodl, S. Simon, P. F. Stadler, I. Stoger & T. H. Struck, 2013a. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. Mol Phylogenet Evol 69: 352-64.

Bernt, M., A. Donath, F. Juehling, F. Externbrink, C. Florentz, G. Fritzsch, J. Puetz, M. Middendorf & P. F. Stadler, 2013b. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69: 313-319.

Bleidorn, C., L. Podsiadlowski & T. Bartolomaeus, 2006. The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (Annelida, Orbiniidae) – A novel gene order for Annelida and implications for annelid phylogeny. Gene 370: 96-103.

Bolger, A. M., M. Lohse & B. Usadel, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Boore, J. L., 1999. Animal mitochondrial genomes. Nucleic Acids Research 27: 1767-1780.

Boore, J. L., 2004. Complete mitochondrial genome sequence of *Urechis caupo*, a representative of the phylum Echiura. Bmc Genomics 5: 67.

Brockman, S. A. & C. S. McFadden, 2012. The mitochondrial genome of *Paraminabea aldersladei* (Cnidaria: Anthozoa: Octocorallia) supports intramolecular recombination as the primary mechanism of gene rearrangement in octocoral mitochondrial genomes. Genome Biology and Evolution 4: 994-1006.

Castresana, J., 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Molecular Biology and Evolution 17: 540-552.

Golombek, A., S. Tobergte, M. P. Nesnidal, G. Purschke & T. H. Struck, 2013. Mitochondrial genomes to the rescue - Diurodrilidae in the myzostomid trap. Molecular Phylogenetics and Evolution 68: 312-326.

Hints, O. & M. E. Eriksson, 2007. Diversification and biogeography of scolecodontbearing polychaetes in the Ordovician. Palaeogeography, Palaeoclimatology, Palaeoecology 245: 95-114.

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

Iannelli, F., F. Griggio, G. Pesole & C. Gissi, 2007. The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. Bmc Evolutionary Biology 7: 155.

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

Katoh, K., K. Misawa, K. i. Kuma & T. Miyata, 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30: 3059-3066.

Kircher, M., S. Sawyer & M. Meyer, 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research 40: e3.

Kück, P. & G. C. Longo, 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Frontiers in Zoology 11:1-8.

Li, H. & R. Durbin, 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754 - 1760.

Li, Y., K. M. Kocot, C. Schander, S. R. Santos, D. J. Thornhill & K. M. Halanych, 2015. Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida). Molecular Phylogenetics and Evolution 85: 221-229.

Li, S., Y. Chen, M. Zhang, X. Bao, Y. Li, W. Teng, Z. Liu, C. Fu, Q. Wang & W. Liu, 2016. Complete mitochondrial genome of the marine polychaete, *Marphysa sanguinea* (Polychaeta, Eunicida). Mitochondrial DNA Part A 27: 42-43.

Licher, F. 1999, Revision of der Gattung *Typosyllis* Langerhans, 1879 (Polychaeta: Syllidae). Morphologie, Taxonomie und Phylogenie. Abhl. Senckenberg. Naturforsch. Gesells. 551: 1-336.

Luo, Y.-J., N. Satoh & K. Endo, 2015. Mitochondrial gene order variation in the brachiopod *Lingula anatina* and its implications for mitochondrial evolution in lophotrochozoans. Marine Genomics 24, Part 1: 31-40.

Meyer, M. & M. Kircher, 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010: pdb.prot5448-.

Milne, I., G. Stephen, M. Bayer, P. J. A. Cock, L. Pritchard, L. Cardle, P. D. Shaw & D. Marshall, 2013. Using Tablet for visual exploration of second-generation sequencing data. Briefings in Bioinformatics 14: 193 202.

Moritz, C., T. E. Dowling & W. M. Brown, 1987. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. Annual Review of Ecology and Systematics 18: 269-292.

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

Nygren, A, 1999. Phylogeny and reproduction in Syllidae (Polychaeta). Zool. J. Linn. Soc. 126: 365-386.

Nygren, A, 2004. Revision of Autolytinae (Syllidae: Polychaeta). Zootaxa 680: 1-314.

Nygren, A. & P. Sundberg, 2003. Phylogeny and evolution of reproductive modes in Autolytinae (Syllidae, Annelida). Mol. Phyl. Evol. 29: 235-249.

Patra, A. K., Y. M. Kwon, S. G. Kang, Y. Fujiwara & S.-J. Kim, 2016. The complete mitochondrial genome sequence of the tubeworm *Lamellibrachia satsuma* and structural conservation in the mitochondrial genome control regions of Order Sabellida. Marine Genomics 26: 63-71.

Peng, Y., H. C. M. Leung, S. M. Yiu & F. Y. L. Chin, 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28: 1420-1428.

Perna, N. & T. Kocher, 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution 41: 353-358.

Rawlings, T. A., M. J. MacInnis, R. Bieler, J. L. Boore & T. M. Collins, 2010. Sessile snails, dynamic genomes: gene rearrangements within the mitochondrial genome of a family of caenogastropod molluscs. Bmc Genomics 11: 440.

Renaud, G., M. Kircher, U. Stenzel & J. Kelso, 2013. freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. Bioinformatics 29: 1208-1209.

Renaud, G., U. Stenzel & J. Kelso, 2014. leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Research 42: e141.

Richter, S., F. Schwarz, L. Hering, M. Böggemann & C. Bleidorn, 2015. The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). Genome Biology and Evolution 7: 3443-3462.

Rota-Stabelli, O., Z. Yang & M. J. Telford, 2009. MtZoa: A general mitochondrial amino acid substitutions model for animal evolutionary studies. Molecular Phylogenetics and Evolution 52: 268-272.

Sammler, S., C. Bleidorn & R. Tiedemann, 2011. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. Bmc Genomics 12: 35.

Shen, X., X. Ma, J. Ren & F. Zhao, 2009. A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of Phascolosoma esculenta. Bmc Genomics 10: 1-11.

Stach, T., A. Braband & L. Podsiadlowski, 2010. Erosion of phylogenetic signal in tunicate mitochondrial genomes on different levels of analysis. Molecular Phylogenetics and Evolution 55: 860-870.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.

Stamatakis, A., P. Hoover & J. Rougemont, 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology 57: 758-771.

Struck, T. H., 2011. Direction of evolution within Annelida and the definition of Pleistoannelida. Journal of Zoological Systematics and Evolutionary Research 49:340-345.

Vallès, Y. & J. L. Boore, 2006. Lophotrochozoan mitochondrial genomes. Integrative and Comparative Biology 46: 544-557.

Weigert, A. & C. Bleidorn, 2016. Current status of annelid phylogeny. Organisms Diversity & Evolution: 1-18.

Weigert, A., A. Golombek, M. Gerth, F. Schwarz, T. H. Struck & C. Bleidorn, 2016. Evolution of mitochondrial gene order in Annelida. Molecular Phylogenetics and Evolution 94: 196-206.

Wu, Z. G., X. Shen, M. A. Sun, J. F. Ren, Y. J. Wang, Y. L. Huang & B. Liu, 2009. Phylogenetic analyses of complete mitochondrial genome of *Urechis unicinctus* (Echiura) support that echiurans are derived annelids. Molecular Phylogenetics and Evolution 52: 558-562.

Xia, X., 2013. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. Molecular Biology and Evolution 30: 1720-1728.

Yuan, Y., Q. Li, H. Yu & L. Kong, 2012. The complete mitochondrial genomes of six heterodont bivalves (Tellinoidea and Solenoidea): Variable gene arrangements and phylogenetic implications. Plos One 7: e32353.

Zhang, L., P. Sechi, M. Yuan, J. Jiang, Y. Dong & J. Qiu, 2016. Fifteen new earthworm mitogenomes shed new light on phylogeny within the Pheretima complex. Scientific Reports 6: 20096.

Zhong, M., B. Hansen, M. P. Nesnidal, A. Golombek, K. M. Halanych & T. H. Struck, 2011. Detecting the symplesiomorphy trap: A multigene phylogenetic analysis for terebelliform annelids. Bmc Evolutionary Biology 11: 369.

Zhong, M., T. H. Struck & K. M. Halanych, 2008. Phylogenetic information from three mitochondrial genomes of Terebelliformia (Annelida) worms and duplication of the methionine tRNA. Gene 416: 11-21.

5.9. The Evolution of annelids Reveals two adaptive routes to the interstitial realm

Torsten Hugo Struck[1&2], Golombek, Anja[1&3], Weigert, Anne[4&5], Franke, Franziska Anni[2&4], Westheide, Wilfried[6], Purschke, Günter[3], Bleidorn, Christoph[4], Halanych, Kenneth Michael[2]

[1] Centre of Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig Bonn 53113; Germany
[2] Department of Biological Sciences; Auburn University; Auburn, AL 36849; USA
[3] FB05 Biology/Chemistry; University of Osnabrück; Osnabrück 49069; Germany
[4] Institute of Biology; University of Leipzig; Leipzig 04103; Germany
[5] Max Planck Institute for Evolutionary Anthropology; Leipzig 04103; Germany
[6] University of Osnabrück; Wallenhorst 49134; Germany

Corresponding author:
Torsten H. Struck; Zoological Research Museum Alexander Koenig; Adenauerallee 160; 53113 Bonn; Germany; torsten.struck.zfmk@uni-bonn.de; Phone: +49-228-9122401; Fax: +49-228-9122212

Many animals permanently inhabit the marine interstitium, the space between sand grains [1, 2]. Different evolutionary scenarios may explain the existence of interstitial animals [3, 4]. These scenarios include that (1) this environment is the ancestral habitat of bilaterians [5, 6] or interstitial taxa evolved from larger ancestors (2) by miniaturization or (3) progenesis [3]. The first view mirrors the former hypothesis that interstitial annelids, called archiannelids, were at the base of the annelid radiation [7]. Based on morphological data, however, progenesis is generally favoured for interstitial annelids today [3, 4, 8]. Herein our phylogenomic approach revealed that interstitial archiannelids are robustly placed into two groups nested within the annelid tree. Evolution of the first group comprising among others Dinophilidae is, as expected, best explained by progenesis. In contrast, the second group comprising Protodrilida and Polygordiidae appears to have evolved by stepwise miniaturization adapting from coarser to finer sediments. Hence, besides progenesis [3, 4] miniaturization, which was thought to be too slow for an adaptation to the interstitium [3], is an important second route towards adaptation to interstitial environments. Both progenesis and miniaturization should be considered when investigating the evolution of interstitial taxa from other animal groups [1, 3].

Results and Discussion

Interstitial environments are inhabited by numerous metazoan taxa (e.g. platyhelminths, nematodes, kinorhynchs, copepodes, gastrotrichs) including many annelid taxa, which among others comprised also the so-called archiannelids (Protodrilida, Polygordiidae, Dinophilidae, Diurodrilidae, Nerillidae, and *Apharyngtus*) [8, 9]. Hypotheses regarding evolution of interstitial species can be categorized into three scenarios [3]. First, inhabiting the interstitium was assumed to be ancestral for Bilateria; a life style for example shown by gastrotrichs and gnathostomulids [5, 6]. Likewise, the “Archiannelida” concept proposed that the interstitial annelid taxa mentioned exhibit the ancestral condition of Annelida [7, 10]. However, recent phylogenomic studies [11] placed the non-interstitial lineages Oweniidae and Magelonidae at the base of the annelid tree, but these analyses did not include any archiannelid taxon. The second scenario suggests that progenesis was the evolutionary process for adapting to the interstitium [3]. That is, larval or juvenile stages of a larger ancestor temporarily inhabiting the interstitium arrested the somatic development and became sexually mature. Thus, they inhabited the interstitium permanently. The third, often neglected, scenario suggested that the interstitium was colonized by miniaturization via gradual, step-by-step decrease in body size

from a much larger adult ancestor, which had an infaunal or epibenthic life history [3]. Whereas miniaturized species should resemble adult ancestors, progenetic species should resemble larval or juvenile stages of their ancestors [3].

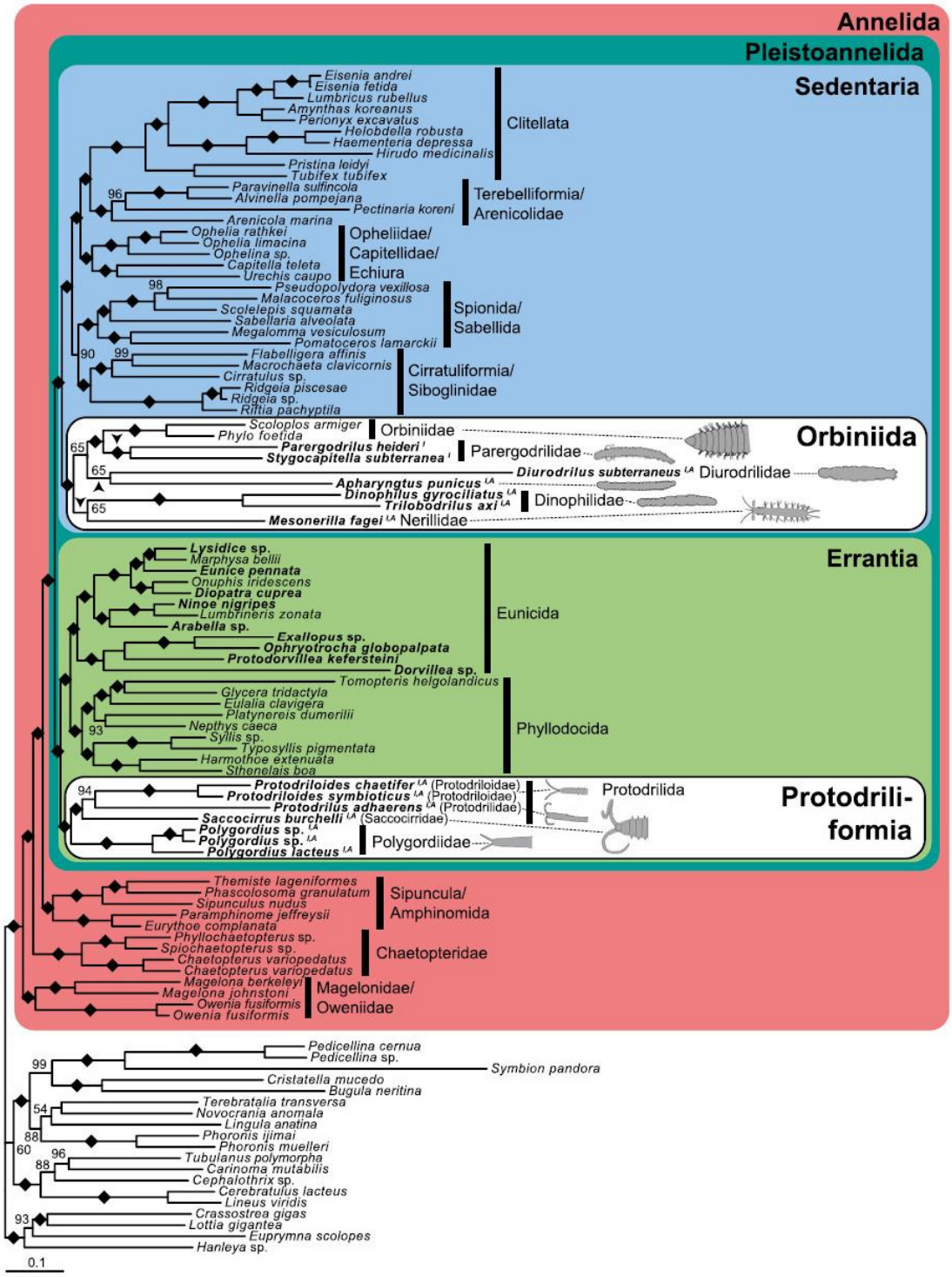


Fig. 1: Tree of Maximum Likelihood (ML) analysis of the largest dataset d01 with 100 species, 189,193 amino acid positions and 41.2% sequence coverage using RAxML [30]. Bootstrap values above 50 are shown at the branches with values of 100% depicted as diamonds. Higher taxonomic units are indicated and species for which new data have been generated are in bold. Drawings of relevant taxa are displayed, but are not to scale. Superscript I indicates interstitial species and A former archiannelids. Arrow heads indicate possible events of progenesis. For testing of alternative hypotheses see Table S1 and for density plots of different bias measurements Figure S3.

Progenesis is currently assumed for all archiannelid taxa (taxa labelled with an A in Fig. 1) [3, 4, 8, 9, 12-14]. However, the conclusion of progenesis based on morphological data alone entails the risk of circular reasoning [4].
Characters of the adult stage of a taxon are compared with larval or juvenile characters and not with adult ones of another taxon assuming progenesis [15], followed by the subsequent conclusion of progenesis. In contrast, some molecular-phylogenetic studies found archiannelids as part of the basal annelid radiation albeit given weak nodal support [10, 16] pointing to the possibility of an interstitial ancestry of Annelida. Hence, we investigated whether these interstitial annelids still show a putatively ancestral condition of Annelida, or evolved by progenesis or miniaturization.
We applied a phylogenomic approach, generating transcriptome sequence data for 12 interstitial archiannelid species representing Protodrilida, Polygordiidae, Dinophilidae, Diurodrilidae, Nerillidae, and *Apharyngtus*, 2 additional interstitial species (i.e., Parergodrilidae) as well as 9 additional annelid taxa using Illumina HiSeq2000 sequencing technology and a modified RNA amplification method [6]. All sequence data were deposited in NCBI SRA and are available via the BioProject PRJNA282709. These data were complemented with data of 77 annelid and lophotrochozoan outgroup species of a previoud study [11]. Phylogenetic reconstructions based on the largest datasets (d01 in Table 1) resulted in significantly supported phylogenetic relationships within Annelida generally congruent with previous analyses [11, 17] (Figure 1) except for placement of Cirratuliformia/Siboglinidae. In our analyses this clade is sister to Sabellida/Spionida, while in most analyses of Weigert et al. [11] it is sister to Orbiniidae. Although the reconstructed topology of annelids is generally robust among studies, this change illustrates that increased taxon sampling may be necessary to place all annelid taxa securely.
In our analyses, interstitial annelid taxa were placed into two major groups with significant nodal support of 100% for each placement (Figure 1). First, Protodrilida and Polygordiidae were placed together as sister to the other Errantia. In contrast to morphology84 based hypotheses, neither Protodrilida nor Polygordiidae were closely related to any sedentarian taxon, specifically to either Spionida and Opheliidae, respectively, assuming progenesis [4]. Hypothesis testing significantly rejected these morphology-based hypotheses (Table S1). Second, Dinophilidae, Diurodrilidae, *Apharyngtus,* and Nerillidae were placed in a clade together with Orbiniidae and Parergodrilidae within Sedentaria. For these four archiannelid taxa a close relationship to Errantia or its subtaxon Eunicida had been proposed based on morphological data; again assuming progenesis [4, 10]. For example, Dinophilidae and Diurodrilidae were considered progenetic Eunicida due to their resemblance to the eunicidan polytroch larvae [3, 4, 8, 9, 12-14]. Such a relationship is not supported by our data and was significantly rejected by hypothesis testing (Table S1). Previous analyses using nuclear rRNA genes also challenged a close relationship of these taxa to Eunicida, but lacked strong support for alternative placements [13, 14, 18].
We also thoroughly checked if tree reconstruction artifacts affected the topology and especially placement of interstitial taxa [6, 19, 20]. Therefore, influence of erroneously assigned paralogous sequences, cross-contamination, branch-length heterogeneity, overall evolutionary rate, amino acid composition, compositional heterogeneity, and the degree of shared missing data was assessed. First, we showed that paralogous sequences or cross101 contamination [6, 19] were not influencing placement of interstitial taxa as exclusion of potentially affected sequences (Table S2) did not alter results (d02 in Table 1). Second, interstitial species were not clustered as in Figure 1 when hierarchical clustering based on metric values of potential biases was applied (e.g., degree of missing data per species) (Figures S1 & S2). Third, we generated more conservative but smaller datasets by excluding biased gene partitions or species (Figure S3, Table 1). If interstitial species were among the excluded (i.e., biased) species, they were re-included to determine the effect of the corresponding bias on the topological position.

Table 1: Characteristics of the datasets generated based on different criteria and bootstrap support for the three hypotheses Orbiniida, Protodriliformia and Protodriliformia placed within Errantia. Exclusion of partitions and species was based on density plots. Partitions or species, which were part of the skewed right tails, were excluded. If interstitial species were among the excluded (i.e., biased) species each excluded interstitial species was re-included. The number of species (= Spe.) and positions (= Pos.), the percentage of sequence coverage (= Cov.) as well as the d value of MARE [32], the sensitivity criterion and the corresponding threshold (= Thres.) values determined using TreSpEx [33] and BaCoCa [34], respectively, are provided for each dataset. Moreover, it has been indicated when species were re-included (= Species reincl.). Bootstrap values at the significance level of 95 or higher are in bold and of 70 and higher are in italics. n/a = not applicable. For more detailed results of the cross-contamination and paralogy screening see Table S2 and for the position of Diurodrilidae Table S3.

Data set #	MARE d	Sensitivity criterion	Thres.	Species reincl.	Pos. #	Spe. #	Cov. %	Orbiniida	Protodriliformia	Protodriliformia within Errantia
1	1.0	none	n/a	n/a	189193	100	41.23	**100**	**100**	**100**
2	1.0	contamination & paralogy	n/a	n/a	189193	100	41.21	**100**	**100**	**100**
3	1.5	Missing data per partition	n/a	n/a	116990	100	44.87	*94*	**100**	**100**
4	2.0	Missing data per partition	n/a	n/a	80669	100	47.33	*76*	*77*	*75*
5	2.0	w/o *Mesonerilla*	n/a	n/a	80669	99	47.32	**98**	**100**	**100**
6	2.0	w/o *Protodriloides* & *Protodrilus*	n/a	n/a	80669	97	47.10	**99**	**100**	*93*
7	1.0	Branch length Heterogeneity	72.78	n/a	143036	100	42.95	**100**	**100**	**100**
8	1.0	Evolutinary rate	1.335	n/a	169392	100	41.88	**100**	**100**	**100**
9	1.0	LB score	50.89	n/a	189193	98	41.77	**100**	**100**	**100**
10	1.0	LB score	29.24	n/a	189193	93	42.34	**99**	**100**	**99**
11	1.0	LB score	16.89	n/a	189193	86	42.00	**100**	**100**	**100**
12	1.0	LB score	16.89	*Apharyngtus punicus*	189193	87	42.04	**100**	**100**	**100**
13	1.0	LB score	16.89	*Dinophilus gyrociliatus*	189193	87	42.62	**100**	**100**	**100**
14	1.0	LB score	16.89	*Diurodrilus subterraneus*	189193	87	41.67	**100**	**100**	**100**
15	1.0	LB score	16.89	*Trilobodrilus axi*	189193	87	42.46	**100**	**100**	**100**
16	1.0	compositial heterogeneity	0.00121	n/a	189193	97	42.44	**100**	**100**	**100**
17	1.0	compositial heterogeneity	0.00087	n/a	189193	92	44.50	**100**	**100**	**100**
18	1.0	compositial heterogeneity	0.00064	n/a	189193	80	49.24	**100**	**100**	**100**
19	1.0	compositial heterogeneity	0.00064	*Protodriloides chaetifer*	189193	81	48.69	**100**	**100**	**100**
20	1.0	compositial heterogeneity	0.00064	*Trilobodrilus axi*	189193	81	49.64	**100**	**100**	**100**
21	1.0	Missing data per species	0.566	n/a	189193	44	75.60	**100**	**100**	**100**
22	1.0	Missing data per species	0.566	*Diurodrilus subterraneus*	189193	45	74.23	**100**	**100**	**100**
23	1.0	Missing data per species	0.566	*Mesonerilla fagei*	189193	45	74.88	**100**	**100**	**100**
24	1.0	Missing data per species	0.566	*Protodriloides chaetifer*	189193	45	74.03	**100**	**100**	**100**

Phylogenetic reconstructions based on these datasets recovered the same results with respect to the placement of interstitial annelid taxa as shown in Figure 1 with generally significant support of 100% (Figure 2, d03-24 in Table 1). Hence, in contrast to previous studies [3, 4, 8, 9, 12-14, 18], our analyses robustly placed interstitial annelid taxa.

Relationships within the two clades including interstitial taxa were also stable. Monophyly of both, Polygordiidae and Protodrilida, is recovered, usually with significant bootstrap support ≥

95%. Within the other clade, Parergodrilidae is always sister to Orbiniidae with maximal support. *Apharyngtus* is sister to Diurodrilidae, and together they are sister to Orbiniidae/Parergodrilidae in all analyses. Dinophilidae and Nerillidae is always sister to these four taxa (Figures 1 & 2). However, nodal support for the position of Diurodrilidae is unstable; a situation possibly due to biased base composition drawing Diurodrilidae towards Dinophilidae reducing bootstrap support at associated nodes (Table S3). Excluding *Diurodrilus* increased bootstrap support to 100% for both clades Dinophilidae/Nerillidae and *Apharnygtus*/Orbiniidae/Parergodrilidae.

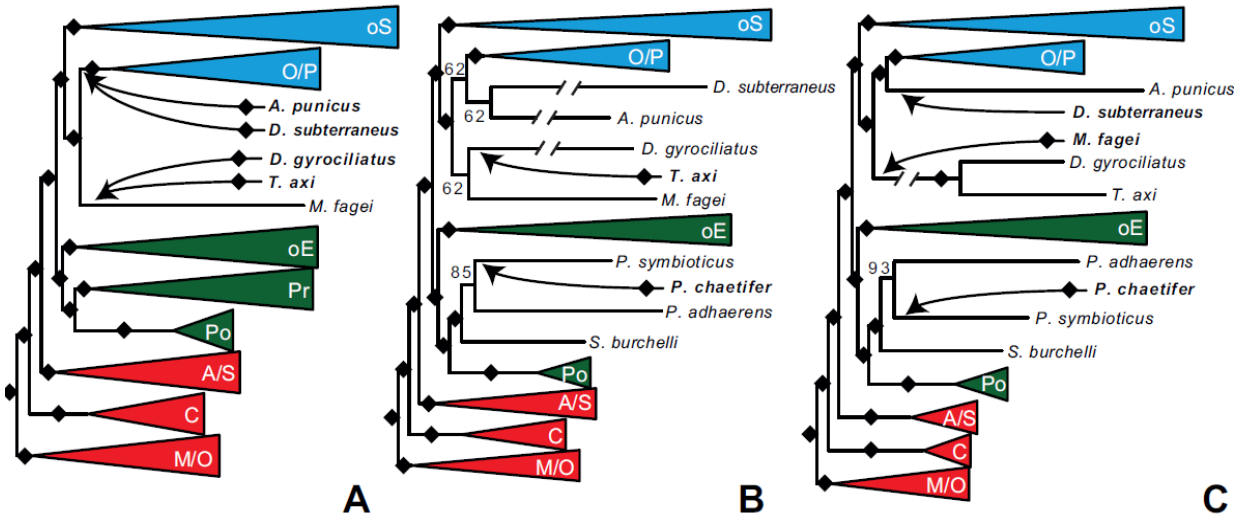


Fig. 2: Trees of ML analyses from which biased species were excluded. Exclusion of species was based on density plots and species, which were part of the skewed right tails of normal distributions of theses biases, were excluded (Figure S3). If interstitial species were among the excluded (i.e., biased) species each excluded interstitial species was re-included to determine the effect of the bias on its position. That is, if this species was placed differently in analyses when all other biased species were excluded. (A) Long-branched species using the LB score (d11, 86 species), (B) compositional heterogeneity using RCFV values (d18, 80) and (C) degree of missing data (d21, 44). Positions of re-included species plus bootstrap support for this position indicated by arrows. Higher taxonomic units except for the interstitial annelids are collapsed and outgroups not shown. Bootstrap values >50 shown, with values of 100% as diamonds. A/S = Amphinomidae/Sipuncula; C = Chaetopteridae; M/O = Magelonidae/Oweniidae; oE = other Errantia; O/P = Orbiniidae/Parergodrilidae; oS = other Sedentaria; Po = Polygordiidae; Pr = Protodrilida. For the results of the clustering analyses see Figures S1 and S2.

These results have interesting implications for annelid evolution as interstitial taxa are not part of the basal radiation of lineages. Hence, inhabiting the interstitial realm and possessing a simple body organization is not an ancestral trait of Annelida. The "Archiannelida" concept has to be rejected, thereby confirming previous studies [3, 10].

Nerillidae, Dinophilidae, Diurodrilidae, and *Apharyngtus* are closely related to Orbiniidae and Parergodrilidae in our analyses with strong nodal support. Therefore, we name this new clade Orbiniida as the first described species of this clade was an orbiniid (Figure 1). Ancestral state (AS) reconstructions of Orbiniida (Table S4, Figure S4A & B) differed only in two characters from the reconstruction of the last common ancestor (LCA) of Sedentaria, which is remarkably similar to reconstructions of the large LCA of Annelida [11]. Prostomial palps were absent and the head encompasses two or more rings behind the prostomium. However, both characters have to be critically reviewed as reversions to the original state have occurred. In Nerillidae prostomial palps are present and in Orbiniidae the prostomium is followed by one ring as adults [8, 9]. Although several small-sized orbiniid species with two rings in the adult stage are known, recent analyses [21] supported the hypothesis that these species evolved independently by progenesis within Orbiniidae [21, 22]. Therefore, the adult stage of the large LCA of Orbiniida

resembled the LCA of Sedentaria by possessing palps and only one ring in addition to parapodia and chaetae.
Parergodrilidae, Dinophilidae, Diurodrilidae, and *Apharyngtus* are of small size and lack palps, parapodia and chaetae (except for Parergodrilidae). Hence, these taxa do not esemble the adult stage LCA of Orbiniida. Together with their small size, Dinophilidae, Diurodrilidae and *Apharyngtus* possess ciliary rings and bands resembling polytroch larvae of Orbiniidae, which temporarily inhabit the interstitial realm [18, 22]. Moreover, heads of Parergodrilidae, Dinophilidae and *Apharyngtus* incorporate the prostomium and more than one ring as is also observed in larval and juvenile stages of Orbiniidae. Thus, given that several larval or juvenile characters of orbiniids persist in these four taxa, evolution by progenesis seems likely (Figure 3A). As the state reconstruction revealed a large-size LCA for Orbiniida similar to the LCA of Sedentaria, independent progenesis leading to Dinophilidae, Diurodrilidae/*Apharyngtus* and Parergodrilidae (arrows in Figure 1) is a more likely explanation than a single progenesis event at the base of Orbiniida followed by subsequent reversion to large size. A conclusion further substantiated by the fact that several independent progenesis events are known within Orbiniidae [21]. Dinophilidae and Diurodrilidae/*Apharyngtus* likely evolved by progenesis from an earlier developmental stage than Parergodrilidae (Figure 3A) as the latter bears chaetae. Whereas resemblance to polytroch larvae and size support the hypothesized progenesis of these taxa, except for their very small size (< 1mm) in the range of larval stages [3] strong evidence is lacking for Nerillidae (Figure 3A). Unlike orbiniids, nerillids possess three antennae at the prostomium [9, 23].

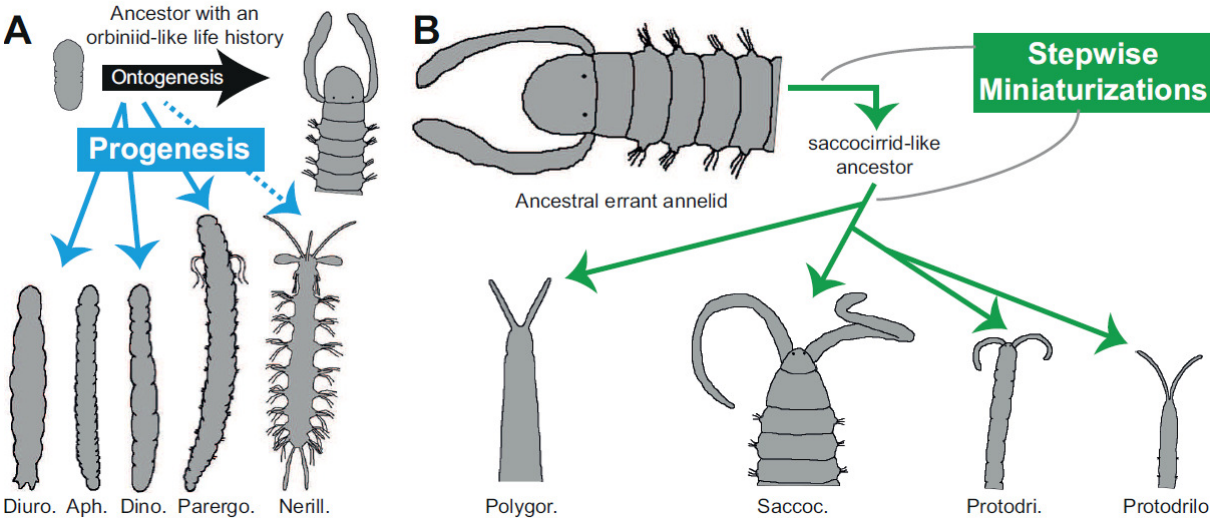


Fig. 3: Two different scenarios explain the evolution of interstitial annelids. (A) progenesis and (B) stepwise miniaturization. Drawings are not to scale. The dashed line indicates weak evidence only due to size. Aph. = Apharyngtus; Dino. = Dinophilidae; Diuro. = Diurodrilidae; Nerill. = Nerillidae; Parergo. = Parergodrilidae; Polygor. = Polygordiidae; Protodri. = Protodrilidae; Protodrilo. = Protodriloidae; Saccoc. = Saccocirridae. For the results of ancestral state reconstruction using maximum likelihood mapping and a modified morphological data matrix of Weigert et al. [11] in Mesquite [31] see Table S4 and Figure S4.

In contrast, Polygordiidae and Protodrilida do not resemble developmental stages of related taxa, and progenesis seems unlikely. Several species of Polygordiidae and Protodrilida are large being several centimetres long [9, 24]. Instead, we propose that miniaturization explains their evolution (Figure 3B), thereby contradicting the previous hypothesis of progenesis [4]. Similar to the annelid LCA [11, 17, 25], the errant LCA was reconstructed to be a larger epibenthic or infaunal annelid consisting of a prostomium with palps and eyes followed by one ring, homonymous segments bearing parapodia with simple chaetae and a pygidium with cirri (Table

S4). Saccocirridae (Protodrilida) possess a prostomium with palps and eyes followed by one ring and homonymous segments with parapodia bearing simple chaetae similar to the reconstructed ancestor. Polygordiidae species, like saccocirrids, are large in the range of centimetres inhabiting coarse sediments like gravel [26] but show only internal signs of segmentation. Protodriloidae (Protodrilida) are considerably smaller with a body length of about one centimetre without obvious external segmentation except for chaetae in *Protodriloides chaetifer* [9]. Similarly for Protodrilidae (Protodrilida), a body length of about one centimetre as well as complete reduction of parapodia and chaetae are regarded as ancestral traits [24]. During the course of evolution within Protodrilidae, species inhabiting finer sediments decrease further in size and reduce characters like eyes [9, 24]. Whereas interstitial annelids usually are too small allowing production of a sufficient number of oocytes for development via a planktonic larva [27], Saccocirridae, Polygordiidae and some Protodrilidae are capable to produce enough oocytes necessary for such a mode of development and only in smaller species direct development occurs [9, 27].

Reconstructions of the LCA of Polygordiidae and Protodrilida were similar to recent Saccocirridae except for the absence of parapodia and presence of pygidial cirri (Table S4). However, parapodia would have to be regained in Saccocirridae. As absence of parapodia is not well supported (proportional likelihood *p* of 0.56; Suppl. Table 9, Figure S4C) regaining a complex character like parapodia seems less plausible than independent losses in Polygordiidae and Protodrilidae/Protodriloidae. Therefore, the LCA was most likely saccocirrid-like evolving from an infaunal ancestor that inhabited coarse sediments. In the lineage that lead to Polygordiidae, character traits were further reduced [26], while within Protodrilida additional size-reduction occurred allowing the inhabitation of finer sediments [8, 9, 24] (Figure 3B). As the LCA resembles the protodrilidan Saccocirridae, we name this clade Protodriliformia (Figure 1).

The most prominent characteristic of interstitial environments is the small space, which organisms inhabit, enforcing a rigid requirement for small body size [3, 4]. The probability of a successful permanent invasion is supposedly much higher starting from a stage already adapted to the interstitium (e.g., juveniles) as in progenesis than from a larger stage adapted to an infaunal or epifaunal life via miniaturization [3]. Therefore, progenesis has generally been favored over miniaturization to explain the evolution of interstitial annelids [3, 4]. Besides cases of progenetic evolution in Orbiniida, several interstitial species, which independently originated by progenetic evolution, can also be found in Eunicida and possibly Hesionidae [3, 4, 12]. Stepwise miniaturization occurs within Protodriliformia [contra 4] and Pisionidae [3]. In the light of our results, the interstitial species of Syllidae [28] also more likely evolved by miniaturization as they are morphologically very similar to adult stages of non-interstitial syllids. Hence, miniaturization might be another evolutionary trajectory, whose importance is similar to the one of progenesis. Thus, there are two different evolutionary routes to adapt to the interstitium from larger ancestors. Explanations of how interstitial lineages evolve from within a group of taxa that possess larger body size typically focus on progenetic mechanisms [3] whereas the process of miniaturization is often neglected.

In view of our findings on annelids, miniaturization should be taken into account more often when investigating evolution of interstitial taxa from larger ancestors within other animal groups as well. One other explanation proposed for some taxa (e.g., Platyhelminthes, Gastrotricha and Gnathifera) is the idea that they descended for a common bilaterian ancestor that had an interstitial life style [5, 6], but some have argued that the interstitial nature of such taxa was secondarily derived [29]. The debate about taxa like Platyhelminthes, Gastrotricha and Gnathifera raises the question, how likely are secondary reductions in body size (to the point of being interstitial) over the course of evolutionary history. From a parsimonious perspective, recent phylogenomic analyses were most congruent with the former hypothesis of retaining the ancestral bilaterian life style [5, 6]. For support of the latter hypothesis of secondary reductions

additional evidence like traces of larval, juvenile or adult character traits suggesting that they evolved from a larger ancestor is needed. To date such unequivocal evidence is lacking for these taxa (i.e., Platyhelminthes, Gastrotricha and Gnathifera). Comparatively in the case of annelids, our phylogenomic analyses provided a robust evolutionary framework allowing alternative hypotheses of body size evolution to be tested. When combined with knowledge of larval, juvenile or adult character traits, both progenesis and miniaturization emerge as important evolutionary processes in Annelida.

Author Contributions

T.H.S. conceived this study. A.G. and K.H. took the lead on data collection. G.P. and W.W. aided in the data collections. A.G., A.W., F.A.F, and K.H. generated the RNAseq libraries. T.H.S. took the lead on data analyses. A.W. and C.B. aided in the data analyses. T.H.S., G.P., C.B. and K.H. mainly contributed to writing the manuscript.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (grants DFG-STR 683/5-2 and DFG-STR 683/8-1 to T.H.S., DFG-BL 787/5-1 to C.B.) as well as by the USA National Science Foundation (DEB-1036537). We also thank Birgit Nickel and Matthias Meyer (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany) for their assistance in sequencing using Illumina Hi-Seq 2000. This is contribution #XXX to AU Marine Biology Program and # XX to The Molette Biological Laboratory.

References

1. Giere, O. (2009). Meiobenthology - The microscopic motile fauna of aquatic sediments, (Berlin Heidelberg: Springer-Verlag).
2. Noodt, W. (1974). Anpassungen an interstielle Bedingungen: Ein Faktor in der Evolution höherer Taxa der Crustacea. Faun.-ökol. Mitt. *4*, 445-452.
3. Westheide, W. (1987). Progenesis as a principle in meiofauna evolution. J. Nat. Hist. *21*, 843-854.
4. Struck, T.H. (2006). Progenetic species in polychaetes (Annelida) and problems assessing their phylogenetic affiliation. Integr. Comp. Biol. *46*, 558–568.
5. Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., et al. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. R. Soc. B: Biol. Sci. *276*, 4261-4270.
6. Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., et al. (2014). Platyzoan paraphyly based on phylogenomic data supports a non-coelomate ancestry of Spiralia. Mol. Biol. Evol. *31*, 1833-1849.
7. Hermans, C.O. (1969). The systematic position of the Archiannelida. Syst. Zool. *18*, 85-102.
8. Worsaae, K., and Kristensen, R.M. (2005). Evolution of interstitial Polychaeta (Annelida). Hydrobiologia *535-536*, 319-340.
9. Westheide, W. (2008). Polychaetes: Interstitial families., Second Edition Edition, (Shrewsbury: Field Studies Council).
10. Struck, T.H. (2012). Phylogeny of Annelida. In Zoology Online. (Berlin, Boston: DeGruyter), p. 23 pp. 12
11. Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C., et al. (2014). Illuminating the base of the annelid tree using transcriptomics. Mol. Biol. Evol. *31*, 1391-1401.
12. Eibye-Jacobsen, D., and Kristensen, R.M. (1994). A new genus and species of Dorvilleidae (Annelida, Polychaeta) from Bermuda, with a phylogenetic analysis of Dorvilleidae, Iphitimidae and Dinophilidae. Zool. Scr. *23*, 107-131.
13. Worsaae, K., and Rouse, G.W. (2008). Is *Diurodrilus* an annelid? J. Morphol. *269*, 1426-1455.
14. Golombek, A., Tobergte, S., Nesnidal, M.P., Purschke, G., and Struck, T.H. (2013). Mitochondrial genomes to the rescue – Diurodrilidae in the myzostomid trap. Mol. Phylogenet. Evol. *68*, 312–326.
15. Wiens, J.J., Bonett, R.M., and Chippindale, P.T. (2005). Ontogeny Discombobulates Phylogeny: Paedomorphosis and Higher-Level Salamander Relationships. Syst. Biol. *54*, 91 - 110.

16. Struck, T.H., Nesnidal, M.P., Purschke, G., and Halanych, K.M. (2008). Detecting possibly saturated positions in 18S and 28S sequences and their influence on phylogenetic reconstruction of Annelida (Lophotrochozoa). Mol. Phylogenet. Evol. *48*, 628-645.
17. Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., et al. (2011). Phylogenomic analyses unravel annelid evolution. Nature *471*, 95–98.
18. Struck, T.H., Halanych, K.M., and Purschke, G. (2005). Dinophilidae (Annelida) is most likely not a progenetic Eunicida: Evidence from 18S and 28S rDNA. Mol. Phylogenet. Evol. *37*, 619–623.
19. Struck, T.H. (2013). The impact of paralogy on phylogenomic studies – A case study on annelid relationships. PLoS ONE *8*, e62892.
20. Simmons, M.P. (2012). Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics *28*, 208-222.
21. Bleidorn, C., Hill, N., Ersèus, C., and Tiedemann, R. (2009). On the role of character loss in orbiniid phylogeny (Annelida): Molecules vs. morphology. Mol. Phylogenet. Evol. *52*, 57-69.
22. Rouse, G.W., and Pleijel, F. (2001). Polychaetes, (Oxford: University Press).
23. Worsaae, K. (2005). Phylogeny of Nerillidae (Polychaeta, Annelida) as inferred from combined 18S rDNA and morphological data. Cladistics *21*, 143-162.
24. Martínez, A., Di Domenico, M., Rouse, G.W., and Worsaae, K. (2015). Phylogeny and systematics of Protodrilidae (Annelida) inferred with total evidence analyses. Cladistics *31*, 250–276.
25. Struck, T.H. (2011). Direction of evolution within Annelida and the definition of Pleistoannelida. J. Zool. Syst. Evol. Res. *49*, 340-345.
26. Ramey-Balci, P.A., Fiege, D., and Purschke, G. (2012). Polygordiida: Polygordiidae Czerniavsky, 1881. In Zoology Online. (Berlin, Boston: DeGruyter).
27. Westheide, W. (1984). The concept of reproduction in polychaetes with small body size: adaptations in interstitial species. In Polychaete Reproduction., Volume 29, H.-D. Pfannenstiel and A. Fischer, eds. (Stuttgart: Fortschritte der Zoologie), pp. 265-287.
28. San Martin, G., and Aguado, M.T. (2012). Family Syllidae. In Zoology Online. (Berlin, Boston: DeGruyter).
29. Brinkman, H., and Philippe, H. (2008). Animal phylogeny and large-scale sequencing: progress and pitfalls. J. Syst. Evol. *46*, 274–286.
30. Stamatakis, A. (2006). RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.
31. Maddison, W.P., and Maddison, D.R. (2009). Mesquite: a modular system for evolutionary analysis., 2.71 Edition.
32. Meusemann, K., von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P., Ebersberger, I., Walzl, M., Pass, G., Breuers, S., et al. (2010). A phylogenomic approach to resolve the arthropod Tree of Life. Mol. Biol. Evol. *27*, 2451-2464.
33. Struck, T.H. (2014). TreSpEx – Detection of misleading signal in phylogenetic reconstructions based on tree information. Evolutionary Bioinformatics *10*, 51-67.
34. Kück, P., and Struck, T.H. (2014). BaCoCa - A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. *70*, 94–98.

6. References

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.* USA 97, 4453 – 4456.

Aguado, M.T., San Martin, G., Siddall, M.E. (2012). Systematics and evolution of syllids (Annelida, Syllidae). *Cladistics* 28, 234 – 250.

Aguado, M.T., Glasby, C.J., Schroeder, P.C., Weigert, A., Bleidorn, C. (2015a). The making of a branching annelid: an analysis of complete mitochondrial genome and ribosomal data of *Ramisyllis multicaudata*. *Sci. Rep.* 5, 12072.

Aguado, M.T., Helm, C., Weidhase, M., Bleidorn, C. (2015b). Description of a new syllid species as a model for evolutionary research of reproduction and regeneration in annelids. *Org. Divers. Evol.* 15, 1 – 21.

Aguado, M.T., Richter, S., Sontowski, R., Golombek, A., Struck, T.H., Bleidorn, C. (2016). Syllidae mitochondrial gene order is unusually variable for Annelida. *Gene* 594, 89 – 96.

Ahlrichs, W.H. (1997). Epidermal ultrastructure of *Seison nebaliae* and *Seison annulatus*, ans a comparison of epidermal structures within Gnathifera. *Zoomorpholgy* 117, 41 – 48.

Altenhoff, A.M., Dessimoz, C. (2012). Inferring orthology and paralogy. *Methods Mol. Biol.* 855, 259 – 279.

Avise, J.C. (2000). Phylogeography – The History and Formation of Species. Harvard Univerity Press, Cambridge, Mass.

Ax, P. (1967). *Diurodrilus ankeli nov*. spec. (Archiannelida) von der nordamerikanischen Pazifikküste. Ein Beitrag zur Morphologie, Systematik und Verbreitung der Gattung Diurodrilus. *Z. Morphol. Oeko Tiere* 60. 5 - 16.

Ax, P. (1985). The position of the Gnathostomulida and Platyhelminthes in the phylogenetic system of Bilateria. In: Morris, S.C., George, J.D., Gibson, R., Platt, H.M. (eds.). *The origins and relationships of lower invertebrates*, pp. 168 – 180. The Systematics Association, Cambridge University Press, Cambridge.

Barnes, R.D. (1968). Invertebrate Zoology. W.B. Saunders Co., Philadelphia, PA.

Bartolomaeus, T. (1998). Chaetogenesis in polychaetous Annelida – significance for annelid systematics and the position of the Pogonophora. *Zoology* 100, 348 – 364.

Bartolomaeus, T., Purschke, G., Hausen, H. (2005). Polychaete phylogeny based on morphological data – a comparison of current attempts. *Hydrobiologia* 535–536, 341 – 356.

Bensasson, D., Zhang, D., Hartl, D.L., Hewitt, G.M. (2001) Mitochondrial pseudogenes: evolution's

misplaced witnesses. *Trends Ecol. Evol.* 16, 314 – 321.

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

Bernt, M., Braband, A., Schierwater, B., Stadler, P.F. (2013a). Genetic aspects of mitochondrial genome evolution. *Mol. Phylogenet.Evol.* 69, 328 – 338.

Bernt, M., Bleidorn, C., Braband, A., Dambach, J., Donath, A., Fritzsch, G., Golombek, A., Hadrys, H., Jühling, F., Meusemann, K., Middendorf, M., Misof, B., Perseke, M., Podsiadlowski, L., von Reumont, B., Schierwater, B., Schlegel, M., Schrödl, M., Simon, S., Stadler, P.F., Stöger, I., Struck, T.H. (2013b). A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Mol. Phylogenet. Evol.* 69, 352 – 364.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Putz, J., Middendorf, M., Stadler, P.F. (2013c). MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313 – 319.

Bleidorn, C. (2007). The role of character loss in phylogenetic reconstruction as exemplified for the Annelida. *J. Zool. Syst. Evol. Res.* 45, 299 – 307.

Bleidorn, C. (2008). Lophotrochozoan relationships and parasites. A snap-shot. *Parasite* 15, 329 – 332.

Bleidorn, C., Podsiadlowski, L., Bartolomaeus, T. (2006). The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (Annelida, Orbiniidae) – a novel gene order for Annelida and implications for annelid phylogeny. *Gene* 370, 96 – 103.

Bleidorn C., Eeckhaut, I., Podsiadlowski, L., Schult, N., McHugh, D., Halanych, K.M., Milinkovitch, M.C., Tiedemann, R. (2007). Mitochondrial genome and nuclear sequence data support Myzostomida as part of the annelid radiation. *Mol. Biol. Evol.* 24, 1690 – 1701.

Bleidorn, C., Lanterbecq, D., Eeckhaut, I., Tiedemann, R. (2009a). A PCR survey of Hox genes in the myzostomid *Myzostoma cirriferum*. *Dev. Genes Evol.* 219, 211 – 216.

Bleidorn, C., Podsiadlowski, L., Zhong, M., Eeckhaut, I., Hartmann, S., Halanych, K.M., Tiedemann, R. (2009b). On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evol. Biol.* 9, 150.

Bleidorn, C., Hill, N., Erséus, C., Tiedemann, R. (2009c). On the role of character loss in orbiniid phylogeny (Annelida): molecules vs. morphology. *Mol. Phylogenet. Evol.* 52, 57 – 69.

Boore, J.L. (1999). Animal mitochondrial genomes. *Nucl. Acids Res.* 27, 1767 – 1780.

Boore, J.L. (2006). The complete sequence of the mitochondrial genome of *Nautilus macromphalus* (Mollusca; Cephalopoda). *BMC Genom*. 7, 182.

Boore, J.L., Brown, W.M. (1994). Mitochondrial genomes and the phylogeny of mollusks. *Nautilus* 108, 61 – 78.

Boore, J.L.; Brown, W.M. (1998). Big trees from little genomes: mitochondrial gene order as a

phylogenetic tool. *Curr. Opin. Genet. Dev.* 8, 668 – 674.

Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L., Brown, W.M. (1995). Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376, 163 – 165.

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

Bourlat, S.J., Rota-Stabelli, O., Lanfear, R., Telford, M.J. (2009). The mitochondrial genome structure of *Xenoturbella bocki* (phylum Xenoturbellida) is ancestral within the deuterostomes. *BMC Evol. Biol.* 9, 107.

Bridge, D., Cunningham, C.W., Schierwater, B., DeSalle, R., Buss, L.W. (1992). Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. *Proc. Natl. Acad. Sci.* USA 89, 8750 – 8753.

Brinkman, H., Philippe, H. (2008). Animal phylogeny and large-scale sequencing: progress and pitfalls. *J. Syst. Evol.* 46, 274 – 286.

Brusca, R.C., Brusca, G.J. (2003). Invertebrates, second ed. Sinauer Associates Inc., Sunderland, Massachusetts.

Cameron, C.B., Garey, J.R., Swalla, B.J. (2000). Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proc. Natl. Acad. Sci.* USA 97, 4469 – 4474.

Cannon, J.T., Vellutini, B.C., Smith, J. 3rd, Ronquist, F., Jondelius, U., Hejnol, A. (2016) Xenacoelomorpha is the sister group to Nephrozoa. Nature 530 (7588), 89 – 93.

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. Biol. Rev. 73, 203 – 266.

Clark, R.B. (1969). Systematics and phylogeny: Annelida, Echiura, Sipuncula. In: Florkin M, Scheer BT, eds. Chemical Zoology, Vol.4. New York: Academic Press, 1 – 68.

Clary, D.O., Wolstenholme, D.R. (1984). The *Drosophila* mitochondrial genome. *Oxf. Surv. Eukaryot. Genes* 1, 1 – 35.

Colgan, D.J., Hutchings, P.A., Brown, S. (2001). Phylogenetic relationships within the Terebellomorpha. *J. Mar. Biol. Assoc. U.K.* 81, 765 – 773.

Couso, J.P. (2009). Segmentation, metamerism and the Cambrian explosion. *Int. J. Dev. Biol.* 53, 1305 – 1316.

Dales, R.P. (1963). Annelids. London: Hutchinson University Library.

Dales, R.P. (1977). The polychaete stomatodeum and phylogeny. In: Reish D.J., Fauchald K., eds. Essasy on Polychaetous Annelids in Memory of Dr. Olga Hartman. Los Angeles: Allan Hancock Foundation, University of Southern California, 525 – 546.

Danovaro, R., Dell'anno, A., Pusceddu, A., Gambi, C., Heiner, I., Kristensen, R.M. (2010). The first Metazoa living in permanently anoxic conditions. *BMC Biol.* 8, 30.

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

de Quatrefages, A.M. (1866). *Histoire Naturelle des Annelides, Marine et d'Eau Douce. Annelides et Gephyriens* Vol. 1 (Librairie Encyclopédique de Roret).

De Robertis, E.M. (2008). The molecular ancestry of sementation mechanisms. *Proc. Natl. Acad. Sci. U S A* 105, 16411 – 16412.

de Rosa R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte A., et al. (1999). HOX genes in brachiopods and priapulids and protostome evolution. *Nature* 399, 772 – 776.

Dordel, J., Fisse, F., Purschke, G., Struck, T.H. (2010). Phylogenetic position of Sipuncula derived from multi-gene and phylogenomic data and its implication for the evolution of segmentation. *J. Zool. Evol. Res*. 48, 197 – 207.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., *et al*. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745 – 750.

Ebersberger, I., Strauss, S., von Haeseler, A. (2009). HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol. Biol.* 9, 157.

Edgecombe, G., Giribet, G., Dunn, C., Hejnol, A., Kristensen, R., Neves, R., Rouse, G., Worsaae, K., SØrensen, M. (2011). Higher-level metazoan relationships: recent progress and remaining questions. *Org. Divers. Evol.* 11, 151 – 172.

Eeckhaut, I., McHugh, D., Mardulyn, P., Tiedemann, R., Monteyne, D., Jangoux, M., Milinkovitch, M.C. (2000). Myzostomida: a link between trochozoans and flatworm? *Proc. R. Soc. Lond. B* 267, 1383 – 1392.

Eeckhaut, I, Lanterbecq, D. (2005). Myzostomida: a review of the phylogeny and ultrastructure. *Hydrobiologia* 535-356, 253 – 275.

Eernisse, D.J., Peterson, K.J. (2004). The history of animals. In *Assembling the tree of life* (eds J. Cracraft & M.J. Donoghue), Oxford, UK: Oxford University Press.

Ehlers, U., Sopott-Ehlers, B. (1997). Ultrastructure of the subepidermal musculature of *Xenoturbella bocki*, the adelphotaxon of the Bilateria. *Zoomorphology* 117, 71 – 79.

Eibye-Jacobsen, D., Kristensen, R.M. (1994). A new genus and species of Dorvilleidae (Annelida, Polychaeta) from Bermuda, with a phylogenetic analysis of Dorvilleidae, Iphitimidae and Diophilidae. *Zool. Scr*. 23, 107 – 131.

Erséus, C. (2005). Phylogeny of oligochaetous Clitellata. *Hydrobiologia* 535-536, 357 - 372.

Fauchald, K. (1974). Polychaete phylogeny: a problem in protostome evolution. *Syst. Zool*. 23, 493 – 506.

Fauchald, K. (1977). The polychaete worms: definitions and keys to the orders, families and genera. Natural History Museum of Los Angeles County, *Science* Series 28, 1 - 190.

Fauchald, K., Rouse, G.W. (1997). Polychaete systematics: Past and present. *Zool. Scr.* 26, 71 - 138.

Fauvel, P. (1923). Polychètes errantes. *Faune de France* 5, 1 - 488.

Fauvel, P. (1927). Polychètes sédentaires. *Faune de France* 16, 1 - 494.

Galtier, N., Nabholz, B., Glemin, S., Hurst, G.D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18, 4541 – 4550.

Garstang, W. (1928). The morphology of the Tunicata and its bearing on the phylogeny of the Chordata. Q. J. Microsc. Sci. 72, 51 – 187.

Gerlach, S.A. (1952). *Diurodrilus benazzii*, ein neuer Archiannelide aus dem Küstengrundwasser des Mittelmeeres. *Zool. Anz.* 149, 185 – 188.

Gerlach, S.A. (1953). Zur Kenntnis der Archianneliden des Mittelmeeres. *Kieler Meeresforsch* 9, 248 - 251.

Giere, O. (2009). Meiobenthology – The Microscopic Motile Fauna of Aquatic Sediments (Berlin: Springer-Verlag).

Giribet, G. (2008). Assembling the lophotrochozoan (= spiralian) tree of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1513 – 1522.

Giribet, G., Distel, D.L., Polz, M., Sterrer, W., Wheeler, W.C. (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.

Glasby, C.J., Hutchings, P.A., Hall, K. (2004). Assessment of monophyly and taxon affinities within the polychaete clade Terebelliformia (Terebellida). *J. Mar. Biol. Assoc. U.K.* 84, 961 – 971.

Golombek, A., Tobergte, S., Nesnidal, M.P., Purschke, G., Struck, T.H. (2013). Mitochondrial genomes to the rescue – Diurodrilidae in the myzostomid trap. *Mol. Phylogenet. Evol.* 68, 312 – 326.

Golombek, A., Tobergte, S., Struck, T.H. (2015). Elucidating the phylogenetic position of Gnathostomulida and first mitochondrial genomes of Gnathostomulida, Gastrotricha and Polycladida (Pltyhelminthes). Mol. Phylogenet. Evol. 86, 49 – 63.

Gould, S.J. (1977). Ontogeny and Phylogeny. Cambridge, MA: The Belknap Press of Harvard University Press.

Grube, A.E. (1850). Die Familien der Anneliden. *Archive für Naturgeschichte*, Berlin, 16, 249 – 364.

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

Halanych, K.M. (2004). The new view of animal phylogeny. *Annu. Rev. Ecol. Syst.* 35, 229 – 256.

Hall, K.A., Hutchings, P.A., Colgan, D.J. (2004). Further phylogenetic studies of the Polychaeta using 18S rDNA sequence data. *J. Mar. Biol. Ass. U.K.* 84, 949 – 960.

Hankeln, T., Wey-Fabrizius, A.R., Herlyn, H., Witek, A., Weber, M., Nesnidal, M.P., Struck, T.H. (2014). Phylogeny of platyzoan taxa based on molecular data. In: Wägele, J.W., Bartolomaeus, T. (Eds.), Deep Metazoan Phylogeny: The backbone of the Tree of Life – New Insights from Analyses of Molecules, Morphology, and Theory of Data Analysis. De Gruyter, Berlin, pp. 105 – 125.

Hassanin, A., Leger, N., Deutsch, J. (2005). Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* 54, 277 – 298.

Haszprunar, G. (1996). Plathelminthes and Plathelminthomorpha – paraphyletic taxa. *J. Zool. Syst. Evol. Res.* 34, 41 – 48.

Hausdorf, B., Helmkampf, M., Meyer, A., Witek, A., Herlyn, H., Bruchhaus, I., Hankeln, T., Struck, T.H., Lieb, B. (2007). Spiralian phylogenomics supports the resurrection of Bryozoa comprising Ectoprocta and Entoprocta. *Mol. Biol. Evol.* 24, 2723 – 2729.

Hausdorf, B., Helmkampf, M., Nesnidal, M.P., Bruchhaus, I. (2010). Phylogenetic relationships within the lophiphorate lineages (Ectoprocta, Brachiopoda, and Phoronida). *Mol. Phylogenet. Evol.* 55, 1121 – 1127.

Hausen, H. (2005). Chaetae and chaetogenesis in polychaetes (Annelida). *Hydrobiologia* 535-536, 37 – 52.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. Biol. Sci*. 276, 4261 – 4270.

Helm, C., Bernhart, S.H., Siederdissen, C.H.Z.., Nickel, B., Bleidorn, C. (2012). Deep sequencing of small RNAs confirms an annelid affinity of Myzostomida. *Mol. Phylogenet. Evol.* 64, 198 – 203.

Helmkampf, M, Bruchhaus, I., Hausdorf, B. (2008). Phylogenomic analyses of lophophorates (brachiopods, phoronids and bryozoans) confirm the Lophotrochozoa concept. *Proc. R. Soc. Lond. B. Biol. Sci.* 275, 1927 – 1933.

Herlyn, H., Ehler, U. (1997). Ultrastructure and function of the pharynx of *Gnathostomula paradoxa* (Gnathostomulida). *Zoomorphology* 117, 135.

Hermans, C.O. (1969). The systematic position of the Archiannelida. Syst. Biol. 18, 85 – 102.

Hessling, R. (2002). Metameric organization of the nervous system on developmental stages of *Urechis caupo* (Echiura) and its phylogenetic implications. *Zoomorphology* 121, 221 – 234.

Hessling, R. (2003). Novel aspects of the nervous system of *Bonellia viridis* (Echiura) revealed by the combination of immunohistochemistry, confocal laser-scanning microscopy and three dimensional reconstruction. *Hydrobiologia* 496, 225 – 239.

Hessling, R., Westheide, W. (2002). Are Echiura derived from a segmented ancestor? Immunohistochemical analysis of the nervous system in developmental stages of *Bonellia viridis*. *J. Morphol.* 252, 100 – 113.

Hyman, L.H. (1951). The invertebrates. Vol. 2. Platyhelminthes and Rhynchocoela: the Acoelomate bilaterian. New York: McGraw-Hill.

Isrealsson, O. (1997). Xenoturbella's molluscan relatives and molluscan embryogenesis. *Nature* 390, 32.

Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H. (2006). Phylogenomics: The beginning of incongruence? *Trends Genet.* 22, 225 – 231.

Jekely, G., Colombelli, J., Hausen, H., Guy, K., Stelzer, F., Nedelec, F., Arendt, D. (2008). Mechanism of phototaxis in marine zooplankton. *Nature* 456, 359 – 399.

Jenner, R.A. (2004a). When molecules and morphology clash: reconciling conflicting phylogenies of the Metazoa by considering secondary character loss. *Evol. Dev.* 6, 372 – 378.

Jenner, R.A. (2004b). Towards a phylogeny of the Metazoa: evaluating alternative phylogenetic positions of Platyhelmithes, Nemertea, and Gnathostomulida, with a critical reappraisal of cladistic characters. *Cont. Zool.* 73, 3 – 163.

Jennings, R.M., Halanych, K.M. (2005). Mitochondrial genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): evidence for conserved gene order in Annelida. *Mol. Biol. Evol.* 22, 210 – 222.

Kocot, K.M. (2016). On 20 years of Lophotrochozoa. Org. Divers. Evol. 16, 329 – 343.

Kristensen, R.M., Nørrevang, A. (1977). On the fine structure of *Rastrognathia macrostoma* gen. et sp. n. placed in Rastrognathiidae fam. n. (Gnathostomulida). *Zool. Scr.* 6, 27 – 41.

Kristensen, R.M., Niilonen, T. (1982). Structural studies on *Diurodrilus* Remane (Diurodrilidae fam.n.), with description of *Diurodrilus westheidei* sp.n. from the arctic interstitial meiobenthos, W. Greenland. *Zool. Scr*. 11, 1 – 12.

Kristensen, R.M., Funch, P. (2000). Micrognathozoa: A new class with complicated jaws like those of Rotifera and Gnathostomulidae. *J. Morphol.* 246, 1 – 49.

Kristof, A., Wollesen, T., Wanninger, A. (2008). Segmental mode of neural patterning in Sipuncula. *Curr. Biol.* 18, 1129 – 1132.

Kvist, S., Siddall, M.E. (2013). Phylogenomics and Annelida revisited: a cladistic approach using genome wide expressed sequence tag data mining and examining the effects of missing data. *Cladistics* 29, 435 – 448.

Lanterbecq, D., Bleidorn, C., Michel, S., Eeckhaut, I. (2008). Locomotion and fine structure of parapodia in *Myzostoma cirriferum* (Myzostomida). *Zoomorphology* 127, 59 – 68.

Lavrov, D.V., Pett, W., Voigt, O., Wörheide, G., Forget, L., Lang, B.F., Kayal, E. (2013). Mitochondrial DNA of *Clathrina clathrus* (Calcarea, Calcinea): six linear chromosomes, fragmented rRNAs, tRNA editing, and a novel genetic code. *Mol. Biol. Evol.* 30, 865 – 880.

Li, S., Chen, Y., Zhang, M., Bao, X., Li, Y., Teng, W., Liu, Z., Fu, C., Wang, Q., Liu, W. (2016). Complete mitochondrial genome of the marine polychaete, *Marphysa sanguinea* (Polychaeta, Eunicida). *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27, 42 – 43.

Licher, F. (1999). Revision of der Gattung *Typosyllis* Langerhans, 1879 (Polychaeta: Syllidae). Morphologie, Taxonomie und Phylogenie. Abh. Senckenb. Naturforsch. Ges. 551, 1 – 336.

Luo, Y.-J., Satoh, N., Endo, K. (2015). Mitochondrial gene order variation in the brachiopod *Lingula anatina* and its implications for mitochondrial evolution in lophotrochozoans. *Mar. Gnomics* 24 (Part 1), 31 – 40.

Mallatt, J., Winchell, C.J. (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, J., inchell, C.J. (2007). Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star. *Mol. Phylogenet. Evol.* 43, 1005 – 1022.

Marletaz, F., Martin, E., Perez, Y., Papillon, D., Caubit, X., Lowe, C.J., Freeman, B., Fasano, L., Dossat, C., Wincker, P. (2006). Chaetognath phylogenomics: a protostome with deuterostome-like development. *Curr. Biol.* 16, R577 – R578.

Matus, D.Q., Copley, R.R., Dunn, C.W., Hejnol, A., Eccleston, H., Halanych, K.M., Martindale, M.Q., Telford, M.J. (2006). Broad taxon and gene sampling indicate that chaetognaths are protostomes. *Curr. Biol.* 16, R575 – R576.

McHugh, D. (1997). Molecular evidence that echiurans and pogonophorans are derived annelids. *Proc. Natl. Acad. Sci. U. S. A.* 94, 8006 – 8009.

McHugh, D. (2000). Molecular phylogeny of the Annelida. *Can. J. Zool.* 78, 1873 - 1884.

McHugh, D. (2005). Molecular systematics of polychaetes (Annelida). *Hydrobiologia* 535-536, 309 - 318.

Metschnikoff, V.E. (1881). Über die systematische Stellung von *Balanoglossus. Zool. Anz.* 4, 139 – 157.

Müller, M.C.M., Westheide, W. (2000). Structure of the nervous system of *Myzostoma cirriferum*

(Annelida) as revealed by immunohistochemistry and cLSM analyses. *J. Morphol*. 245, 87 – 98.

Müller, M.C.M., Westheide, W. (2002). Comparative analysis of the nervous systems in presumptive progenetic dinophilid and dorvilleid polychaetes (Annelida) by immunohistochemistry and cLSM. *Acta Zool.* 83, 33 – 48.

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

Mwinyi, A., Bailly, X., Bourlat, S.J., Jondelius, U., Littlewood, D.T.J., Podsiadlowski, L. (2010). The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of *Symsagittifera roscoffensis*. *BMC Evol.* Biol. 10.

Nesnidal, M.P., Helmkampf, M., Bruchhaus, I., Hausdorf, B. (2010). Compositional heterogeneity and phylogenomic inference of metazoan relationships. *Mol. Biol. Evol.* 27, 2095 – 2104.

Nesnidal, M.P., Helmkampf, M., Meyer, A., Witek, A., Bruchhaus, I., Ebersberger, I., Hankeln, T., Lieb, B., Struck, T.H., Hausdorf, B. (2013). New phylogenomic data support the monophyly of Lophophorata and an Ectoproct-Phoronid clade and indicate that Polyzoa and Kryptrochozoa are caused by systematic bias. *BMC Evolutionary Biology* 13, 253.

Nielsen, C. (2010). After all: *Xenoturbella* is an acoelomorph! *Evol. Dev.* 12, 241 – 243.

Nielsen, C. (2012). Animal Evolution – Interrelationships of the Living Phyla. Oxford University Press Inc., New York.

Noodt, W. (1974). Anpassungen an interstitielle Bedingungen. Ein Faktor in der Evolution höherer Taxa der Crustacea. *Faun. – ökol. Mitt.* 4, 445 – 452.

Nosenke, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonada, M., Müller, W.E., et al., (2013). Deep metazoan phylogeny: when different genes tell different stories. *Mol. Phylogenet. Evol.* 67, 223 – 233.

Nygren, A. (1999). Phylogeny and reproduction in Syllidae (Polychaeta). *Zool. J. Linnean Soc.* 126, 365 – 386.

Nygren, A. (2004). Revision of Autolytinae (Syllidae, Polychaeta). *Zootaxa* 680, 1 – 314.

Nygren, A., Sundberg, P. (2003). Phylogeny and evolution of reproductive modes in Autolytinae (Syllidae, Annelida). *Mol. Phylogenet. Evol.* 29, 235 – 249.

Osca, D., Irisarri, I., Todt, C., Grande, C., Zardoya, R. (2014). The complete mitochondrial genome of *Scutopus ventrolineatus* (Mollusca: Chaetodermomorpha) supports the Aculifera hypothesis. BC Evol. Biol. 14, 197.

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.

Paps, J., Bagunà, J., Riutort, M. (2009a). Bilaterian phylogeny: a broad sampling of 13 nuclear genes provides a new Lophotrochozoa phylogeny and supports a paraphyletic basal Acoelomorpha. *Mol. Biol. Evol.* 26, 2397 – 2406.

Paps, J., Bagunà, J., Riutort, M. (2009b). Lophotrochozoan internal phylogeny: new insights from an up to-date analysis of nuclear ribosomal genes. *Proc. Roy. Soc. B: Biol. Sci.* 276, 1245 – 1254.

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

Paxton, H. (2000). Family Diurodrilidae. In: Beesly, P.L., Ross, G.J., Glasby, C.J. (Eds.), Polychaetes & Allies: The Southern Synthesis. Fauna of Australia, Polychaeta, Myzostomida, Pogonophora, Echiura, Sipuncula, vol. 4A. CSIRO Publishing, Melbourne, pp. 104- 105.

Perna, N.T., Kocher, T.D. (1995). Patterns of nucleotide composition at fourfold degenerate sitesof animal mitochondrial genomes. *J. Mol. Evol.* 41, 353 – 358.

Perseke, M., Hankeln, T., Weich, B., Fritzsch, G, Stadler, P., Israelsson, O., Bernhard, D., Schlegel, M. (2007). The mitochondrial DNA of *Xenoturbella bocki*: genomic architecture and phylogenetic anayse. Theory Biosci. 126, 35 – 42.

Perseke, M., Bernhard, D., Fritzsch, G., Brümmer, F., Stadler, P.F., Schlegel, M. (2010). Mitochondrial genome evolution in Ophiuroidea, Echinoidea, and Holothuroidea: insights in phylogenetic elationship of Echinodermata. *Mol. Phylogenet. Evol.* 56, 201 – 211.

Perseke, M., Hetmank, J., Bernt, M., Stadler, P.F., Bernhard, D. (2011). The enigmatic mitochondrial genome of *Rhabdopleura compacta* (Pterobranchia) reveals insights into selection of an efficient tRNA system and supports monophyly of Ambulacraria. BMC Evol. Biol. 9, 90 – 99.

Perseke, M., Golombek, A., Schlegel, M., Struck, T.H. (2013). The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. *Mol. Phylogenet. Evol.* 66, 898 – 905.

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

Philippe, H., Lartillot, N., Brinkmann, H. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* 22, 1246 – 1253.

Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., et al., (2009). Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* 19, 706 – 712.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T., Manuel, M., Wörheide, G., Baurain, D.

(2011a). Resolving difficult phylogenetic questions. why more sequences are not enough. *PLoS Biol.* 9, e1000602.

Philippe, H., Brinkmann, H., Colpey, R.R., Moroz, L.L., Nakano, H., Poustka, A.J., Wallberg, A., Peterson, K.J., Telford, M.J. (2011b). Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. *Nature* 4701, 255 – 258.

Pick, K.S., Philippe, H., Schreiber, F., Erpenbeck, D., Jackson, D.J., Wrede, P., Wiens, M., Alie, A., et al., (2010). Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol. Biol. Evol.* 27, 1983 – 1987.

Purschke, G. (1985a). Anatomy and ultrastructure of ventral pharyngeal organs and their phylogenetic importance in Polychaeta (Annelida) – I. The pharynx of the Dinophilidae. *Zoomorphology* 105, 223 – 239.

Purschke, G. (1985b). Anatomy and ultrastructure of ventral pharyngeal organs and their phylogenetic importance in Polychaeta (Annelida) – I. The pharynx of the Nerillidae. *Mikrofauna* 2, 23 – 60.

Purschke, G. (1999). Terrestrial polychaetes – models for the evolution of the Clitellata (Annelida)? *Hydrobiologia* 406, 87 – 99.

Purschke, G., Jouin, C. (1988). Anatomy and ultrastructure of the ventral pharyngeal organs of *Saccocirrus* (Saccocirridae) and *Protodriloides* (Protodriloidae fam. n.) with remarks on the phylogenetic relationships with Protodrilida (Annelida: Polychaetes). *J. Zool.* 215, 405 – 432.

Purschke, G., Hessling, R., Westheide, W. (2000). The phylogenetic position of the Clitellata and the Echiura – on the problematic assessment of absent characters. *J. Zool. Syst. Evol. Res.* 38, 165 – 173.

Raible, F., Tessmar-Raible, K., Osoegawa, K., Wincker, P., Jubin, C., Balavoine, G., Ferrier, D., Benes, V., de Jong, P., Weissenbach, J., Bork, P., Arendt, D. (2005). Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*. *Science* 310, 1325 – 1326.

Rasmussen, M., Kellis, M. (2007). Accurate gene-tree reconstruction by learning gene- and species specific substitution rates across multiple complete genomes. *Genome Res.* 17, 1932 – 1942.

Reisinger, E. (1960). Was ist *Xenoturbella*? *Z. Wiss. Zool.* 164, 188 – 198.

Remane, A. (1925). Diagnosen neuer Archianneliden (zugleich 3. Beitrag zur Fauna der Kieler Bucht). *Zool. Anz.* 65, 15 – 17.

Riser, N.W. (1984). General observations on the intertidal interstitial fauna of New Zealand. Tane 30, 239 – 250.

Rivera, A., Weisblat, D.A. (2009). And Lophotrochozoa makes three: *Notch/Hes* signaling in annelid segmentation. *Dev. Genes Evol.* 219, 37 – 43.

Rouse, G.W. (1999). Trochophore concepts: ciliary bands and the evolution of larvae in spiralian

Metazoa. *Biol. J. Linn. Soc.* 66, 411 – 464.

Rouse, G.W. (2000). Bias? What bias? The evolution of downstream larval-feeding in animals. *Zool. Scr*. 29, 213 – 236.

Rouse, G.W., Fauchald, K. (1997). Cladistics and polychaetes. *Zool. Scr.* 26, 139 – 204.

Rouse, G.W., Pleijel, F. (2001). Polychaetes. Oxford University Press.

Rubinoff, D., Holland, B.S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst. Biol.* 54, 952 – 961.

Schaeffer, B. (1987). Deuterostome monophyly and phylogeny. Evol. Biol. 21, 179 – 235.

Seaver, E.C. (2003). Segmentation: mono- or polyphyletic? *Int. J. Dev. Biol.* 47, 583 – 595.

Shadel, G.S., Clayton, D.A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* 66, 409 - 435.

Singh, T.R., Tsagkogeorga, G., Delsuc, F., Blanquart, S., Shenkar, N., Loya, Y., Douzery, E.J.P., Huchon, D. (2009). Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *BMC Genomics* 10, 534.

Sørensen, M.V. (2003). Further structures in the jaw apparatus of *Limnognathia maerski* (Micrognathozoa), with notes on the phylogeny of the Gnathifera. *J. Morphol*. 2255, 131 – 145.

Sørensen, M.V., Tyler, S., Hoge, M.D., Funch, P. (2003). Organization of pharyngeal hard parts and musculature in *Gnathostomula armata* (Gnathostomulida: Gnathostomulidae). Can. J. Zool. 81, 1147 – 1463.

Stach, T. (2008). Chordate phylogeny and evolution: a not so simple three-taxon problem. J. Zool. 276, 117 – 141.

Stach, T., Braband, A., Podsaidlowski, L. (2010). Erosion of phylogenetic signal in tunicate mitochondrial genomes on different levels of analysis. *Mol. Phylogenet. Evol.* 55, 860 – 870.

Steinauer, M.L., Nickol, B.B., Broughton, R., Orti, G. (2005). First sequenced mitochondrial genome from the phylum Acanthocephala (Leptorhynchoides thecatus) and its phylogenetic position within Metazoa. *J. Mol. Evol*. 60, 706 – 715.

Sterrer, W., Mainitz, M., Rieger, R.M. (1985). Gnathostomulida: enigmatic as ever. In: Morris, S.C., George, J.D., Gibson, R., Platt, H.M. (eds.). *The origins and relationships of lower invertebrates*, pp. 181 – 191. The Systematics Association, Cambridge University Press, Cambridge.

Storch, V. (1968). Zur vergleichendne Anatomie der segmentalen Muskelsysteme und zur Verwandtschaft der Polychaeten-Familien. *Z. Morph. Tiere* 63, 251 – 342.

Struck, T.H. (2006). Progenetic species in polychaetes (Annelida) and problems assessing their

phylogenetic affiliation. *Integr. Comp. Biol.* 46, 558 – 568.

Struck, T.H. (2011). Direction of evolution within Annelida and the definition of Pleistoannelida. *J. Zool. Syst. Evol. Res.* 49, 340 – 345.

Struck, T.H. (2012). Phylogeny of Annelida. In Handbook of Zoology Online, G. Purschke, M. Böggemann, and W. Westheide, eds. (Berlin: DeGruyter).

Struck, T.H. (2013). The impact of paralogy on phylogenomic studies – a case study on annelid relationships. *PLoS One* 8, e62892.

Struck, T.H., Westheide, W., Purschke, G. (2002). Progenesis in Eunicida ("Polychaeta"; Annelida) – seperate evolutionary events? Evidence from molecular data. *Mol. Phylogenet. Evol.* 25, 190 – 199.

Struck, T.H., Purschke, G. (2005). The sister group relationship of Aeolosomatidae and Potamodrilidae (Annelida: "Polychaeta") – a molecular phylogenetic approach based on 18S rDNA and Cytochrome Oxidase I. *Zool. Anz.* 243, 281 – 293.

Struck, T.H., Fisse, F. (2008). Phylogenetic position of Nemertea derived from phylogenomic data. *Mol. Biol. Evol.* 25, 728 – 736.

Struck, T.H., Halanych, K.M., Purschke, G. (2005). Dinophilidae (Annelida) is most likely not a progenetic Eunicida: evidence from 18S and 28S rDNA. *Mol. Phylogenet. Evol.* 37, 619 – 623.

Struck, T.H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., Halanych, K.M. (2007). Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evol. Biol.* 7, 57.

Struck, T.H., Nesnidal, M.P., Purschke, G., Halanych, K.M. (2008). Detecting possibly saturated positions in 18S and 28S sequences and their influence on phylogenetic reconstruction of Annelida (Lophotrochozoa). *Mol. Phylogenet. Evol.* 48, 628 - 645.

Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., *et al.* (2011). Phylogenomic analyses unravel annelid evolution. *Nature* 471, 195 – 198.

Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kück, P., Herlyn, H., Hankeln, T. (2014a). Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of Spiralia. *Mol. Biol. Evol.* 31, 1833 – 1849.

Struck, T.H., Purschke, G., Dordel, J., Hösel, H., Nesnidal, M.P., Diersing, F., Bleidorn, C., Paul, C., Hill, N., Tiedemann, R., Selbig, J., Hartmann, S. (2014b). Phylogeny and evolution of Annelida based on molecular data. In: Wägele, J.W., Bartholomaeus, T. (Eds.), Deep Metazoan Phylogeny: The backbone of the Tree of Life – New Insights from Analyses of Molecules, Morphology, and Theory of Data Analysis. De Gruyter Berlin, 143 – 160.

Struck, T.H., Golombek, A., Weigert, A., Franke, F.A., Westheide, W., Purschke, G., Bleidorn, C., Halanych, K.M. (2015). The evolution of annelids reveals two adaptive routest o the interstitial realm. *Curr. Biol.* 25, 1993 – 1999.

Swalla, B.J. (2006). Building divergent body plans with similar genetic pathways. *Heredity* 97, 235 – 243.

Swalla, B.J., Smith, A.B. (2008). Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. *Philos. Trans. Roy. Soc. Lond. B. Biol. Sci.* 363, 1557 – 1568.

Telford, M.J., Wise, M.J., Gowri-Shankar, V. (2005). Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from bilaterian. *Mol. Biol. Evol.* 22, 1129 – 1136.

Tessmar-Raible, K., Arendt, D. (2003). Emerging systems: between vertebrates and arthropods, the Lophotrochozoa. *Curr. Opin. Genet. Dev.* 13, 331 – 340.

Tomer, R., Denes, A.S., Tessmar-Raible, K., Arendt, D. (2010). Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell* 142, 800 – 809.

Turbeville, J.M. (1986). An ultrastructural analysis of coelomogenesis in the hoplonemertine *Prosorhochmus americanus* and the polychaete *Magelona* sp. *J. Morphol.* 187, 51 – 60.

Tzetlin, A., Purschke, G. (2005). Pharynx and intestine. *Hydrobiologia* 535 – 536, 199 – 225.

Vallès, Y., Boore, J. (2006). Lophotrochozoan mitochondrial genomes. *Integr. Comp. Biol.*, 46, 544 – 57.

Wägele, J.W., Bartolomaeus, T. (Eds.) (2014). Deep Metazoan Phylogeny: The backbone of the Tree of Life – New Insights from Analyses of Molecules, Morphology, and Theory of Data Analysis. De Gruyter, Berlin.

Weigert, A., Bleidorn, C. (2016) Current status of annelid phylogeny. Organisms, Diversity and Evolution 16, 345-362.

Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C., Struck, T.H. (2014). Illuminating the base of the annelid tree using transcriptomics. *Mol. Biol. Evol.* 31, 1391 – 1401.

Weigert, A., Golombek, A., Gerth, M., Schwarz, F., Struck, T.H., Bleidorn, C. (2016). Evolution of mitochondrial gene order in Annelida. *Mol. Phylogenet. Evol.* 94, 196 – 206.

Westheide, W. (1985). The systematic position of the Dinophilidae and the archiannelid problem. In: Conway Morris S, George JS, Gibson R, Platt HM, editors. The origin and relationships of Lower Invertebrates, Vol. 397. Oxford: Clarendon Press. pp 310 – 326.

Westheide, W. (1987). Progenesis as a principle in meiofauna evolution. *J. Nat. Hist.* 21, 843 – 854.

Westheide, W. (1990). Polychaetes: Interstitial Families. Oegstgeest: Universal Book Services. pp 1 –

152.

Westheide, W. (1997). The direction of evolution within the Polychaeta. *J. Nat. Hist.* 31, 1 – 15.

Westheide, W. (2008). Polychaetes: Interstitial Families, Second Edition (Field Studies Council).

Westheide, W., McHugh, D., Purschke, G., Rouse, G. (1999). Systematization of the Annelida: Different approaches. *Hydrobiologia* 402, 291 – 307.

Wey-Fabrizius, A.R., Podsiadlowski, L., Herlyn, H., Hankeln, T. (2013). Platyzoan mitochondrial genomes. *Mol. Phylogenet. Evol.* 69, 365 – 375.

Wey-Fabrizius, A.R., Herlyn, H., Rieger, B., Rosenkranz, D., Witek, A., Welch, D.B.M., Ebersberger, I., Hankeln, T. (2014). Transcriptome data reveal syndermatan relationships and suggest the evolution of endoparasitism in Acanthocephala via an epizoic stage. *PLoS ONE* 9, e88618.

Wiens, J., Bonett, R., Chippindale, P. (2005). Ontogeny discombobulates phylogeny: paedomorphosis and higher-level salamander relationships. *Syst. Biol.* 54, 91 – 110.

Winchell, C.J., Sullivan, J., Cameron, C.B., Swalla, B.J., Mallatt, J. (2002). Evaluationg hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol. Biol. Evol.* 19, 762 – 776.

Witek, A., Herlyn, H., Meyer, A., Boell, L., Bucher, G., Hankeln, T. (2008). EST based phylogenomics of Syndermata questions monophyly of Eurotatoria. *BMC Evol. Biol.* 8, 345 – 355.

Witek, A., Herlyn, H., Ebersberger, I., Mark Welch, D.B., Hankeln, T. (2009). Support for the monophyletic origin of Gnathifera from phylogenomics. *Mol. Phylogenet. Evol*. 53, 1037 – 1041.

Wolstenholme, D.R. (1992). Animal mitochondrial DNA: Structure and Evolution. *Int. Rev. Cytol.* 141, 173 – 216.

Worsaae, K. (2005). Phylogeny of Nerillidae (Polychaeta, Annelida) as inferred from combined 18S rDNA and morphological data. *Cladistics* 21, 143 – 162.

Worsaae, K., Kristensen, R.M. (2005). Evolution of interstitial Polychaeta (Annelida). *Hydrobiologia* 535 - 536, 319 - 340.

Worsaae, K., Rouse, G.W. (2008). Is *Diurodrilus* an annelid? *J. Morphol.* 269, 1426 - 1455.

Xu, W., Jameson, D., Tang, B., Higgs, P.G. (2006) The relationship between the rate of molecular evolution and the rate of genome rearrangements in animal mitochondrial genomes. *J. Mol. Evol.* 63, 375 – 392.

Yokobori, S., Iseta, T., Asakawa, S., Sasaki, T., Shimizu, N., Yamagishi, A., Oshima, T., Hirose, E. (2008). Complete nucleotide sequences of mitochondrial genomes of two solitary entoprocts, Loxocorone allax and Loxosomella aloxiata: implications for lophotrochozoan phylogeny. *Mol. Phylogenet. Evol.* 47, 612 – 628.

Yuan, Y., Li, Q., Yu, H., Kong, L. (2012). The complete mitochondrial genomes of six heterodont bivalves (Tellenoidea and Solenoidea): variable gene arrangements and phylogenetic implications. *PLoS One* 7, e32353.

Zhong, M., Struck, T.H., Halanych, K.M. (2008). Phylogenetic information from three mitochondrial genomes of Terebelliformia (Annelida) worms and duplication of the methionine tRNA. *Gene* 416, 11 – 21.

Zhong, M., Hansen, B., Nesnidal, M., Golombek, A., Halanych, K.M., Struck, T.H. (2011). Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. *BMC Evolutionary Biology* 11, 369.

Zimmer, R.L. (1973). Morphological and developmental affinities of the lophophorates. *In:* Larwood, G.P. (Ed.), Living and Fossil Bryozoa. Academic Press, New York, 593 – 599.

Zrzavý, J. (2003). Gastrotricha and metazoan phylogeny. *Zool. Scripta* 32, 61 – 81.

Zrzavý, J., Mihulka, S., Kepka, P., Bezdek, A., Tietz, D. (1998). Phylogeny of the metazoan based on morphological and 18S ribosomal DNA evidence. *Cladistics* 14, 249 – 285.

Zrzavý, J., Hypsa, V., Tietz, D.F. (2001). Myzostomida are not annelids: molecular and morphological support for a clade of animals with anterior sperm flagella. *Cladistics* 17, 170 – 198.

Zrzavý, J., Riha, P., Pialek, L., Janouskovec, J. (2009). Phylogeny of Annelida (Lophotrochozoa): total evidence analysis of morphology and six genes. *BMC Evol. Biol.* 9, 189.

7. Publikationen

Artikel:

Zhong, M., Hansen, B., Nesnidal, M., Golombek, A., Halanych, K.M., Struck, T.H. (2011). Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. *BMC Evolutionary Biology* 11, 369.

Perseke, M., Golombek, A., Schlegel, M., Struck, T.H. (2013). The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. *Mol. Phylogenet. Evol.* 66, 898 – 905.

Golombek, A., Tobergte, S., Nesnidal, M.P., Purschke, G., Struck, T.H. (2013). Mitochondrial genomes to the rescue – Diurodrilidae in the myzostomid trap. *Mol. Phylogenet. Evol.* 68, 312 – 326.

Bernt, M., Bleidorn, C., Braband, A., Dambach, J., Donath, A., Fritzsch, G., Golombek, A., Hadrys, H., Jühling, F., Meusemann, K., Middendorf, M., Misof, B., Perseke, M., Podsiadlowski, L., von Reumont, B., Schierwater, B., Schlegel, M., Schrödl, M., Simon, S., Stadler, P.F., Stöger, I., Struck, T.H. (2013). A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Mol. Phylogenet. Evol.* 69, 352 – 364.

Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kück, P., Herlyn, H., Hankeln, T. (2014). Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of Spiralia. *Mol. Biol. Evol.* 31, 1833 – 1849.

Golombek, A., Tobergte, S., Struck, T.H. (2015). Elucidating the phylogenetic position of Gnathostomulida and first mitochondrial genomes of Gnathostomulida, Gastrotricha and Polycladida (Pltyhelminthes). Mol. Phylogenet. Evol. 86, 49 – 63.

Struck, T.H., Golombek, A., Weigert, A., Franke, F.A., Westheide, W., Purschke, G., Bleidorn, C., Halanych, K.M. (2015). The evolution of annelids reveals two adaptive routest o the interstitial realm. *Curr. Biol.* 25, 1993 – 1999.

Weigert, A., Golombek, A., Gerth, M., Schwarz, F., Struck, T.H., Bleidorn, C. (2016). Evolution of mitochondrial gene order in Annelida. *Mol. Phylogenet. Evol.* 94, 196 – 206.

Aguado, M.T., Richter, S., Sontowski, R., Golombek, A., Struck, T.H., Bleidorn, C. (2016). Syllidae mitochondrial gene order is unusually variable for Annelida. *Gene* 594, 89 – 96.

Beiträge auf Tagungen:

2010: 1st graduate meeting of the study group Zoological Systematics of the DZG, Hamburg, "Systematics in the Age of Genomics", Vortrag: A. Golombek

Phylogenetic position of the supposedly paedomorphic species *Apharyngtus punicus* and Dinophilidae (Annelida) using mitogenomic data.

2010: 103. Jahrestagung der Deutschen Zoologischen Gesellschaft, Hamburg, Poster: A. Golombek, G. Purschke, T.H. Struck

Phylogenetic position of the supposedly paedomorphic species *Apharyngtus punicus* and Dinophilidae (Annelida) using mitogenomic data.

2011: "Deep Metazoan Phylogeny 2011 - new data, new challanges", München, Poster: A. Golombek, G. Purschke, T.H. Struck

Phylogenetic position of the interstitial annelid taxa A*pharyngtus punicus* and Dinophilidae of supposedly paedomorphic origin.

2012: 13. GfBS-Jahrestagung, Bonn, Poster: A. Golombek, G. Purschke, T.H. Struck

The phylogenetic position of the former archiannelidan taxa Protodrilidae and

Polygordiidae within Annelida

2012: Graduiertentreffen der Fachgruppe Zoologische Systematik, Leipzig, "Beyond tree reconstruction – applications for phylogenetic trees", Vortrag: A. Golombek, T.H. Struck

The phylogeny of the former archiannelidan taxa Protodrilidae and Polygordiidae based on molecular data.

2012: 105. Jahrestagung der Deutschen Zoologischen Gesellschaft, Konstanz, Vortrag: A. Golombek, G. Purschke, S. Tobergte, T.H. Struck

Interstitial annelid taxa: addressing their phylogeny with complete mitochondrial genomes using Next Generation Sequencing.

2013: BioSyst.EU 2013 Global Systematics!, Wien, Vortrag: A. Golombek, K.M. Halanych, G. Purschke, T.H. Struck

The phylogeny of interstitial annelid taxa using Next Generation Sequencing: when morphological data is not enough.

2013: XIth International Polychaete Conference, Sydney, Vortrag: A. Golombek, G. Purschke, T.H. Struck

Mitochondrial Genomes to the Rescue? – Phylogeny of interstitial annelid Taxa

8. Danksagung

Mein Dank gilt Herrn apl. Professor Dr. Günther Purschke und Herrn Professor Dr. Torsten Struck (auch wenn er jetzt in Oslo am naturhistorischen Museum arbeitet) für die Möglichkeit in der AG Zoologie/Entwicklungsbiologie an der Universität Osnabrück promovieren zu dürfen und für die Bereitstellung meines Themas. Vor allem danke ich den beiden für Ihre unermüdliche Unterstützung während der gesamten Zeit meiner Promotion, vor allem im letzte Jahr. Der Weg bis hier hin war nicht immer einfach und ich konnte mich vor allem in den schwierigen Phasen immer auf das Verständnis und die Hilfe von beiden verlassen.

Meinen Dank an Herrn Prof. Dr. Struck kann ich eigentlich gar nicht in Worte fassen. Zum einen danke ich ihm für bereitwillige Übernahme das Zweitgutachtens meiner Arbeit. Aber vor allem danke ich ihm für seine geduldige und freundliche Art und das er mich immer wieder ermutigt hat nicht aufzugeben. Ich habe so unheimlich viel von ihm gelernt.

Ebenso danke ich Herrn Prof. Dr. Achim Paululat und Frau Dr. Katharina Psathaki für ihre Bereitschaft meiner Prüfungskommission anzugehören.

Ein ganz besonderer Dank gilt der gesamten Arbeitsgruppe Zoologie/Entwicklungsbiologie der Universität Osnabrück. Ich möchte ich mich herzlich für die tolle Arbeitsatmosphäre bedanken. 2011 bin ich gemeinsam mit Prof. Dr. Torsten Struck nach Bonn an das Forschungsmuseum Alexander Koenig gegangen, und der Abschied aus Osnabrück fiel mir nicht leicht. Aber immer wenn ich wieder zurückgekehrt bin, sei es um einen Fortschrittsbericht über den Verlauf meiner Arbeit abzugeben oder um einige Klonierungsarbeiten im Labor durchzuführen, wurde ich immer wieder herzlich in der Arbeitgruppe aufgenommen.
Auch wenn sie nicht mehr in der Arbeitsgruppe sind, so möchte ich dennoch Stephanie Albrecht, Maik Drechsler, Mareike Panz, Christine Lehmacher und Bettina Abeln für eine tolle Zeit danken. Sie hatten immer ein offenes Ohr und die Zeit im Labor mit ihnen war immer lustig. Danken möchte ich auch Herrn Dr. Heiko Harten der mir immer mit Rat und Tat zur Seite stand.
Ein besonderer Dank geht an Arianne Wilmes. Wir kennen uns seit dem ersten Tag unseres Studiums und wir haben viele Dinge zusammen erlebt.

Dankeschön sagen möchte ich auch der Arbeitsgruppe am Forschungsmuseum Alexander Koenig in Bonn. Ich habe dort zwei Jahre verrbracht und es war eine sehr schöne Zeit. Vor allem möchte ich Frau Claudia Etzbauer danken, die mir immer zugehört hat egal ob es um private oder berufliche Dinge ging, und die mir im Labor bei allen Angelegenheiten immer hilfrich zur Seite stand. Ebenso danken möchte ich Sandra Meid und Alexander Donath durch die meine Zeit in Bonn unvergesslich war.
Auch möchte ich meinen Mitbewohnern aus meiner Zeit in Osnabrück und Bonn danken. Es war ein tolles Zusammenleben und wir haben viele schöne Stunden miteinander verbracht.

Meiner besten Freundin Bine möchte ich einfach für „Alles“ danken. Sie ist immer für mich da (auch wenn wir uns nur sehr selten sehen) und ich bin so froh und stolz sie als Freundin zu haben. Ebenso möchte ich all meinen Freunden danken, die immer für mich da sind und mit denen mich so viel verbindet.

Ein ganz besonderes Dankeschön geht an meine Famile. Ich möchte an dieser Stelle für die bedingungslose Unterstützung und all das entgegengebrachte Verständnis in all den Jahren Danke sagen. Sie sind immer für mich da und haben nie am mir gezweifelt. Ohne sie hätte ich es nicht geschafft! Ich hab Euch sehr lieb!!

Ein ganz besonderer Dank geht an Daniel Pulte. Wir sind diesen langen Weg gemeinsam gegangen (er war von Anfang an dabei). Und auch wenn ich den letzten Abschnitt jetzt alleine gehe, weil sich unsere Wege getrennt haben, so war er immer für mich da, hat mich angespornt und hat mir in den schwierigen Momenten den Raum gegeben um wieder weiter zu machen. Dankeschön für alles.

9. Lebenslauf

Persönliche Daten

Name:	Golombek
Vorname:	Anja
Geburtsdatum:	21.01.1985
Geburtsort:	Lutherstadt Wittenberg

Schulische und akademische Ausbildung

2010 – 2018	Promotion an der Universität Osnabrück in der Abteilung Zoologie/Entwicklungsbiologie Thema: “Inferring the phylogeny of problematic metazoan taxa using mitogenomic and phylogenomic data.”
2004 – 2010	Biologiestudium an der Universität Osnabrück Erwerb des Hochschulgrades: Diplom-Biologin Schwerpunkte: Zoologie, Entwicklungs- und Neurobiologie Thema: "Die phylogenetische Stellung der vermutlich paedomorphen Art *Apharyngtus punicus* basierend auf nuklearen und mitochondrialen Markern."
1995 – 2004	Gymnasium Jessen in Jessen (Elster) Abschluss: Abitur
1991 – 1995	Grundschule Max Lingner in Jessen (Elster)

10. Eidesstattliche Erklärung

Erklärung über die Eigenständigkeit der erbrachten wissenschaftlichen Leistung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Meinen spezifischen Beitrag zu den Publikationen in meiner Dissertation sowie die besonderen, wichtigen Beiträge anderer Autoren habe ich im Folgenden beschrieben:

Zhong, M., Hansen, B., Nesnidal, M., Golombek, A., Halanych, K.M., Struck, T.H. (2011):

1. Generierung, Sequenzierung und Annotation aller nuklearen Daten mit der Unterstützung von B. Hansen & M. Nesnidal
2. M. Zhong: Durchführung der Generierung, Sequenzierung und Annotation aller mitochondrialer Genome
3. alle Autoren haben das Manuskript gelesen und bestätigt

Perseke, M.[1], Golombek, A.[1], Schlegel, M., Struck, T.H. (2013):
([1] geteilte Erstautorenschaft)

1. Konzeption der Arbeit und der durchgeführten Analysen gemeinsam mit M. Perseke
2. Schreiben des Manuskripts unter Mitarbeit von M. Perseke
3. alle Autoren haben das Manuskript gelesen und bestätigt
4. die phylogenetischen Analysen erfolgten im Rahmen eines „Special issue"-Projekts (zugehörig zum Priority Program „Deep Metazoan Phylogeny") und wurden von M. Bernt, A. Donath und den Mitarbeitern der Bioinformatikgruppe von P.F. Stadler durchgeführt und in diese Veröffentlichung eingefügt

Golombek, A.[1], Tobergte, S.[1], Nesnidal, M.P., Purschke, G., Struck, T.H. (2013):
([1] gleichmäßige Beteiligung beider Autoren)

1. Konzeltion der Arbeit
2. Generierung, Sequenzierung und Annotation des mitochondrialen Genoms von *Diurodrilus subterraneus*
3. Zusammenstellung der Datensätze, Konzeption und Durchführung der phylogenetischen Analysen

4. Schreiben des Manuskripts mit der Unterstützung von T.H. Struck
5. S. Tobergte: Generierung und Zusammenstellung der 28S rDNA von *Diurodrilus subterraneus*
6. M.P. Nesnidal: Generierung und Zusammenstellung der rDNA von Arten der Platyzoa
7. alle Autoren haben das Manuskript gelesen und bestätigt

Bernt, M., Bleidorn, C., Braband, A., Dambach, J., Donath, A., Fritzsch, G., Golombek, A., Hadrys, H., Jühling, F., Meusemann, K., Middendorf, M., Misof, B., Perseke, M., Podsiadlowski, L., von Reumont, B., Schierwater, B., Schlegel, M., Schrödl, M., Simon, S., Stadler, P.F., Stöger, I., Struck, T.H. (2013):

1. bei dieser Veröffentlichung handelt es ich um eine Kooperation verschiedener Arbeitsgruppen, in welcher die Phylogeny der Bilateria mittels mitochondrialer Genome untersucht wurde (zugehörig zum Priority Program „Deep Metazoan Phylogeny“)
2. alle Autoren haben das Manuskript gelesen und bestätigt und einen gleichen Anteil an der Konzeption der durchgeführten phylogenetischen und analytischen Analysen und dem konzeptionellen Aufbau dieser Studie
3. die Analysen erfolgten im Rahmen eines „Special issue“-Projekts (Deep Metazoan Phylogeny) und wurden von M. Bernt, A. Donath und den Mitarbeitern der Bioinformatikgruppe von P.F. Stadler durchgeführt und in diese Veröffentlichung eingefügt

Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kück, P., Herlyn, H., Hankeln, T. (2014):

1. Mitarbeit an der Generierung der cDNA libraries
2. alle Autoren haben das Manuskript gelesen und bestätigt

Golombek, A., Tobergte, S., Struck, T.H. (2015):

1. Konzeption der Arbeit und der durchgeführten phylogentischen Analysen
2. Generierung, Sequenzierung und Annotation der mitochondrialen Genome der in dieser Studie analysierten Arten der Platyzoa in Zusammenarbeit mit S. Tobergte
3. Zusammenarbeit beim Schreiben des Manuskripts mit T.H. Struck
4. alle Autoren haben das Manuskript gelesen und bestätigt

Struck, T.H., Golombek, A., Weigert, A., Franke, F.A., Westheide, W., Purschke, G., Bleidorn, C., Halanych, K.M. (2015):

1. hauptverantwortlich für die Datensammlung gemeinsam mit K.M. Halanych
2. Generierung der RNA-seq libraries gemeinsam mit A. Weigert, F.A. Franke & K.M. Halanych
3. alle Autoren haben das Manuskript gelesen und bestätigt

Weigert, A., Golombek, A., Gerth, M., Schwarz, F., Struck, T.H., Bleidorn, C. (2016):

1. Generierung, Sequenzierung und Annotation mitochondrialer Genome gemeinsam mit M. Gerth & F. Schwarz
2. alle Autoren haben das Manuskript gelesen und bestätigt

Aguado, M.T., Richter, S., Sontowski, R., Golombek, A., Struck, T.H., Bleidorn, C. (2016):

1. Mithilfe bei der Generirung, Sequenierung und Annotation mitochondrialen Genome
2. alle Autoren haben das Manuskript gelesen und bestätigt

Alle oben aufgeführten Personen sind als Koautoren auf den entsprechenden Publikationen geführt.

Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

_____ _____

(Ort, Datum) (Anja Golombek)