Molecular Phylogenetics and Evolution 80 (2014) 79–87

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

A transcriptome approach to ecdysozoan phylogeny

Janus Borner [a], Peter Rehm [a], Ralph O. Schill [b], Ingo Ebersberger [c], Thorsten Burmester [a,*]

[a] Institute of Zoology and Zoological Museum, University of Hamburg, D-20146 Hamburg, Germany
[b] Zoology, Biological Institute, University of Stuttgart, Germany
[c] Department for Applied Bioinformatics, University of Frankfurt, Institute for Cell Biology and Neuroscience, Germany

ARTICLE INFO

Article history:
Received 18 March 2014
Revised 15 July 2014
Accepted 1 August 2014
Available online 12 August 2014

Keywords:
Bayesian analyses
Ecdysozoa
Indel
Next generation sequencing
Phylogenomics

ABSTRACT

The monophyly of Ecdysozoa, which comprise molting phyla, has received strong support from several lines of evidence. However, the internal relationships of Ecdysozoa are still contended. We generated expressed sequence tags from a priapulid (penis worm), a kinorhynch (mud dragon), a tardigrade (water bear) and five chelicerate taxa by 454 transcriptome sequencing. A multigene alignment was assembled from 63 taxa, which comprised after matrix optimization 24,249 amino acid positions with high data density (2.6% gaps, 19.1% missing data). Phylogenetic analyses employing various models support the monophyly of Ecdysozoa. A clade combining Priapulida and Kinorhyncha (i.e. Scalidophora) was recovered as the earliest branch among Ecdysozoa. We conclude that Cycloneuralia, a taxon erected to combine Priapulida, Kinorhyncha and Nematoda (and others), are paraphyletic. Rather Arthropoda (including Onychophora) are allied with Nematoda and Tardigrada. Within Arthropoda, we found strong support for most clades, including monophyletic Mandibulata and Pancrustacea. The phylogeny within the Euchelicerata remained largely unresolved. There is conflicting evidence on the position of tardigrades: While Bayesian and maximum likelihood analyses of only slowly evolving genes recovered Tardigrada as a sister group to Arthropoda, analyses of the full data set, and of subsets containing genes evolving at fast and intermediate rates identified a clade of Tardigrada and Nematoda. Notably, the latter topology is also supported by the analyses of indel patterns.

1. Introduction

The phylogenetic relationships of animal phyla are one of the most hotly debated topics of zoology. Resolving early evolutionary events also has fundamental impact on the understanding of animal biology. Based on phylogenetic analyses of rRNA sequences, Aguinaldo and colleagues (Aguinaldo et al., 1997) defined the superphylum "Ecdysozoa", which comprises the molting phyla Arthropoda, Onychophora (velvet worms), Tardigrada (water bears), Nematoda (roundworms), Nematomorpha (horsehair worms), Priapulida (penis worms), Kinorhyncha (mud dragons) and Loricifera. Ecdysozoa include the most species-rich animal phylum (Arthropoda) and thus outnumber the other protostome superphylum (Lophotrochozoa) and the Deuterostomia (Telford et al., 2008). In addition to the process of molting of the three-layered cuticle, which is controlled by ecdysteroid hormones, Ecdysozoa share only few other morphological characters ("synapomorphies"), including the lack of ciliated epithelia and the absence of spiral cleavage (Giribet and Ribera, 1998; Schmidt-Rhaesa et al., 1998; Telford et al., 2008, 2009). The "Ecdysozoa" hypothesis is at odds with the more "traditional" animal systematics, which holds the view of a close relationship of panarthropods (Arthropoda plus Onychophora and Tardigrada) and annelids (which are now regarded as members of the superphylum "Lophotrochozoa"), and a common origin of animals with a coelomate body cavity (Westheide and Rieger, 1996; Brusca and Brusca, 2003).

The monophyly of Ecdysozoa has received support from molecular phylogenetic studies using selected genes (Mallatt et al., 2004; Webster et al., 2006; Bourlat et al., 2008; Dunn et al., 2008; Telford et al., 2008; Hejnol et al., 2009). Still, several approaches that applied large datasets deriving from whole genomes suggested that *Drosophila melanogaster* (Arthropoda) is closer related to humans than to *Caenorhabditis elegans* (Nematoda), thereby supporting the Coelomata concept (Blair et al., 2002; Wolf et al., 2004; Philip et al., 2005; Ciccarelli et al., 2006; Rogozin et al., 2007). However, others have argued that this topology was the result of long branch

Abbreviations: EST, expressed sequence tag; ML, maximum likelihood; mya, million years ago; SIC, simple indel coding.

* Corresponding author. Address: Institute of Zoology and Zoological Museum, Biocenter Grindel, University of Hamburg, Martin-Luther-King-Platz 3, D-20146 Hamburg, Germany. Fax: +49 40 42838 3937.
E-mail address: thorsten.burmester@uni-hamburg.de (T. Burmester).

http://dx.doi.org/10.1016/j.ympev.2014.08.001
1055-7903/

attraction (LBA), which positions e.g. the nematode *C. elegans* close to the root (Copley et al., 2004; Irimia et al., 2007). In fact, inclusion of additional taxa, a procedure that tends to reduce the effect of LBA on phylogenetic tree reconstruction, consistently recovered Ecdysozoa (Philippe et al., 2005; Webster et al., 2006; Roeding et al., 2007; Dunn et al., 2008; Lartillot and Philippe, 2008; Meusemann et al., 2010; Campbell et al., 2011).

While the Ecdysozoa concept has become widely accepted, the relationships within the Ecdysozoa are not well resolved (for review, see: Telford et al., 2008, 2009; Schmidt-Rhaesa, 2013). There is general agreement that Arthropoda and Onychophora are closely related phyla (Westheide and Rieger, 1996; Brusca and Brusca, 2003) and that Nematomorpha are associated with Nematoda (Nielsen, 1995; Dunn et al., 2008; Telford et al., 2008; Schmidt-Rhaesa, 2013). Otherwise, ecdysozoan relationships are disputed. For example, tardigrades have been traditionally considered to be allied with Arthropoda (Westheide and Rieger, 1996; Brusca and Brusca, 2003), a topology that is tentatively supported by a shared microRNA (Campbell et al., 2011), shared structures of the nervous system (Mayer et al., 2013) and *engrailed* expression patterns (Gabriel and Goldstein, 2007). Molecular studies using large-scale sequence alignments suggested that tardigrades may be more closely related to Nematoda (Giribet, 2003; Roeding et al., 2007; Lartillot and Philippe, 2008; Meusemann et al., 2010), although this topology may also be attributed to long-branch attraction (Rota-Stabelli et al., 2011). The worm-like ecdysozoan phyla (i.e., Nematoda, Nematomorpha, Priapulida, Kinorhyncha and Loricifera) have been referred to as "Cycloneuralia" (Schmidt-Rhaesa, 2013). This classification is at odds with studies that e.g. found the priapulids as sister taxon of all other Ecdysozoa (Webster et al., 2006; Lartillot and Philippe, 2008).

The poor resolution of ecdysozoan relationships is most likely due to the lack of data from important taxa. Because of their enormous biological, ecological and biomedical importance, a huge amount of sequences has been generated from Arthropoda and Nematoda, whereas the other ecdysozoan phyla are considerably undersampled. While the sequencing of specifically selected genes for molecular phylogenetic purposes is a tedious procedure that usually leads to short multiple sequence alignments, more recent molecular phylogenetic studies mostly rely on expressed sequence tags (ESTs). We approach to resolve the relationships among Ecdysozoa by obtaining transcriptomes of key taxa employing next generation sequencing. In addition to the phylogenetic approach based on multigene alignments, we traced the evolution of Ecdysozoa by analyzing indel patterns.

2. Materials and methods

2.1. Species collection and RNA isolation

New transcriptome data from eight ecdysozoan species were generated (see also Supplemental Table S1). Specimens of five chelicerates were used in this study: *Gluvia dorsalis* (Solifugae), *Mastigoproctus giganteus* (Uropygi), *Euphrynichus bacillifer* (Amblypygi), *Phalangium opilio* (Opiliones), *Chelifer cancroides* (Pseudoscorpiones). Additionally, transcriptomes of the tardigrade *Echiniscus testudo* (Echiniscoidea), the priapulid *Halicryptus spinulosus* (Halicryptomorphida) and the kinorhynch *Pycnophyes kielensis* (Homalorhagida) were sequenced. Total RNA of each species was extracted according to Holmes and Bonner (Holmes and Bonner, 1973).

2.2. Transcriptome sequencing

cDNA libraries were constructed using a modified template-switching (SMART) procedure (Mint-Universal cDNA synthesis kit, Evrogen, Russia) and sequenced with the 454 GS FLX Titanium chemistry (Roche). Each cDNA library was sequenced in a half PicoTiterPlate (Roche) according to the manufacturer's protocol. Transcriptome sequencing of *G. dorsalis*, *M. giganteus*, *E bacillifer*, *P. opilio*, *C. cancroides* and *H. spinulosus* was carried out at the Max Planck Institute for Molecular Genetics, Berlin, Germany. The transcriptomes of *E. testudo* and *P. kielensis* were sequenced by LGC Genomics GmBH (Berlin, Germany). Vector-clipping, trimming and quality checking of raw sequence reads and assembly into contigs were performed at the Center for Integrative Bioinformatics (CIBIV), Vienna, Austria. The transcriptomes were checked for possible contaminations with various BLAST-based approaches by cross-comparisons and searches with known protein sequences. Raw data have been deposited in the NCBI Sequence Read Archive (SRA) and assembled contig sequences are available from the Transcriptome Sequences Database (TSA) (BioProject IDs PRJNA236247, PRJNA236248, PRJNA236250, PRJNA236252, PRJNA236253, PRJNA236410, PRJNA236410, PRJNA258412).

2.3. Taxon sampling and orthology assignment

In addition to the assemblies of the six ecdysozoan species, gene predictions of all ecdysozoan genome projects were added to the dataset and transcriptome data of all ecdysozoan species which contained more than 1000 contigs were obtained from the Deep Metazoan Phylogeny (DMP) database (http://www.deep-phylogeny.org/). If more than two species from the same order fulfilled these criteria, only the top two species were selected. The resulting dataset comprises 63 species: 50 ecdysozoans, nine other protostomes and four deuterostomes.

Orthology assignment was performed employing the HaMStR pipeline (Ebersberger et al., 2009); http://sourceforge.net/projects/hamstr). A reference set of 1253 orthologous sequence clusters was used in the analysis, which is based on the proteomes of seven primer taxa: *Apis melifera*, *Caenorhabditis elegans*, *Capitella capitata*, *Daphnia pulex*, *Helobdella robusta*, *Lottia gigantea* and *Schistosoma mansoni* (http://www.deep-phylogeny.org). HaMStR was run with the -strict option using sequentially all seven primer taxa as reference species for the reverse BLAST search. A candidate sequence was only then accepted as an ortholog if it obtained the corresponding reference protein as best hit in all seven BLAST searches.

2.4. Multiple sequence alignments and generation of datasets

Each group of orthologous proteins was aligned individually using MAFFT L-INS-i v7.013 (Katoh and Standley, 2013). Trailing and leading gaps were coded as missing data in each gene alignment. Poorly aligned sections were eliminated by Gblocks v0.91b (settings: −b2 = 41 [65% of the number of sequences] −b3 = 10 −b4 = 5 −b5 = a; Talavera and Castresana, 2007). Alignments of orthologous proteins with less than 50% taxon coverage were removed from the dataset. Finally, the individual alignments were concatenated into a single supermatrix.

In an additional approach, the dataset was divided into three partitions based on the average substitution rate of each gene. To avoid skewed results due to missing data, only the taxa for which sequence data from all genes was available were used for the assessment of substitution rates (*A. melifera*, *C. elegans*, *C. capitata*, *D. pulex*, *H. robusta*, *L. gigantea* and *S. mansoni*). The substitution rates were calculated as the average sum of pairwise scores of all positions in the alignment according to a PAM150 matrix. Positions with gaps were ignored. The individual alignments were concatenated into three subsets (slow, intermediate and fast), processed with Gblocks and used for tree reconstruction as described.