A Phylogenomic Approach to Resolve the Arthropod Tree of Life

Karen Meusemann,†,[1] Björn M. von Reumont,†,[1] Sabrina Simon,[2] Falko Roeding,[3] Sascha Strauss,[4] Patrick Kück,[1] Ingo Ebersberger,[4] Manfred Walzl,[5] Günther Pass,[6] Sebastian Breuers,[7] Viktor Achter,[7] Arndt von Haeseler,[4] Thorsten Burmester,[3] Heike Hadrys,[2,8] J. Wolfgang Wägele,[1] and Bernhard Misof*†,[3]

[1]Zoologisches Forschungsmuseum Alexander Koenig, Molecular Biology Unit, Bonn, Germany
[2]Stiftung Tieraerztliche Hochschule Hannover, Institute of Ecology & Evolution, Hannover, Germany
[3]Biocenter Grindel & Zoological Museum, University of Hamburg, Hamburg, Germany
[4]Center for Integrative Bioinformatics Vienna, Max F Perutz Laboratories, University of Vienna, Medical University of Vienna, University of Veterinary Medicine of Vienna, Vienna, Austria
[5]Department of Theoretical Biology, University of Vienna, Vienna, Austria
[6]Department of Evolutionary Biology, University of Vienna, Vienna, Austria
[7]Regional Computing Center of Cologne (RRZK), The University of Cologne, Cologne, Germany
[8]Department of Ecology and Evolutionary Biology, Yale University
†These authors contributed equally to this work.

***Corresponding author:** E-mail: bernhard.misof@uni-hamburg.de.

Associate editor: Barbara Holland

Abstract

Arthropods were the first animals to conquer land and air. They encompass more than three quarters of all described living species. This extraordinary evolutionary success is based on an astoundingly wide array of highly adaptive body organizations. A lack of robustly resolved phylogenetic relationships, however, currently impedes the reliable reconstruction of the underlying evolutionary processes. Here, we show that phylogenomic data can substantially advance our understanding of arthropod evolution and resolve several conflicts among existing hypotheses. We assembled a data set of 233 taxa and 775 genes from which an optimally informative data set of 117 taxa and 129 genes was finally selected using new heuristics and compared with the unreduced data set. We included novel expressed sequence tag (EST) data for 11 species and all published phylogenomic data augmented by recently published EST data on taxonomically important arthropod taxa. This thorough sampling reduces the chance of obtaining spurious results due to stochastic effects of undersampling taxa and genes. Orthology prediction of genes, alignment masking tools, and selection of most informative genes due to a balanced taxa–gene ratio using new heuristics were established. Our optimized data set robustly resolves major arthropod relationships. We received strong support for a sister group relationship of onychophorans and euarthropods and strong support for a close association of tardigrades and cycloneuralia. Within pancrustaceans, our analyses yielded paraphyletic crustaceans and monophyletic hexapods and robustly resolved monophyletic endopterygote insects. However, our analyses also showed for few deep splits that were recently thought to be resolved, for example, the position of myriapods, a remarkable sensitivity to methods of analyses.

Key words: arthropod phylogeny, phylogenomics, expressed sequence tags, supermatrix, matrix saturation, relative informativeness.

Introduction

Extensive sequence data from genome and expressed sequence tag (EST) projects were recently used to infer a deep metazoan phylogeny (Bourlat et al. 2006; Roeding et al. 2007; Delsuc et al. 2008; Dunn et al. 2008; Philippe et al. 2009; Hejnol et al. 2009). These phylogenomic studies consistently place arthropods within the superphylum Ecdysozoa. These studies are, however, sparse in their sampling of arthropods. Large groups like pancrustaceans are represented by only a few taxa, and important taxa from chelicerates, myriapods, crustaceans, or hexapods are completely missing. EST studies presenting a broader arthropod taxon sampling focus on pancrustacean and hexapod relationships (Timmermans et al. 2008) or on relationships within pterygote insects (Simon et al. 2009). Other studies are essentially restricted to multigene analyses comprising larger arthropod data sets. Regier et al. (2008) analyzed 62 arthropod taxa covered by mainly three genes, but only for a small subset of 13 taxa were all 68 gene regions present. This multigene matrix, however, had 71% missing data. A large proportion of missing data within a supermatrix might cause problems for phylogenetic inference (Sanderson 2007; Wiens and Moen 2008). The most recent study (Regier et al. 2010) relies on selected 62 nuclear protein coding genes for 75 arthropod taxa. Important taxa assumed to be positioned at basal splits, like proturans (Hexapoda), are still missing and their data set at an amino acid level is relatively small (ca. 13,000 amino acids). Much attention was drawn to large arthropod data sets inferred from ribosomal RNA (rRNA) genes (Mallatt and Giribet 2006; von Reumont et al. 2009). Drawbacks of the rRNA-based studies include a lack of robust signal or conflicts in the data (see von Reumont et al. 2009).

Table 1. Species for Novel EST Data in the Present Study.

Species	Group	Accession No.	RNA Extraction	cDNA Library Construction	No. of EST Raw Data	Proc. EST Sequences	No. of EST Contigs
Peripatopsis sedgwicki	ON	FN232766–FN243241	Urea–phenol	CloneMiner	10,611	10,476	3,452
Endeis spinosa	CH, Pycnogonida	FN211278–FN215339	Urea–phenol	CloneMiner	4,063	4,062	2,672
Limulus polyphemus	CH, Xiphosura	FN224411–FN232765	Urea–phenol	Creator SMART	8,435	8,355	4,050
Archispirostreptus gigas	MY, Diplopoda	FN194820–FN198827	Urea–phenol	Creator SMART	4,032	4,008	2,299
Pollicipes pollicipes	CR, Cirripedia	FN243242–FN247432	Absolutely RNA (Strategene)	CloneMiner	4,224	4,191	1,721
Tigriopus californicus	CR, Copepoda	FN247433–FN252183	Trizol (Invitrogen)	Creator SMART	5,024	5,006	2,598
Triops cancriformis	CR, Branchiopoda	FM868344–FM872274	Trizol (Invitrogen)	Creator SMART	3,981	3,930	2,542
Acerentomon franzi	HE, Protura	FN186135–FN190445	Absolutely RNA (Strategene)	CloneMiner	4,600	4,565	1,995
Campodea cf. *fragilis*	HE, Diplura	FN203025–FN211277	Absolutely RNA (Strategene)	CloneMiner	8,375	8,253	6,407
Anurida maritima	HE, Collembola	FN190447–FN194819	Trizol (Invitrogen)	Creator SMART	4,391	4,373	3,504
Lepismachilis y-signata	HE, Archaeognatha	FN219557–FN224410	Absolutely RNA (Strategene)	CloneMiner	4,895	4,854	2,288
Ischnura elegans[a]	HE, Odonata	FN215340–FN219556	RNAeasy (Quiagen)	Creator SMART	4,219	4,217	3,194
Baetis sp.[a]	HE, Ephemeroptera	FN198828–FN203024	RNAeasy (Quiagen)	Creator SMART	4,225	4,197	3,035

Accession no., accession numbers; proc. EST sequences, number of ESTs after processing; ON, Onychophora; CH, Chelicerata; MY, Myriapoda; CR, Crustacea; HE, Hexapoda.
[a] Simon et al. (2009).

Despite this recent progress, these studies fail to completely resolve the arthropod tree of life, leaving many important questions open.

To alleviate the limitations of previous studies, we compiled a more comprehensive set of 233 taxa (214 euarthropod taxa plus 3 onychophorans, 2 tardigrades, and 14 outgroup taxa) and 775 putative orthologous genes that cover 350,356 amino acid positions. We contribute data of 11 new EST projects from velvet worms, millipedes, sea spiders, barnacles, copepods, branchiopods, proturans, diplurans, springtails, and bristletails. Recently published data on dragon- and mayflies (Simon et al. 2009) were also added. These 13 projects fill critical gaps in the published data (table 1). Previous phylogenomic analyses have shown that beside massive accumulation of data, several additional elements must be part of the analysis pipeline: careful selection of orthologs, consideration of data quality, reduction of data gappiness, and model fitting (Roeding et al. 2007; Dunn et al. 2008; Hartmann and Vision 2008; Philippe et al. 2009). Consequently, we used recently developed tools for ortholog gene prediction (Ebersberger et al. 2009, see supplementary fig. 1, Supplementary Material online) and alignment masking (Misof and Misof 2009), which facilitate a completely reproducible data analysis. Moreover, we applied new heuristics of selecting an optimal data set from a supermatrix to increase the number of taxa with potentially informative genes (supplementary fig. 2, Supplementary Material online); this contrasts with other recent studies (Dunn et al. 2008; Regier et al. 2010) that rely on presence|absence matrices. The logic behind our approach is to reduce effects of poorly represented taxa and of uninformative genes by identifying and filtering these prior to tree reconstruction (see Methods and supplementary figs. 3–5, Supplementary Material online). This preprocessing improves the signal-to-noise ratio in the data and considerably helped to reduce the effort spent in tree reconstructions. Retention of taxa and genes in the supermatrix was based on their contribution to the overall informativeness and the matrix saturation of the data matrix (=number of present gene entries in relation to the total size of the matrix) prior to tree reconstructions, thus allowing a better exploration of tree space.

Materials and Methods

Molecular Techniques

For thirteen arthropod species, cDNA libraries were constructed. Total RNA was prepared with standard kits from tissue or complete specimens preserved in RNA later or liquid nitrogen and stored at −80 °C, or total RNA was directly prepared from living specimens using Urea–phenol following Holmes and Bonner (1973) (table 1). For crustaceans and apterygote hexapods, RNA preparation was conducted by the Max Planck Institute for Molecular Genetics (MPIMG), Berlin, Germany. The cDNA libraries were constructed using CloneMiner (Invitrogen) or Creator SMART (Clontech, Heidelberg, Germany) at the MPIMG; cDNA libraries for pterygote insects were normalized (Simon et al.

2009). From cDNA libraries, ESTs were generated by sequencing clones from the 5′ end on the automated capillary sequencer system ABI 3730XL (Applied Biosystems, Darmstadt, Germany) using BIGDYE chemistry (Applied Biosystems). Between 3,930 and 10,476 sequences were processed from cDNA libraries (table 1). All single EST sequences were deposited in EMBL (http://www.ebi.ac.uk/embl/) after being quality checked and assembled into unique transcripts (contigs), whereby two projects on pterygote insects originally sequenced for this arthropod study have recently been published (Simon et al. 2009).

Sequence Processing and Orthology Assignment

We preprocessed new EST data (table 1) with LUCY (Chou and Holmes 2001). EST data available for 190 additional euarthropods (myriapods, chelicerates, pancrustaceans) plus 2 onychophorans, 2 tardigrades, and selected species of nematodes, annelids, and molluscs (in total 216 species) were extracted from public databases, dbEST (NCBI), the Gene Index Project or the NCBI Trace Archive (supplementary table 1, Supplementary Material online). We screened all EST sequences for contamination and low-quality ends of sequences. Subsequently, overlapping ESTs from the same taxon were assembled into contigs using the TGICL package (Pertea et al. 2003). For the orthology prediction with HaMStR (Ebersberger et al. 2009), all contigs were translated into amino acid sequences in all reading frames. In total, 244 species were "hamstred" of which 28 species were "proteome" species. Thirteen species were used as primer taxa (supplementary fig. 1 and supplementary table 1, Supplementary Material online). Sequences of vertebrate species were additionally used to train profile Hidden Markov Models (Ebersberger et al. 2009) but excluded in further phylogenetic analyses for computational reasons. Eight *Drosophila* proteome species were also excluded for computational reasons. The HaMStR search identified 775 putative orthologous genes for our original data set (233 species).

Alignments and Alignment Masking

Inferred amino acid sequences of all 775 putative orthologous genes were aligned (supplementary fig. 2, Supplementary Material online) with MAFFT *L-INSI* (Katoh and Toh 2008). The data set comprised 222 euarthropods, 3 onychophorans, 2 tardigrades, 3 vertebrates, 8 nematodes, 3 annelids, and 3 molluscs. Excluding randomly similar aligned sections can make phylogenetic analyses more reliable prior to tree reconstruction (Castresana 2000; Misof and Misof 2009; Kück et al. 2010). We therefore identified randomly similar sections for all gene alignments separately for each of the 775 genes with ALISCORE on the amino acid level (Misof and Misof 2009; Kück et al. 2010) using default settings and maximal number of pairwise comparisons. In total, 57.62% of originally 826,633 amino acid positions were excluded to increase the signal-to-noise ratio. For each gene, only sequences comprising more than one half of the sequence information were included in the ALISCORE analyses. We masked each alignment with ALICUT (http://www.utilities.zfmk.de) by excluding all randomly similar alignment positions. All masked alignments were concatenated to a masked superalignment comprising 233 taxa and 350,356 amino acid positions.

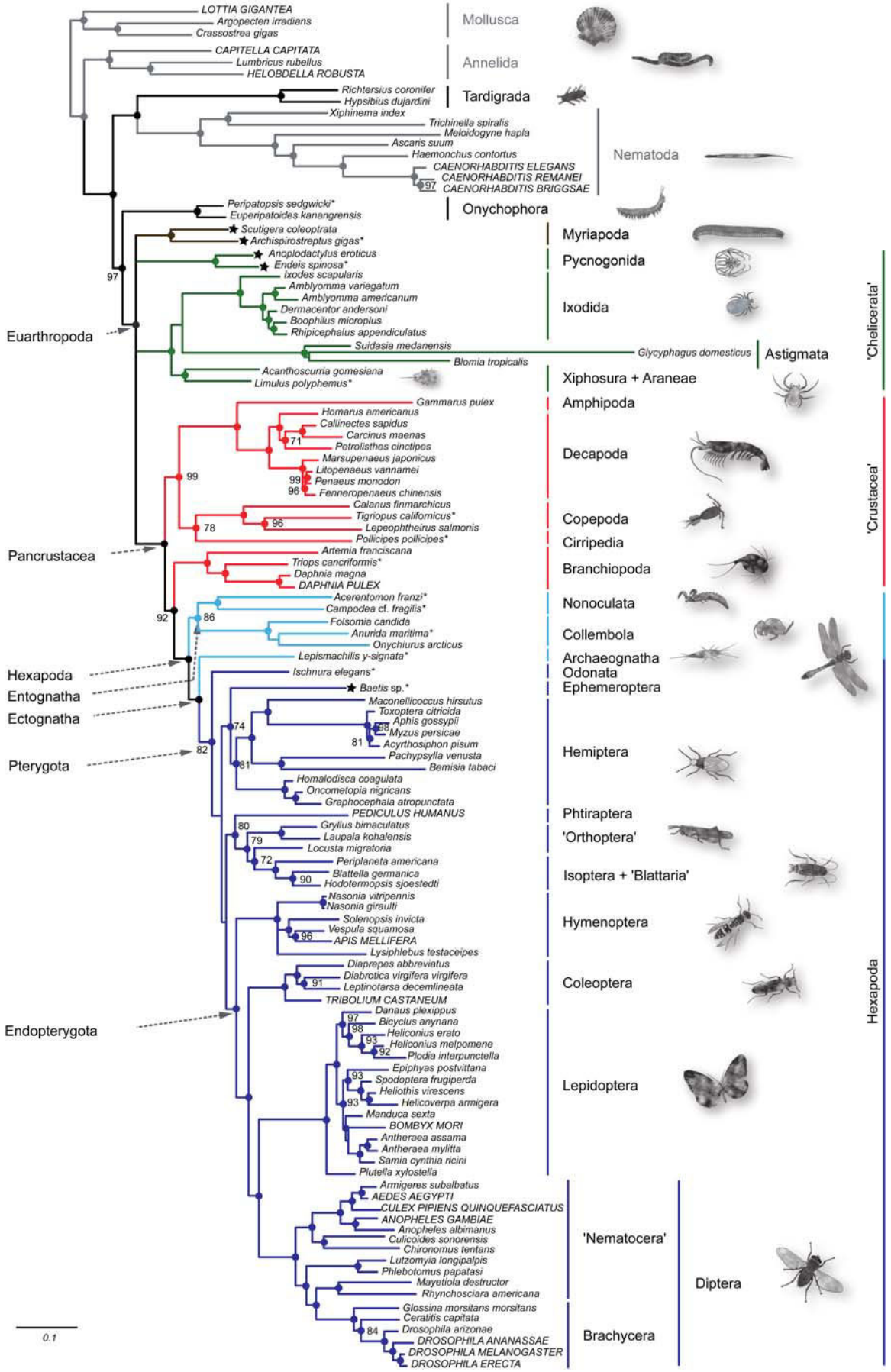
Selecting an Optimal Subset Using New Reduction Heuristics

With the software MARE (MAtrix REduction) (http://mare.zfmk.de), the relative informativeness of each single gene within a superalignment was calculated based on weighted geometry quartet mapping (Nieselt-Struwe and von Haeseler 2001), extended to amino acid data. Each gene received a value of informativeness between 0.0 and 1.0, reflecting the relative number of resolved quartet trees (supplementary fig. 3, Supplementary Material online). A data availability matrix indicating present (1) and absent (0) genes was then transformed into a matrix of potential information content of each taxon and gene by multiplying availability (0|1) with scores of informativeness. Relative information content of each gene was calculated as the average value over all taxa including missing taxa. The total average information content P (relative informativeness) of a supermatrix was calculated as the sum of relative information content of all genes in relation to the number of taxa (see supplementary fig. 4, Supplementary Material online). To select an optimal subset of taxa and genes with high total average information content, we used a simple hill climbing procedure. Reduction starts with dropping either taxon (row) or gene (column) with the lowest average information content, generating a new matrix. In case of ties, genes are excluded. Consequently, taxa or genes with lowest average information content will be discarded from the matrix, yielding a selected optimal subset (SOS) of taxa and genes with increased relative information content (supplementary fig. 5, Supplementary Material online). We defined the copepod *Tigriopus* and the chilopod *Scutigera* as taxon constraints; thus, they were not dropped from the submatrix. Copepods are discussed as a sister group to hexapods (Mallatt and Giribet 2006; von Reumont et al. 2009), and *Scutigera* was the only representative of chilopods (Myriapoda). Therefore, we constrained matrix reduction to retain both species as key taxa.

In order to reach an optimum of matrix reduction, we defined an optimality function $f(\mathrm{P})$, which takes into account that size reduction of an original matrix B and low total average informativeness of a reduced matrix B' are penalized

$$f(\mathrm{P}) = 1 - |(\lambda - \mathrm{P}^{\alpha\times(1-\mathrm{P})})| \quad \text{if } \mathrm{P} < 1, \qquad (1)$$

with α as a scaling factor (default set to $\alpha = 3$), λ as the size ratio between reduced B' and original matrix B (matrix size defined as #taxa × #genes). P is maximized, if $\mathrm{P} = 1$, reduction stops. The optimality function favors reduction of matrices to high average information content. The connectivity between taxa was set to a minimum number of two overlapping genes and taxa. This means that two sets of taxa must share at least two taxa with both genes. Finally, the original superalignment was rewritten based on the SOS. Details of the new reduction algorithm will be published



elsewhere (Misof B, Meyer B, von Reumont BM, Kück P, Meusemann K, unpublished data).

Phylogenetic Analyses

We conducted maximum likelihood (ML) analyses using RAxML Pthreads 7.0.0 (Stamatakis 2006b; Ott et al. 2007) for 1) the original data set (unreduced supermatrix) comprising 233 taxa, 775 genes, and 350,356 amino acid positions and 2) the SOS comprising 117 taxa and 129 genes with an alignment length of 37,476 amino acid positions. The data matrices have been deposited at Treebase (http://purl.org/phylo/treebase/phylows/study/TB2/S10507).

We applied ML tree search and rapid bootstrapping within one step (-f a, 1,000 bootstrap replicates) on the SOS. For the original concatenated supermatrix, we conducted ten single ML tree searches and separate bootstrapping (100 replicates). We chose the ML tree with the best likelihood value to plot bootstrap values (supplementary fig. 6, Supplementary Material online). All ML analyses were calculated with the PROTMIX (Stamatakis 2006a) substitution model and the WAG matrix (Whelan and Goldman 2001).

Bayesian analyses for the SOS were inferred using PhyloBayes version 2.3c (Lartillot et al. 2008) running the CAT mixture model (Lartillot and Philippe 2004). We ran 25 Markov chain Monte Carlo for 20,000 cycles each, sampling every cycle. All parameter values were checked for convergence to define the burn-in (5,000 cycles). To infer a majority rule consensus (mrc) tree, we checked the discrepancy observed across all bipartitions (*maxdiff* value) of all chains by pairwise comparison and comparing "triple"-chain combinations with the *bpcomp* tool. Harmonic means of the likelihood values of each chain (burn-in excluded) were calculated. To infer the Bayesian mrc tree, we included three chains showing the lowest *maxdiff* value (0.186) while featuring the best likelihood values (harmonic means) of all "triple-chain combinations" (table 2). All trees were rooted with Mollusca.

To identify "unstable" taxa, we calculated leaf stability indices (Thorley and Wilkinson 1999) from the collected bootstrap trees of the ML analysis using Phyutility (Smith and Dunn 2008). We defined a threshold of <95% as unstable. All analyses ran for several months on Linux Clusters, HP ProLiant DL380 G5 blades (Dual quad core Intel Xeon E5345, 2.33 GHz, 2x 4MB L2-cache, 1333 MHz Bus, 32 GB RAM), of the ZFMK (molecular unit) and the RRZK (Regional Computing Center of Cologne) utilizing HPC resources (HP ProLiant, Dual quad core Intel Xeon E5345, 2.33 GHz, 2x 4MB L2-cache, 1333 MHz Bus, 32 GB RAM). RRZK resources were provided by the SuGI (Sustainable Grid Infrastructure) project (Project leader: V. Achter, University of Cologne funded by the BMBF).

Consensus Network of Single Bayesian Topologies

Due to differences between single topologies of the 25 PhyloBayes (Lartillot et al. 2008) chains, we computed a consensus network (Holland and Moulton 2003) with SplitsTree 4.8 (Huson and Bryant 2006). This is a method to identify contradictory signal that cannot be displayed with a simple mrc tree. To visualize conflicts and contradictory signal, we chose a threshold of 0.01 and incorporated averaged edge weights.

Results and Discussion

The SOS

Our SOS includes 117 taxa with 101 euarthropods, 2 onychophorans, 2 tardigrades, and 12 outgroup taxa (supplementary table 1, Supplementary Material online). The data set comprises 129 genes of which 32 genes coded for ribosomal proteins and 97 for nonribosomal proteins (supplementary table 2, Supplementary Material online). The relative information content of genes ranges from 0.42 to 0.92, with an average of 0.7 (supplementary tables 1 and 2, Supplementary Material online). The concatenated masked alignment spans 37,476 amino acid positions (supplementary fig. 4, Supplementary Material online). The relative informativeness rises 4-fold from 0.10 (original data set) to 0.43 (SOS) (supplementary figs. 3–5, Supplementary Material online). Matrix saturation (genes with a relative information content <0.04 considered as missing) increases 3-fold from originally 17.6–62.3% in the SOS. Taxa in the SOS cover on average 84 genes (minimum 35 and maximum 129). Each gene is, on average, present in 76 taxa (minimum 46 and maximum 109 taxa per gene).

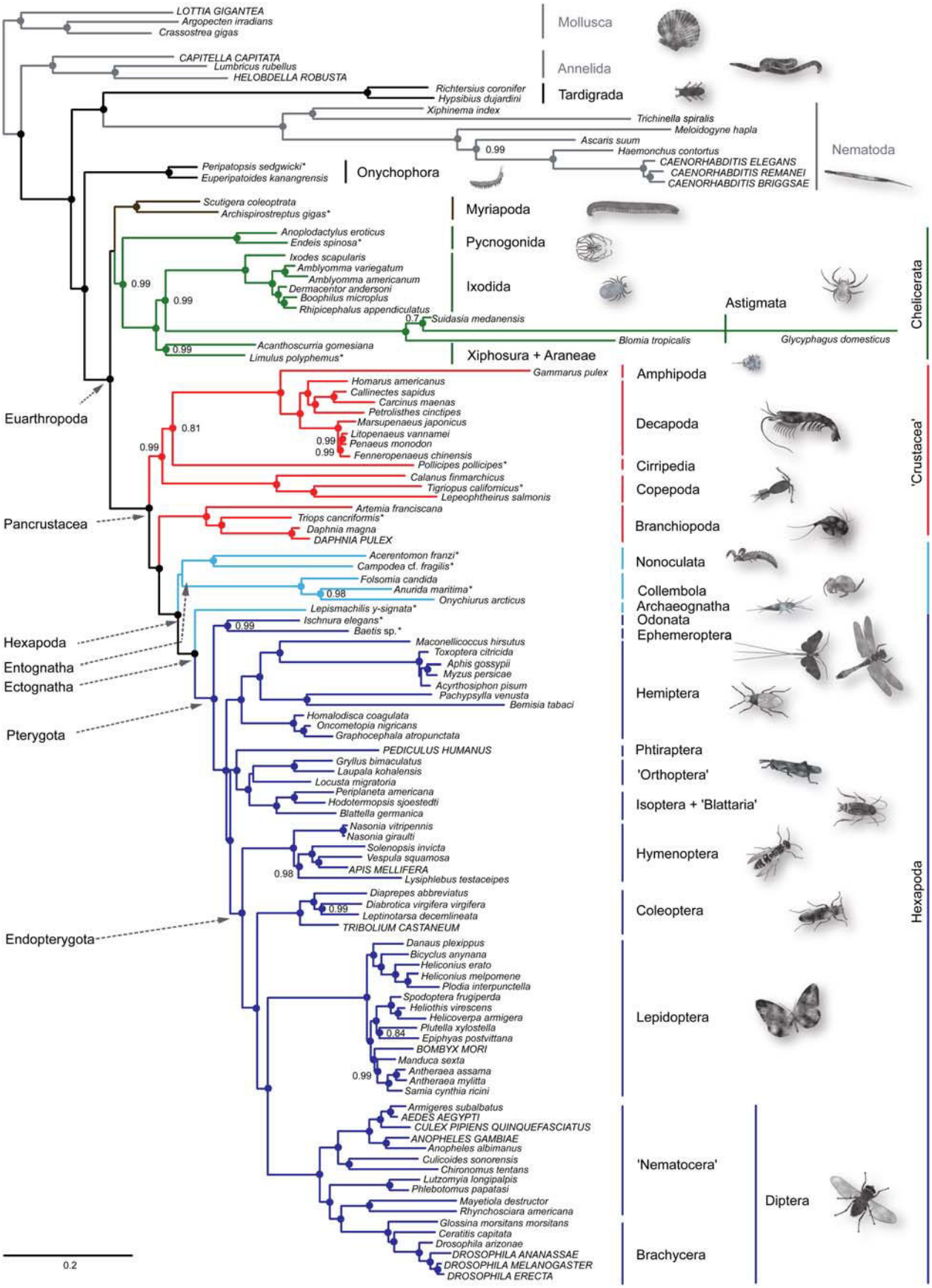
ML and Bayesian tree reconstruction of the SOS resolved arthropod relationships with several strongly supported nodes (figs. 1 and 2 and table 3). In contrast, the tree based on the original supermatrix is in many respects unresolved or shows low support values (supplementary fig. 6, Supplementary Material online). This comparison suggests that the strategy to compute an SOS is successful, for example, improves tree robustness and clades that are widely accepted in the literature (e.g., Hexapoda, Ectognatha, Endopterygota, Coleoptera, Lepidoptera), which was not the case for the unreduced data set. Thus, the discussion of the phylogenetic relationships focuses on the SOS.

Incongruences in Bayesian Analyses

The 25 Bayesian runs did not converge on a single topology (see Methods; fig. 3). Some clades, for example

←

FIG. 1. Phylogram of 117-taxon ML analysis. RAxML tree (majority rule) of the SOS, PROTMIX substitution model + WAG matrix. Support values are derived from 1,000 bootstrap replicates. Support values <70: not shown; support values = 100: represented by a dot only. Quotation marks indicate nonmonophyly. Asterisks indicate EST taxa contributed by the authors. Unstable taxa (leaf stability index <0.95) are marked by a star in front of the taxon name. Color code: molluscs, annelids, and nematodes, lighter grey; tardigrades and onychophorans, black; myriapods, brown; chelicerates, green; crustaceans, red; basal hexapods, light blue; and pterygote insects, dark blue.



(Onychophora, Euarthropoda), Pancrustacea, Branchiopoda as a sister group to Hexapoda and Nonoculata (Protura, Diplura), emerged in all chains with maximal support. Other clades differed between consensus trees inferred from single chains. These incongruences were caused by unstable positions of few taxa (fig. 3): 1) Mandibulata (Myriapoda + Pancrustacea) were found maximally supported in consensus trees of two runs. Both runs show comparatively low harmonic means of likelihoods. In all other runs, myriapods clustered with chelicerates with negligible to moderate support (posterior probability pP 0.52–0.89). 2) The barnacle *Pollicipes* (Cirripedia, Crustacea) emerged as a sister group to copepods in only one run (pP 0.51). However, the alternative clade (*Pollicipes* + Malacostraca) (fig. 2) showed a wide range from 0.56 to 0.96 pP in other runs. 3) The bristletail *Lepismachilis* (Archaeognatha) was inferred as a sister group to Blattaria + Isoptera in several runs, showing moderate or low support (pP 0.52–0.82). Additionally, *Pediculus* (Phthiraptera) emerged as a sister group to this clade (pP 1.0) in these runs. Likelihoods (harmonic means), however, were lower compared with runs used for our Bayesian consensus tree (fig. 2), and results of these runs were rejected after a Bayes factor test (Kaas and Raftery 1995; Nylander et al. 2004). 4) Among butterflies (Lepidoptera), five different topologies with distinctive clades were found. Differences occurred among Yponomeutoidea, Papilionoidea, Pyraloidea, Tortricoidea, and Noctuidea. Incongruent consensus trees might reflect different local optima despite extensive sampling.

Are the Enigmatic Tardigrades and Onychophorans Arthropods *sensu latu*?

Chelicerates, myriapods, crustaceans, and hexapods show highly derived differentiations of segments and segmental appendages (Edgecombe 2009). Tardigrades and onychophorans display a mosaic of plesiomorphic and autapomorphic features of segmental differentiation. The evolution of the arthropod bauplan, as, for example, the evolution of segmentation, appendages, and the central nervous system, can thus be understood only if the phylogenetic positions of tardigrades and onychophorans are resolved (see reviews in Budd and Telford 2009 and Edgecombe 2009).

Tardigrades are tiny animals with morphological characters reminiscent of both Arthropoda and Cycloneuralia (the latter named for their circumpharyngeal nerve ring shared by Nematoda, Nematomorpha (horsehair worms, insect parasites), Priapulida (penis worms), Kinorhyncha (mud dragons), and Loricifera; see Giribet 2003; Edgecombe 2009). Arthropod-like characters include the segmented body, limbs, the presence of a peritrophic membrane, and a ladder-like central nervous system (Giribet 2003). In contrast, structures of mouth, pharynx, cuticle, and sensory organs resemble those of Cycloneuralia (Giribet 2003). Traditionally, tardigrades have been allied with arthropods, an assumption that has been corroborated by molecular studies based on rRNA (Mallatt et al. 2004). Such a clade Tardigrada + Onychophora + Euarthropoda (Panarthropoda) would be compatible with the hypothesis of an evolution of segmentation (including differentiation of the muscular tube, etc.), segmented appendages, and a ladder-like central nervous system within this clade. Alternatively, a sister group relationship of tardigrades with Cycloneuralia (nematodes and allies) would imply either a very ancient evolution of a segmented body plan and a loss of these characteristics within derived Cycloneuralia (including a reversal to an undifferentiated muscular tube) or an independent evolution of segmental characters within Cycloneuralia. A robustly resolved position of tardigrades has a strong impact on our interpretation of the evolution of segmentation.

In our analyses, tardigrades (*Hypsibius* and *Richtersius*) emerge as a sister group of nematodes (bootstrap support 100%, pP 1.0), which is in line with recent findings by Roeding et al. (2007), Lartillot and Philippe (2008), and Bleidorn et al. (2009). These studies had been based on different gene selections. In contrast, Dunn et al. (2008), applying the CAT model (Lartillot and Philippe 2004) of amino acid evolution, found tardigrades either as a sister group of arthropods (including onychophorans) or applying the WAG model, as a sister group of nematodes and nematomorphs, in both cases only weakly supported. Phylogenetic analyses based on morphological characters are similarly ambiguous and support contradicting results. Either panarthropods (including tardigrades, Edgecombe 2009) are favored, or an unresolved clade (Tardigrada + Onychophora + Euarthropoda) is represented in Budd and Telford (2009), or tardigrades are positioned outside the (Onychophora + Euarthropoda) clade (Zantke et al. 2008). Currently, there is no conclusive hypothesis compatible with the contradicting morphological and molecular data about the position of tardigrades within the metazoan tree. This clearly impedes our understanding of the evolution of segmentation within Ecdysozoa.

Onychophorans strongly resemble arthropod-like animals with, for example, a reduction of locomotory cilia, a body cavity with a pericardial septum, a heart with ostia, segmental nephridia with sacculi, the presence of clawed ventral appendages, and the absence of metameric larvae. Deviant from arthropods, onychophorans lack, for example,

←

FIG. 2. Phylogram of 117-taxon Bayesian analysis. Bayesian mrc tree of the SOS, 3 chains out of 25 chains, 20,000 cycles each, burn-in: 5,000 cycles. Support values are estimated under the CAT mixture model. The mrc tree is based on the "triple" (three chains) showing lowest *maxdiff* value (0.186), whereas each of these chains had the best harmonic mean of the likelihood values (burn-in excluded) of all possible triple-chain combinations. pP values <0.7: not shown; pP values = 1.0: represented by a dot only. Quotation marks and color code as specified in figure 1; asterisks indicate EST taxa contributed by the authors.

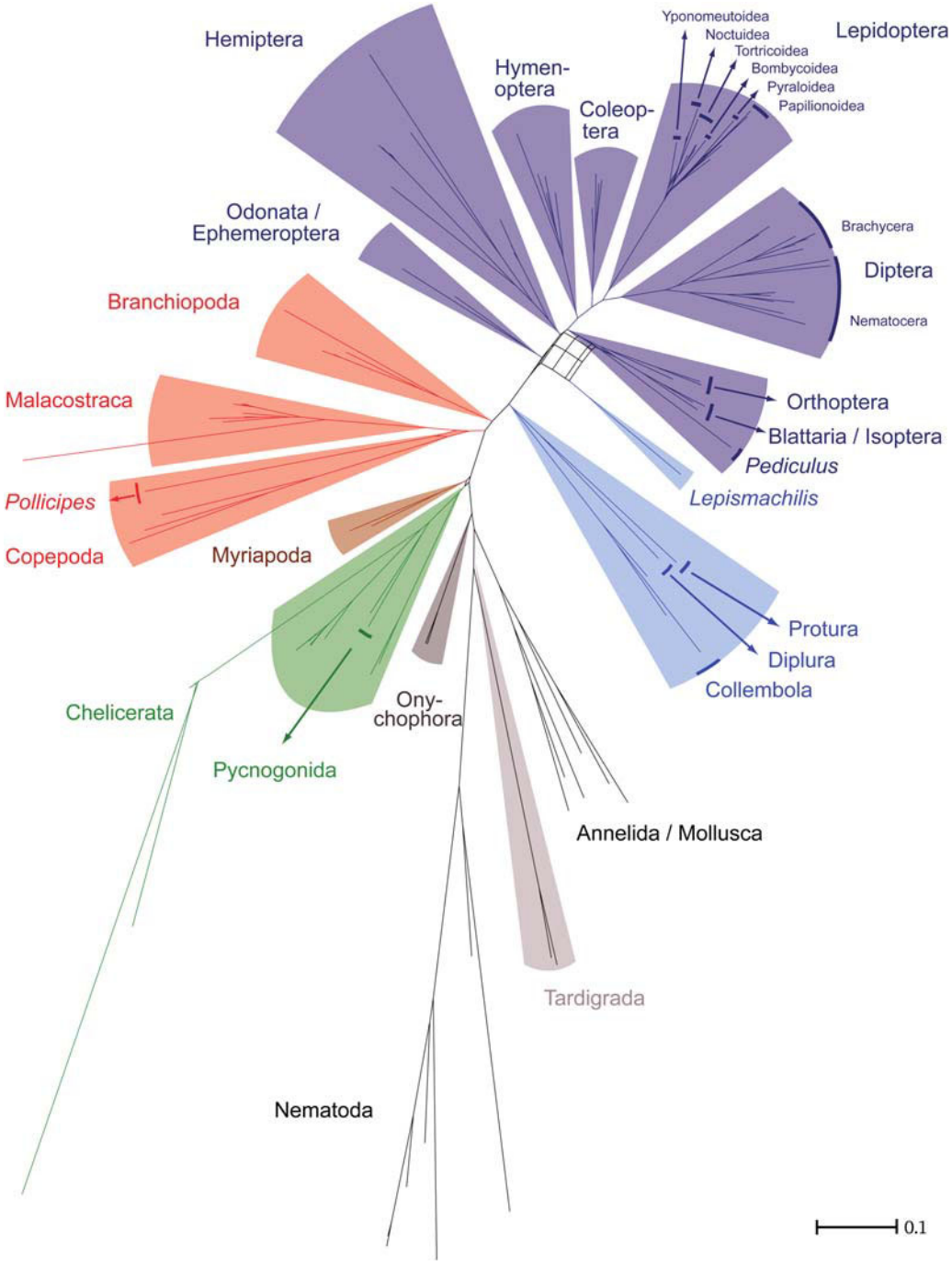


FIG. 3. Consensus network of all 25 PhyloBayes trees. Consensus network of all 25 PhyloBayes chains of the SOS was calculated with SplitsTree 4.8 and visualizes incongruences between 25 topologies (treshold = 0.01, averaged weights). The color code is specified in figure 1.

a complete disintegration of the muscular tube into segmentally arranged muscle systems, segmentally arranged sclerotized exoskeletal structures, and a fully ganglionated organization of the central nervous system. Earlier morphological and molecular analyses have placed onychophorans as either a sister group to Tardigrada + Euarthropoda (Budd and Telford 2009) or sister group to Euarthropoda (Roeding et al. 2007; Dunn et al. 2008; Edgecombe 2009), thus leaving the position of onychophorans unresolved. ML (fig. 1) and Bayesian (fig. 2) analyses of our SOS resolve the position of the onychophorans and show strong support for the clade Onychophora + Euarthropoda. A clade Onychophora + Euarthropoda is compatible with the view that fully differentiated segmentation, including ganglionization of the central nervous system evolved in a common stem-lineage of onychophorans and euarthropods. This view implies that onychophorans primarily lack many characteristics of the euarthropod body organization (Hou and Bergström 1995; Edgecombe 2009). The interpretation of the fossil record of "lobopodian"-grade organisms as possible stem group

Table 2. Log-Likelihood Values (harmonic means) and Chain Combinations of All PhyloBayes Runs for the SOS.

Chain ID	Log Likelihood (harmonic mean, burnin-in excluded)	Chain Combination ("triple" chain)	*Maxdiff* (<0.3)
18	948174.861012454	c04—c18—c20	0.186
04	948217.993492174	c23—c01—c06	0.202933
20	948376.710282837	c21—c23—c08	0.20787
16	948469.642382507	c21—c23—c01	0.18833
05	948525.74067471	c01—c23—c08	0.20787
22	948678.821621205	c21—c08—c01	0.20787
23	948708.71215524	c22—c05—c14	0.23647
21	948752.989770425	c22—c05—c16	0.18653
08	948757.764925626	c22—c14—c16	0.23647
14	948779.209757328	c05—c14—c16	0.1621
01	948865.845517544	All 25 chains	1

Log-likelihood (harmonic means) of all log likelihood values, 20,000 cycles per chain, burn-in (5,000 cycles) excluded; chain combination consisting of three chains each per combination (triple) for which the *maxdiff* value <0.3; *maxdiff*: discrepancy value observed across all bipartitions for the given triple chain (PhyloBayes tool).

representatives of euarthropods is also compatible with this conclusion (Hou and Bergström 1995).

Euarthropoda Including Pycnogonids Favored over the "Cormogonida"

The monophyly of euarthropods is well established, whereas relationships within euarthropods, between myriapods, sea spiders, chelicerates, crustaceans, and hexapods are problematic (compare results of Dunn et al. 2008; Regier et al. 2008, 2010; von Reumont et al. 2009).

Sea spiders (Pycnogonida) represent an extremely aberrant group of arthropods. Earlier morphological and molecular studies have placed sea spiders either as a sister group of Euchelicerata (Bourlat et al. 2008; Brenneis et al. 2008; Dunn et al. 2008) or considered them as the first branch of euarthropods ("Cormogonida" hypothesis, Zrzavý et al. 1998; Maxmen et al. 2005). Although the position of sea spiders is not resolved in the ML tree (fig. 1), the Bayesian tree (fig. 2) shows monophyletic chelicerates including sea spiders with high support (pP 0.99). This result corroborates other phylogenomic analyses (Dunn et al. 2008; Regier et al. 2010, but weakly supported) as well as hox gene and neuroanatomical studies (Jager et al. 2006; Brenneis et al. 2008), which demonstrated the homology of deuterocerebral

Table 3. Selected Clades and Support Values of ML and Bayesian Reconstructions Inferred for the SOS.

Selected Clades	Bootstrap Support (%)	Posterior Probability
(Tardigrada,Nematoda)	100	1
(Onychophora,Euarthropoda)	97	1
((Tardigrada,Nematoda),(Onychophora,Euarthropoda))	100	1
Euarthropoda	100	1
Mandibulata	—	—
Myriochelata	—	0.57
Chelicerata	—	0.99
Euchelicerata	100	1
Pancrustacea	100	1
(Amphipoda,Decapoda)	100	1
(Copepoda,Cirripedia)	78	—
((Amphipoda,Decapoda),(Copepoda,Cirripedia))	99	—
((Amphipoda,Decapoda),Cirripedia)	—	0.81
(((Amphipoda,Decapoda),Cirripedia),Copepoda)	—	0.99
(Branchiopoda,Hexapoda)	92	1
Hexapoda	100	1
Enthognatha	86	0.5
(Collembola,(Protura,Diplura))	86	0.5
Nonoculata: (Protura,Diplura)	100	1
Ectognatha: (Archaeognatha,Pterygota)	100	1
Pterygota	82	1
Chiastomyaria: (Odonata,(Ephemeroptera,Neoptera))	—	—
Paleoptera: (Odonata,Ehemeroptera)	—	0.99
Neoptera	—	1
(Ephemeroptera,Hemiptera)	74	—
Endopterygota	100	1
(Hymenoptera,remaining endopterygote clades)	100	1
(Coleoptera,(Lepidoptera,Diptera))	100	1
(Lepidoptera,Diptera)	100	1

appendages of sea spiders and euchelicerates. It suggests that sea spiders should be included within chelicerates. Our results are inconclusive regarding the position of the pycnogonids, comparing the ML and the Bayesian reconstruction, but the latter agrees with established "nonmolecular" data (Jager et al. 2006; Brenneis et al. 2008) that support pycnogonids as a sister group to Euchelicerata.

The Position of Myriapoda Cause Problems to Address Mandibulata versus Myriochelata

Monophyly of mandibulate arthropods (Myriapoda + Crustacea + Hexapoda) has received substantial support from morphological studies (Richter 2002; Harzsch et al. 2005; Harzsch 2006; Scholtz and Edgecombe 2006; Müller et al. 2007; Bäcker et al. 2008) and from some molecular analyses (Telford et al. 2008; Regier et al. 2008, 2010). Within mandibulates, two alternative clades, either Myriapoda + Hexapoda (Atelocerata, Heymonds 1901, or Tracheata, Pocock 1893) or Crustacea + Hexapoda (Pancrustacea, Zrzavý and Štys 1997) or Tetraconata, Dohle 2001) have been proposed by Grimaldi (2010). Both hypotheses utilize the presence of complex character systems supporting each view (Harzsch 2006; Bäcker et al. 2008; Mayer and Whintington 2009). Molecular evidence, however, has recently accumulated for a clade Myriapoda + Chelicerata, coined Myriochelata (Pisani et al. 2004), or Paradoxopoda (Mallatt et al. 2004). This conflicts with the Mandibulata concept (Mallatt et al. 2004; Roeding et al. 2007; Dunn et al. 2008). At the same time, recent studies have demonstrated a high sensitivity of reconstructing Paradoxopoda with respect to gene choice, taxon sampling, and outgroup selection (Bourlat et al. 2008; Philippe et al. 2009). The most recent study addressing this issue was published by Regier et al. (2010) based on nuclear, mainly nonribosomal protein-coding genes, which again supports Mandibulata. Ribosomal proteins, however, are hardly considered and this result should be interpreted with caution. Furthermore, there is little morphological data supporting a clade Paradoxopoda (Mayer and Whintington 2009) in contrast to data supporting Mandibulata (Wägele 1993; Harzsch 2006; Bäcker et al. 2008). A clade Paradoxopoda would imply the independent evolution of the labium, the loss of the second pair of antennae, and the independent evolution of ectodermal malphigian tubules in myriapods and hexapods.

In our analyses (including ribosomal and nonribosomal single copy genes), the position of myriapods is not resolved. In the Bayesian tree, myriapods emerge as a sister group to chelicerates with low support. In the ML tree, relationships between myriapods, sea spiders, euchelicerates, and pancrustaceans remain unresolved. The results of our phylogenomic analyses and rRNA-based analyses (e.g., von Reumont et al. 2009) indicate that the unstable position of myriapods is not caused by a single myriapod taxon but probably is related to a systematic phenomenon of myriapod molecular evolution. To resolve the myriapod position in the arthropod tree, we therefore need to better understand heterogeneity of substitutional processes among arthropods and to include all myriapod groups in phylogenomic analyses.

Pancrustacea with Branchiopoda as a Sister Group to Hexapoda

Our data support a clade Crustacea + Hexapoda (Pancrustacea, 100% bootstrap support and 1.0 pP). Within crustaceans, relationships are still far from being resolved. Representatives of important crustacean groups are still not covered by EST data. Only few published nonmalacostracan EST projects exist (Branchiopoda, Copepoda, and Cirripedia, presented in this study) (Stillman et al. 2008). Therefore, discussing the sister group of hexapods requires caution, and further EST data for representatives of major crustacean groups (e.g., Remipedia, Leptostraca) are required.

In rRNA-based studies, copepods (Cyclopidae) were found to be a sister group to hexapods (Mallatt and Giribet 2006; Mallatt et al. 2010; von Reumont et al. 2009). In our analyses, Branchiopoda consistently emerge as a sister group to Hexapoda (1.0 pP in the Bayesian approach and moderately supported 92% bootstrap support in ML analyses). This corroborates results of other single- and multigene analyses (Regier et al. 2005; Dunn et al. 2008; Mallatt et al. 2010; Philippe et al. 2009). This well-supported clade Branchiopoda + Hexapoda conflicts with described potential synapomorphies of Malacostraca and Hexapoda (Harzsch 2006), for example, the presence of a third neuropil and chiasmata of the lateral eyes. Ertas et al. (2009) suggest a close relationship of Remipedia and Hexapoda based on hemocyanin. This result is underpinned by neuroanatomical data (Fanenbruck et al. 2004; Fanenbruck and Harzsch 2005). Regier et al. (2010) inferred a clade "Xenocarida" with Remipedia + Cephalocarida as a sister group to Hexapoda, with low support at the amino acid level and high support at nucleotide level. Remipedia as the sister group to Cephalocarida is contradicted by new data on Remipedia larvae (Koenemann et al. 2007, 2009). The incongruence between molecular and morphological results concerning the sister group relationship of hexapods cannot be resolved yet. Careful analyses of signal quality in molecular and morphological data are still required, along with more molecular data on Remipedia and Cephalocarida.

Monophyletic Hexapoda, Entognatha, and Ectognatha

Based on morphological analyses, hexapods are assumed to be monophyletic (Dohle 2001; Bitsch and Bitsch 2004; Harzsch et al. 2005; Harzsch 2006; Ungerer and Scholtz 2008). The monophyly of ectognathous hexapods (Archaeognatha + pterygote insects, see Hennig 1981; Kristensen 1991) seems well founded by single-gene analyses (e.g., Kjer et al. 2006; Misof et al. 2007; von Reumont et al. 2009), is supported by nuclear protein-coding genes (Regier et al. 2010), and also corroborated by our phylogenomic data; this clade "has likewise never been seriously challenged" (Grimaldi 2010).

In contrast, the monophyly of entognathous hexapods (Protura, Diplura, and Collembola) is generally ambiguous (see review of Grimaldi 2010). The interpretation of character states within entognathous hexapods is difficult because of extreme adaptations to subterranean or cryptic

habitats. The presence of many plesiomorphic character states (e.g., presence of fully musculated antennae, abdominal appendages, anameric development [Protura], unsegmented tarsi) gives them an important role in understanding the evolution of hexapods. Our Bayesian and ML analyses recovered Entognatha as a monophyletic group, albeit weakly supported. Within Entognatha, we obtain strong support for a sister group relationship of Protura and Diplura, a clade coined Nonoculata (Luan et al. 2005). This corroborates recent single-gene analyses (Dell'Ampio et al. 2009; Mallatt et al. 2010; von Reumont et al. 2009). Morphological evidence for this clade is still ambiguous (Szucsich and Pass 2008). Our results disagree with inferred relationships of primary wingless hexapods based on mitochondrial data (Nardi et al. 2003; Carapelli et al. 2005, 2007). Those authors proposed the polyphyly of hexapods with a placement of springtails (Collembola) as a sister group to other pancrustacean taxa, implying that features of the hexapod bauplan evolved at least twice. Reanalyses of these mitochondrial data (Delsuc et al. 2003) yielded monophyletic hexapods (although weakly supported). Those analyses, however, never included proturans. Also in recent studies, both Protura and Diplura (e.g., Timmermans et al. 2008; Aleshin et al. 2009), or at least Protura, are missing (Regier et al. 2008, 2010). Including these orders is indispensable to infer deep hexapod relationships. Our analyses based on much more extensive phylogenomic data, including all orders of monocondyl, primary wingless hexapods, yielded strong support for monophyletic hexapods. We conclude that hexapods are monophyletic and that the distinctive bauplan evolved only once.

Relationships among pterygote insects are still disputed. A puzzling problem is the early evolution of winged insects (Whitfield and Kjer 2008). Mayflies, dragonflies, and neopterous winged insects appear early in the fossil record. Morphological and molecular analyses support a clade (Odonata (Ephemeroptera + Neoptera)) coined "Chiastomyaria" (Boudreaux 1979; Kjer 2004), or "Metapterygota" (Ephemeroptera (Odonata + Neoptera)) (see Zhang et al. 2008; Börner 1909), or "Palaeoptera" ((Odonata + Ephemeroptera) Neoptera) (see Hennig 1981; Kukalová-Peck 1983). Most molecular analyses support either a "Chiastomyaria" or "Palaeoptera" clade (see discussion in Simon et al. 2009). A possible explanation for the difficult-to-resolve relationships is an "explosive radiation" once flight evolved (Whitfield and Kjer 2008). Our phylogenomic data are inconclusive in ML tree reconstructions, but strongly support "Palaeoptera" in Bayesian analyses. Convincing morphological synapomorphies for Paleoptera and Neoptera are lacking.

Within neopterous insects, relationships among endopterygote insects are a major focus of scientific activity. For example, it is unclear whether beetles + neuropteridans (Neuropteroidea) branch off first or whether hymenopterans are the sister group to all other endopterygote insects (Kristensen 1999; Kukalová-Peck and Lawrence 2004; Beutel and Pohl 2006; Wiegmann et al. 2009). Our analyses strongly support most orders of Endopterygota (figs. 1 and 2). Several of these clades corroborate previous results based on single nuclear genes (von Reumont et al. 2009). Our phylogenomic approach also unambiguously supports hymenopterans as the sister group to all other endopterygote insects and corroborates previous studies (e.g., Savard et al. 2006; Simon et al. 2009; Wiegmann et al. 2009) in contrast to conclusions based on complete mitochondrial genomes (Castro and Dowton 2005). This result will be extremely important in interpreting and understanding early extinct endopterygote insects and the evolution of this most species-rich group of arthropods.

Conclusions

We show that phylogenomic studies, although raising hope to reach a resolved arthropod tree, still face challenges in interpreting the strength and quality of the phylogenetic signal. We also illustrate unresolved incongruences between morphological and molecular analyses. This, in our opinion, should challenge systematists of every camp to present the strength, quality, and deficiencies of their evidence and work toward resolving outstanding issues.

Supplementary Material

The Supplementary Material containing Supplementary tables 1-2, Supplementary figures 1-6 and Supplementary literature are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjorurnals.org/).

Acknowledgments

We acknowledge Nikola Szucsich, Daniela Bartel, Johannes Dambach, Markus Pennerstorfer, Erich Eder, Francisco Javier Cristobo, Hilke Ruhberg, and Sara Khadjeh for sampling and delivering specimens or tissues. We are also grateful to Michael Kube and Richard Reinhardt (MPIMG, Berlin, Germany) for the construction and sequencing of cDNA libraries and for submitting assembled contigs. We thank Benjamin Meyer, Nikola Szucsich, Janus Borner, Roman Stocsits, and Stefan Grünewald for discussion and help with software applications and Alexandros Stamatakis for RAxML support. Special thanks go to Katharina Misof and Michael Stachowitsch for linguistic help. We also gratefully thank Barbara Holland and anonymous reviewers for providing constructive comments that improved this manuscript. The data matrices are available from Treebase (www.treebase.org) or from http://www.zfmk.de/web/Forschung/Molekularlabor/Datenstze/index.en.html. This is a collaboration of the "Arthropod Network" within the Deutsche Forschungsgemeinschaft (DFG) Priority Program 1174 "Deep Metazoan Phylogeny" (http://www.deep-phylogeny.org). B.M. and K.M. were supported by the DFG grant MI 649/6-2 and J.W.W., P.K., and B.M.v.R. were supported by grants WA530/33 and WA530/34. H.H. and S.Si. were funded by the DFG grant HA 1947/5 and F.R. and T.B. by the DFG grant Bu956/8. S.St. was granted by the DFG project HA 1628/8 and A.v.H. and I.E. were supported by the WWTF. G.P. was funded by

the Austrian Science Foundation (FWF) grant P 20497-B17 and the SuGI-project (S.B. and V.A.) was supported by the BMBF, Germany.

Author's Contributions

F.R. and T.B. delivered EST data for onychophorans, myriapods, pycnogonids, and euchelicerates. B.M.v.R. and J.W.W. supplied ESTs for all new crustacean species. K.M. and B.M. provided ESTs for four hexapod species and M.W. and G.P. enabled the proturan EST project to be conducted. S.Si. and H.H. delivered EST data for two pterygote insects. Processing of EST data and orthologous gene prediction were performed by S.St. and I.E. Alignment masking and new reduction heuristics were developed by B.M. K.M., B.M., and B.M.v.R. designed the study, and analyses were conducted by K.M., B.M.v.R., S.Si., and B.M. P.K. provided Perl-Scripts and was involved in development of reduction heuristics, S.B. and V.A. enabled all Bayesian analyses with technical and computational support. The manuscript was written by B.M., K.M., B.M.v.R., T.B., S.Si., and S.St. with comments and revisions from J.W.W., G.P., I.E., S.B., V.A., and A.v.H. All authors read and approved the final manuscript.

References

Aleshin VV, Mikhailov KV, Konstantinova AV, Nikitina MA, Rusinc LY, Buinova DA, Kedrova OS, Petrov NB. 2009. On the phylogenetic position of insects in the Pancrustacea clade. *Mol Biol (Mosk)*. 43:804–818.

Bäcker H, Fanenbruck M, Wägele JW. 2008. A forgotten homology supporting the monophyly of Tracheata: the subcoxa of insects and myriapods re-visited. *Zool Anz*. 247:185–207.

Beutel RG, Pohl H. 2006. Endopterygote systematics—where do we stand and what is the goal (Hexapoda, Arthropoda)? *Syst Entomol*. 31:202–219.

Bitsch J, Bitsch C. 2004. Phylogenetic relationships of basal hexapods among the mandibulate arthropods: a cladistic analysis based on comparative morphological characters. *Zool Scr*. 33:511–550.

Bleidorn C, Podsiadlowski L, Zhong M, Eeckhaut I, Hartmann S, Halanych KM, Tiedemann R. 2009. On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evol Biol*. 9:150.

Börner C. 1909. Neue Homologien zwischen Crustaceen und Hexapoden. Die Beissmandibel der Insecten und ihre phylogenetische Bedeutung. Archi- und Metapterygota. *Zool Anz*. 34:100–125.

Boudreaux BH. 1979. Arthropod phylogeny: with special reference to insects. New York: John Wiley and Sons Inc.

Bourlat SJ, Juliusdottir T, Lowe CJ, et al. (14 co-authors). 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444:85–88.

Bourlat SJ, Nielsen C, Economou AD, Telford MJ. 2008. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol Phylogenet Evol*. 49:23–31.

Brenneis G, Ungerer P, Scholtz G. 2008. The chelifores of sea spiders (Arthropoda, Pycnogonida) are the appendages of the deutocerebral segment. *Evol Dev*. 10:717–724.

Budd GE, Telford MJ. 2009. The origin and evolution of arthropods. *Nature* 457:812–817.

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

Carapelli A, Nardi F, Dallai R, Boore JL, Liò P, Frati F. 2005. Relationships between hexapods and crustaceans based on four mitochondrial genes. In: Koenemann S, Jenner RA, editors. Crustacean and arthropod relationships. Vol. 16, Crustacean Issues. Boca Raton (FL): *CRC Press*. p. 295–306.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552.

Castro LR, Dowton M. 2005. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). *Mol Phylogenet Evol*. 34:469–479.

Chou HH, Holmes MH. 2001. DNA sequence quality trimming and vector removal. *Bioinformatics* 17:1093–1104.

Dell'Ampio E, Szucsich NU, Carapelli A, Frati F, Steiner G, Steinacher A, Pass G. 2009. Testing for misleading effects in the phylogenetic reconstruction of ancient lineages of hexapods: influence of character dependence and character choice in analyses of 28S rRNA sequences. *Zool Scr*. 38:155–170.

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

Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H. 2008. Additional molecular support for the new chordate phylogeny. *Genesis* 46:592–604.

Dohle W. 2001. Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name "Tetraconata" for the monophyletic unit Crustacea + Hexapoda. Ann Soc Entomol Fr (New Series). 37:85–103.

Dunn CW, Hejnol A, Matus DQ, et al. (18 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.

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 GD. 2009. Palaeontological and molecular evidence linking arthropods, onychophorans, and other Ecdysozoa. *Evol Edu Outreach*. 2:178–190.

Ertas B, von Reumont BM, Wägele JW, Misof B, Burmester T. 2009. Hemocyanin suggests a close relationship of Remipedia and Hexapoda. *Mol Biol Evol*. 26:2711–2718.

Fanenbruck M, Harzsch S. 2005. A brain atlas of *Godzilliognomus frondosus* Yager, 1989 (Remipedia, Godzilliidae) and comparison with the brain of *Speleonectes tulumensis* Yager, 1987 (Remipedia, Speleonectidae): implications for arthropod relationships. *Arthropod Struct Dev*. 34:343–378.

Fanenbruck M, Harzsch S, Wägele JW. 2004. The brain of the Remipedia (Crustacea) and an alternative hypothesis on their phylogenetic relationships. *Proc Natl Acad Sci U S A*. 101:3868–3873.

Giribet G. 2003. Molecules, development and fossils in the study of metazoan evolution; Articulata versus Ecdysozoa revisited. *Zoology (Jena)* 106:303–326.

Grimaldi DA. 2010. 400 million years on six legs: on the origin and early evolution of Hexapoda. *Arthropod Struct Dev*. 39:191–203.

Hartmann S, Vision TJ. 2008. Using ESTs for phylogenomics: can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evol Biol*. 8:95.

Harzsch S. 2006. Neurophylogeny: architecture of the nervous system and a fresh view on arthropod phyologeny. *Integr Comp Biol*. 46:162–194.

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

Hejnol A, Obst M, Stamatakis A, et al. (17 co-authors). 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc Biol Sci*. 276:4261–4270.

Hennig W. 1981. Insect phylogeny. New York: John Wiley and Sons.

Heymonds R. 1901. Die Entwicklungsgeschichte der Scolopender. *Zoologica* 33:1–244, Taf. I–VIII.
Holland B, Moulton V. 2003. Consensus networks: a method for visualising incompatibilities in collections of trees. In: Workshop on Algorithms in Bioinformatics (WABI) 2812, Proceedings. Lecture Notes in *Bioinformatics* Budapest, Hungary. September 15–20, 2003. Berlin: Springer. p. 165–176.
Holmes DS, Bonner J. 1973. Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. *Biochemistry* 12:2330–2338.
Hou X, Bergström J. 1995. Cambrian lobopodians—ancestors of extant onychophorans? *Zool J Linn Soc*. 114:3–19.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 23:254–267.
Jager M, Murienne J, Clabaut C, Deutsch J, Le Guyader H, Manuel M. 2006. Homology of arthropod anterior appendages revealed by hox gene expression in a sea spider. *Nature* 441:506–508.
Kaas RE, Raftery AE. 1995. Bayes factors. J Am Stat. *Assoc*. 90: 773–795.
Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform*. 9:286–298.
Kjer KM. 2004. Aligned 18S and insect phylogeny. *Syst Biol*. 53: 506–514.
Kjer KM, Carle FL, Litman J, Ware J. 2006. A molecular phylogeny of Hexapoda. *Arthropod Syst Phylogeny*. 64:35–44.
Koenemann S, Olesen J, Alwes F, Iliffe TM, Hoenemann M, Ungerer P, Wolff C, Scholtz G. 2009. The post-embryonic development of Remipedia (Crustacea)—additional results and new insights. *Dev Genes Evol*. 219:131–145.
Koenemann S, Schram FR, Bloechl A, Iliffe T. 2007. Post-embryonic development of remipede crustaceans. *Evol Dev*. 9:117–121.
Kristensen NP. 1991. Phylogeny of extant hexapods. In: Naumann ID, Lawrence JF, Nielsen ES, Spradberry JP, Taylor RW, Whitten MJ, Littlejohn MJ, editors. The insects of Australia: a textbook for students and research workers. Melbourne (Australia): CSIRO, Melbourne University Press.
Kristensen NP. 1999. Phylogeny of endopterygote insects, the most successful lineage of living organisms. *Eur J Entomol*. 96: 237–253.
Kück P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wägele JW, 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.
Kukalová-Peck J. 1983. Origin of the insect wing and wing articulation from the arthropodan leg. *Can J Zool*. 61:1618–1669.
Kukalová-Peck J, Lawrence JF. 2004. Relationships among coleopteran suborders and major endoneopteran lineages: evidence from hind wing characters. *Eur J Entomol*. 101:95–144.
Lartillot N, Blanquart S, Lepage T. 2008. PhyloBayes 2.3—a Bayesian software for phylogenetic reconstruction using mixture models. Available from: http://www.phylobayes.org [2.3c current versions]. University of Montreal.
Lartillot N, 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, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos Trans R Soc Lond B Biol Sci*. 363:1463–1472.
Luan Yx, Mallatt JM, Xie Rd, Yang Ym, Yin Wy. 2005. The phylogenetic positions of three basal-hexapod groups (Protura, Diplura, and Collembola) based on on ribosomal RNA gene sequences. *Mol Biol Evol*. 22:1579–1592.
Mallatt J, Craig CW, Yoder MJ. 2010. Nearly complete rRNA genes assembled from across the metazoan animals: effects of more taxa, a structure-based alignment, and paired-sites evolutionary models on phylogeny reconstruction. *Mol Phylogenet Evol*. 55:1–17.
Mallatt J, Giribet G. 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. *Mol Phylogenet Evol*. 40:772–794.
Mallatt JM, Garey JR, Shultz JW. 2004. Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Mol Phylogenet Evol*. 31:178–191.
Maxmen A, Browne WE, Martindale MQ, Giribet G. 2005. Neuroanatomy of sea spiders implies an appendicular origin of the protocerebral segment. *Nature* 437:1144–1148.
Mayer G, Whintington PM. 2009. Velvet worm development links myriapods with chelicerates. *Proc Biol Sci*. 276:3571–3579.
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.
Misof B, Niehuis O, Bischoff I, Rickert A, Erpenbeck D, Staniczek A. 2007. Towards an 18S phylogeny of hexapods: accounting for group-specific character covariance in optimized mixed nucleotide/doublet models. *Zoology (Jena)* 110:409–429.
Müller CHG, Sombke A, Rosenberg J. 2007. The fine structure of the eyes of some bristly millipedes (Penicillata, Diplopoda): additional support for the homology of mandibulate ommatidia. *Arthropod Struct Dev*. 36:463–476.
Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F. 2003. Hexapod origins: monophyletic or paraphyletic? *Science* 299: 1887–1889.
Nieselt-Struwe K, von Haeseler A. 2001. Quartet-mapping, a generalization of the likelihood-mapping procedure. *Mol Biol Evol*. 18:1204–1219.
Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL. 2004. Bayesian phylogenetic analysis of combined data. *Syst Biol*. 53: 47–67.
Ott M, Zola J, Aluru S, Stamatakis A. 2007 Large-scale maximum likelihood-based phylogenetic analyses on the IBM BlueGene/L. In: Proceedings of ACM/IEEE Supercomputing Conference 2007. Reno, NV. ACM, New York, NY.
Pertea G, Huang X, Liang F, et al. (12 co-authors). 2003. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19:651–652.
Philippe H, Derelle R, Lopez P, et al. (20 co-authors). 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol*. 19:706–712.
Pisani D, Poling L, Lyons-Weiler M, Hedges SB. 2004. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. *BMC Biol*. 2:1.
Pocock RI. 1893. On the classification of the tracheate Arthropoda. *Zool Anz*. 16:271–275.
Regier JC, Shultz JW, Ganley ARD, et al. (11 co-authors). 2008. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Syst Biol*. 57:920–938.
Regier JC, Shultz JW, Kambic RE. 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proc Biol Sci*. 272:395–401.
Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Atzhropod relationships revealed by phylogenomic analysis of nuclear protein coding sequences. *Nature* 463:1079–1083.
Richter S. 2002. The Tetraconata concept: hexapod-crustacean relationships and the phylogeny of Crustacea. *Org Divers Evol*. 2: 217–237.
Roeding F, Hagner-Holler S, Ruhberg H, Ebersberger I, von Haeseler Arndt, Kube M, Reinhardt R, Burmester T. 2007. EST sequencing of Onychophora and phylogenomic analysis of Metazoa. *Mol Phylogenet Evol*. 45:942–951.
Sanderson MJ. 2007. Construction and annotation of large phylogenetic trees. *Aust Syst Bot*. 20:287–301.

Savard J, Tautz D, Richards S, Weinstock GM, Gibbs RA, Werren JH, Tettelin H, Lercher MJ. 2006. Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of holometabolous insects. *Genome Res*. 16:1334–1338.

Scholtz G, Edgecombe G. 2006. The evolution of arthropod heads: reconciling morphological, developmental and palaeontological evidence. *Dev Genes Evol*. 216:395–415.

Simon S, Strauss S, von Haeseler A, Hadrys H. 2009. A phylogenomic approach to resolve the basal pterygote divergence. *Mol Biol Evol*. 26:2719–2730.

Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24:715–716.

Stamatakis A. 2006a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. In: 20th International Parallel and Distributed Processing Symposium (IPDPS 2006), Proceedings; 2006 April 25–29; Rhodes Island, Greece. ACM, New York.

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

Stillman JH, Colbourne JK, Lee CE, et al. (12 co-authors). 2008. Recent advances in crustacean genomics. *Integr Comp Biol*. 48: 852–868.

Szucsich NU, Pass G. 2008. Incongruent phylogenetic hypotheses and character conflicts in morphology: the root and early branches of the hexapodan tree. *Mitt Dtsch Ges Allg Angew*. 16:415–429.

Telford MJ, Bourlat SJ, Economou A, Papillon D, Rota-Stabelli O. 2008. The evolution of the Ecdysozoa. *Philos Trans R Soc Lond B Biol Sci*. 363:1529–1537.

Thorley JL, Wilkinson M. 1999. Testing the phylogenetic stability of early tetrapods. *J Theor Biol*. 200:343–344.

Timmermans MJ, D Roelofs D, Mariën J, van Straalen NM. 2008. Revealing pancrustacean relationships: phylogenetic analysis of ribosomal protein genes places Collembola (springtails) in a monophyletic Hexapoda and reinforces the discrepancy between mitochondrial and nuclear DNA markers. *BMC Evol Biol*. 8:83.

Ungerer P, Scholtz G. 2008. Filling the gap between identified neuroblasts and neurons in crustaceans adds new support for Tetraconata. *Proc Biol Sci*. 275:369–376.

von Reumont BM, Meusemann K, Szucsich NU, et al. (13 co-authors). 2009. Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. *BMC Evol Biol*. 9:119.

Wägele JW. 1993. Rejection of the 'Uniramia' hypothesis and implications of the Mandibulata concept. *Zoologische Jahrbücher. Abteilung für Systematik, Ökologie und Geographie der Tiere*. 120:253–288.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18:691–699.

Whitfield JB, Kjer KM. 2008. Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annu Rev Entomol*. 53:449–472.

Wiegmann B, Trautwein M, Kim JW, Cassel B, Bertone M, Winterton S, Yeates D. 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol*. 7:34.

Wiens JJ, Moen DS. 2008. Missing data and the accuracy of Bayesian phylogenetics. *J Syst Evol*. 46:307–314.

Zantke J, Wolff C, Scholtz G. 2008. Three-dimensional reconstruction of the central nervous system of *Macrobiotus hufelandi* (Eutardigrada, Parachela): implications for the phylogenetic position of Tardigrada. *Zoomorphology* 127:21–36.

Zhang J, Zhou C, Gai Y, Song D, Zhou K. 2008. The complete mitochondrial genome of *Parafronurus youi* (Insecta: Ephemeroptera) and phylogenetic position of the Ephemeroptera. *Gene* 424:18–24.

Zrzavý J, Mihulka S, Kepka P, Tietz D. 1998. Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics* 14:249–285.

Zrzavý J, Štys P. 1997. The basic body plan of arthropods: insights from evolutionary morphology and developmental biology. *J Evol Biol*. 10:653–367.