Arthropod Structure & Development 39 (2010) 88–110

Contents lists available at ScienceDirect

Arthropod Structure & Development

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

Arthropod phylogeny revisited, with a focus on crustacean relationships

Stefan Koenemann [a,*], Ronald A. Jenner [b,**], Mario Hoenemann [a], Torben Stemme [a], Björn M. von Reumont [c,***]

[a] Institute for Animal Ecology and Cell Biology, University of Veterinary Medicine Hannover, Bünteweg 17d, D-30559 Hannover, Germany
[b] Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK
[c] Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

ARTICLE INFO

Article history:
Received 10 June 2009
Accepted 14 October 2009

Keywords:
18S rDNA
16S rDNA
Cytochrome *c* oxidase subunit I
Structural alignment
Multiple alignment optimization
Sensitivity analysis

ABSTRACT

Higher-level arthropod phylogenetics is an intensely active field of research, not least as a result of the hegemony of molecular data. However, not all areas of arthropod phylogenetics have so far received equal attention. The application of molecular data to infer a comprehensive phylogeny of Crustacea is still in its infancy, and several emerging results are conspicuously at odds with morphology-based studies. In this study, we present a series of molecular phylogenetic analyses of 88 arthropods, including 57 crustaceans, representing all the major lineages, with Onychophora and Tardigrada as outgroups. Our analyses are based on published and new sequences for two mitochondrial markers, 16S rDNA and cytochrome *c* oxidase subunit I (COI), and the nuclear ribosomal gene 18S rDNA. We designed our phylogenetic analyses to assess the effects of different strategies of sequence alignment, alignment masking, nucleotide coding, and model settings. Our comparisons show that alignment optimization of ribosomal markers based on secondary structure information can have a radical impact on phylogenetic reconstruction. Trees based on optimized alignments recover monophyletic Arthropoda (excluding Onychophora), Pancrustacea, Malacostraca, Insecta, Myriapoda and Chelicerata, while Maxillopoda and Hexapoda emerge as paraphyletic groups. Our results are unable to resolve the highest-level relationships within Arthropoda, and none of our trees supports the monophyly of Myriochelata or Mandibulata. We discuss our results in the context of both the methodological variations between different analyses, and of recently proposed phylogenetic hypotheses. This article offers a preliminary attempt to incorporate the large diversity of crustaceans into a single molecular phylogenetic analysis, assessing the robustness of phylogenetic relationships under varying analysis parameters. It throws into sharp relief the relative strengths and shortcomings of the combined molecular data for assessing this challenging phylogenetic problem, and thereby provides useful pointers for future studies.

1. Introduction

One of the persistent challenges in systematic biology concerns the phylogenetic relationships of Arthropoda (here defined as all extant arthropods and their stem groups, but excluding onychophorans and tardigrades; Panarthropoda refers to the grouping of Arthropoda, Tardigrada and Onychophora). The systematic literature on higher-level relationships within arthropods dwarfs that of any metazoan taxon, with the possible exception of vertebrates. The phylogenetic relationships of the five major traditional groups, Hexapoda, Myriapoda, Crustacea, Chelicerata, and the extinct Trilobitomorpha, has remained a matter of debate since the 19th century (e.g., Latreille, 1817; Pocock, 1893a,b; Lankester, 1904). Although debates about arthropod phylogeny have long been framed in terms of morphological and developmental evidence, current activities show an additional strong focus on molecular data, derived from both mitochondrial and nuclear sources. In this article, we present a series of molecular phylogenetic analyses of arthropod phylogeny based on both published and new sequence data from three loci: 18S rDNA, 16S rDNA, and cytochrome *c* oxidase I (COI). We discuss our results with respect to recent phylogenetic analyses based on molecular and morphological evidence. The specific focus of our analyses is the relationships between the major lineages of crustaceans, which represent over half of the species in our data set. To our knowledge our data set includes the largest sample of crustacean diversity yet analyzed in a single molecular phylogenetic analysis.

* Corresponding author. Tel.: +49 511 953 8881.
** Corresponding author. Tel.: +44 207 942 6885.
*** Corresponding author. Tel.: +49 228 912 2351.
E-mail addresses: stefan.koenemann@tiho-hannover.de (S. Koenemann), r.jenner@nhm.ac.uk (R.A. Jenner), bmvr@arcor.de (B.M. von Reumont).

1467-8039/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.asd.2009.10.003

1.1. The modern fate of early phylogenetic concepts

Several of the familiar higher-level groupings, such as Atelocerata, Tracheata, Uniramia, and Mandibulata, have their origins as far back as the 19th century. For example, Haeckel (1866) erected Tracheata, to which he assigned all arthropods with tracheal breathing, the arachnids, myriapods and insects. Tracheata was redefined by Pocock (1893a,b), who excluded the arachnids. Pocock furthermore considered Myriapoda "an unnatural assemblage of beings", composed of diplopods/pauropods and chilopods/hexapods as the two most closely related groups, and symphylans in an unassigned position ("a question for future discussion"). Based on a detailed comparison of metameric structures, Heymons (1901) continued to support myriapods and hexapods as sister groups, and proposed to unite them under the new name Atelocerata. Today, both concepts, Tracheata and Atelocerata, are usually used as synonyms. Interestingly, in the phylogenetic analysis of combined molecular and morphological evidence of Wheeler et al. (2004), a monophyletic Atelocerata is supported depending on whether selected fossils are included in the analysis. Molecular analyses do not find support for Atelocerata, instead uniting Crustacea and Hexapoda as Pancrustacea (e.g., Regier and Shultz, 1997) or Tetraconata (Dohle, 2001).

Some early hypotheses about the evolutionary relationships of arthropods included other segmented animals, such as onychophorans, as basal arthropods, from which modern, extant forms were believed to have been derived (e.g., Snodgrass, 1935, 1938). Manton (1973) went a step further and proposed the taxon Uniramia to embrace hexapods, myriapods, and onychophorans, three groups characterized by segmented trunks, single-branch limbs, one pair of (first) antennae, and reduced post-oral mouthparts. She considered Crustacea, Chelicerata, and Trilobita to be separate groups from each other and from Uniramia. However, despite apparent support from neuroanatomical data for including Onychophora within Arthropoda (Strausfeld et al., 2006), the Uniramia hypothesis is now generally considered obsolete (see Wägele, 1993). Nevertheless, a recent molecular phylogenetic analysis (Colgan et al., 2008) places Onychophora within Arthropoda, and several phylogenomic analyses (Roeding et al., 2007; Marletaz et al., 2008) place them as a sister group to Chelicerata.

Another early concept of a major arthropod clade goes back to Snodgrass (1935), who erected Mandibulata as a taxon encompassing Crustacea and Atelocerata, groups that both share, in particular, the possession of distinctly shaped mandibles. Although the monophyly of Mandibulata is generally supported by morphological evidence (Vaccari et al., 2004; Wheeler et al., 2004; Giribet et al., 2005), which contradicts the Schizoramia hypothesis that groups chelicerates and crustaceans (Cisne, 1974), it has recently come under fire from molecular phylogenetic analyses that instead united Myriapoda and Chelicerata as a clade Paradoxopoda (or Myriochelata) (Rota-Stabelli and Telford, 2008). However, it is possible that support for Paradoxopoda from mitochondrial evidence is an artifact of outgroup choice (Rota-Stabelli and Telford, 2008), but analyses based on nuclear sequence data may support either Mandibulata or Paradoxopoda (Bourlat et al., 2008; Dunn et al., 2008; Regier et al., 2008).

Although molecular evidence has become a crucial source of data, comparative morphology retains an important role in systematizing both extant and fossil panarthropods. The study of Wheeler et al. (2004) is emblematic for the importance of morphology, especially in showing the power of fossils to influence relationships among extant taxa. This study showed that the inclusion of just a small number of fossil taxa can significantly change the relationships of the major arthropod taxa (alternatively supporting Atelocerata or Pancrustacea) based on morphological or combined molecular and morphological evidence. Our current understanding of the phylogenetic positions and evolution of extinct panarthropod lineages is of course wholly dependent on the deployment of morphological data (Vaccari et al., 2004; Cobbett et al., 2007). Not least, excellent morphological work on fossils has allowed unique insights into the composition of stem-lineages that underpin the extant crown groups of panarthropods (e.g., Walossek and Müller, 1990; Walossek and Müller, 1998; Walossek, 1993; Budd, 1996; Edgecombe, 2004).

1.2. Modern debates and molecular evidence

The application of diverse and increasingly abundant molecular evidence to the problem of higher-level arthropod phylogeny is currently gathering steam on several fronts. First, commonly used nuclear and mitochondrial loci, including 18S rDNA, 28S rDNA, 16S rDNA, COI, elongation factor 1-α, and RNA polymerase II are sequenced for increasing numbers of species across all major extant panarthropod taxa. Second, studies that go beyond these "usual suspects" have started to add valuable independent light on arthropod relationships (Regier et al., 2005, 2008). Third, the development of increasingly sophisticated and powerful sequencing and computational techniques, and the rapidly falling prices of large-scale sequencing will soon take arthropod phylogenetics to the same level as higher-level metazoan phylogenetics. Our study contributes to the first category by analyzing potential phylogenetic signal in a combined data set of three oft-used loci (18S rDNA, 16S rDNA, COI), and goes beyond previous efforts by (1) an increased sampling of taxa within Crustacea, (2) the use of newly developed software to improve the quality of multiple sequence alignments, and (3) the performance of sensitivity analyses to explore the effect of differences in sequence alignment on the phylogenetic results.

A large number of molecular phylogenetic analyses of major arthropod relationships (some also including morphological data) has been published, but despite some emerging consensus many unresolved issues remain (e.g., Giribet, et al., 1996; Fortey and Thomas, 1997; Wheeler, 1997; Zrzavý et al., 1997; Giribet et al., 2001, 2004, 2005; Hwang et al., 2001; Nardi et al., 2003; Regier et al., 2005, 2008; Glenner et al., 2006). As pointed out by Regier et al. (2008), deep arthropod phylogeny shares many of the problems that plague deep metazoan phylogenetics. The original phylogenetic signal has deteriorated significantly over the hundreds of millions of years of independent evolution separating the major taxa, and as data density grows, systematic errors become apparent, making results sensitive to choice of method and data treatment. The studies of Regier et al. (2008) and Reumont et al. (2009) provide clear illustrations of the difficulties involved. Both studies demonstrate that ignoring time-heterogeneous substitution processes in protein data (Regier et al., 2008) or heterogeneous base composition in rRNA data (Reumont et al., 2009) can mislead phylogenetic reconstructions. Studies also show some striking conflicts between mitochondrial and nuclear data, for example with respect to the monophyly of Hexapoda. Consequently, the same recommendations made for future studies of metazoan phylogenetics can be made for higher-level arthropod phylogenetics (e.g. Jenner and Littlewood, 2008), acknowledging that much still needs to be done.

Although a universal consensus remains elusive in this dynamic research area, a provisional consensus can nevertheless be diagnosed in reference to the most recent comprehensive studies (Giribet et al., 2005; Wheeler et al., 2004; Regier et al., 2005, 2008; Bourlat et al., 2008; Timmermans et al., 2008; Budd and Telford, 2009). Arthropoda is monophyletic and comprises at least four extant clades: Pycnogonida, Chelicerata, Pancrustacea (hexapods

and crustaceans), and Myriapoda. The monophyly of Pycnogonida and Chelicerata is well established, whereas the monophyly of Pancrustacea is increasingly well-supported on the basis of molecular evidence. In contrast, the monophyly of Hexapoda and Myriapoda is less certain. Phylogenetic analyses based on mitochondrial sequences have repeatedly questioned hexapod monophyly, suggesting that collembolans do not group with the remaining hexapods. Nevertheless, both hexapod and myriapod monophyly are generally supported in the most comprehensive analyses. Crustacea may be para- or even polyphyletic (Schram and Koenemann, 2004a,b; Regier et al., 2008). We have included Tardigrada and Onychophora as arthropod outgroups, but it should be noted that some phylogenetic analyses in a recent study (Colgan et al., 2008) place both Tardigrada and Onychophora within Arthropoda, and several phylogenomic analyses (Roeding et al., 2007; Marletaz et al., 2008) place Onychophora as a sister group to Chelicerata. However, the inclusion of Tardigrada and Onychophora within Arthropoda in Colgan et al. (2008) is sensitive to method of analysis and data selection, and is robustly contradicted by other molecular phylogenetic analyses (Mallatt and Giribet, 2006; Dunn et al., 2008; Podsiadlowski et al., 2008 for onychophorans; Paps et al., 2009 for tardigrades). The position of Onychophora as a sister group to Chelicerata in Roeding et al. (2007) and Marletaz et al. (2008) may very well be influenced by the absence of Myriapoda in these analyses, and needs further testing.

The phylogenetic relationships of these primary arthropod lineages remain to be established in detail, as do the relationships within these taxa. The most prominent questions that our study aims to address are the following:

- The position of Pycnogonida inside or outside Chelicerata (e.g., Park et al., 2007; reviewed in Dunlop and Arango, 2005);
- Mandibulata vs. Pancrustacea + Paradoxopoda (Myriochelata) (Rota-Stabelli and Telford, 2008; Reumont et al., 2009);
- Monophyly and relationships of Hexapoda and Crustacea within Pancrustacea (e.g., Nardi et al., 2003; Cameron et al., 2004; Cook et al., 2005; Carapelli et al., 2005, 2007 vs. Timmermans et al., 2008);
- The monophyly of Atelocerata (Wheeler et al., 2004);
- The relationships within Crustacea, principally the interrelationships of the major recognized lineages (Martin and Davis, 2001): Maxillopoda, Branchiopoda, Malacostraca, Ostracoda, Remipedia, and Cephalocarida, and to a lesser extent the monophyly and internal relationships of some of these presumed clades.

Although our species sampling allows us to test phylogenetic relationships within the main lineages of Hexapoda, Myriapoda, and Chelicerata, we note that these are included primarily to function as outgroups (and possibly ingroups) to Crustacea.

1.3. Crustacean phylogeny

Progress in resolving phylogenetic relationships is not equal across the major extant arthropod taxa. A conspicuous relative lack of both attention and progress in understanding higher-level phylogenetic relationships characterizes Crustacea compared to hexapods, chelicerates and myriapods. Phylogenetic hypotheses about the evolution of the unparalleled morphological disparity of the major crustacean groups are still chiefly based on morphological evidence (e.g., Dahl, 1963; Schram, 1986; Wilson, 1992; Wills, 1997; Schram and Hof, 1998; Schram and Koenemann, 2004b), with little detailed consensus (Jenner, 2010). Higher-level crustacean molecular phylogenetics was effectively jumpstarted in the late 1980s and 1990s by Lawrence Abele, Trisha Spears and colleagues. In a series of seminal papers they explored crustacean phylogeny based on ribosomal gene sequences, seeding a growing literature. However, so far no comprehensive molecular phylogeny that includes most major taxa has been performed.

Martin and Davis (2001) recognized six major groups of Crustacea: Malacostraca, Branchiopoda, Maxillopoda, Ostracoda, Remipedia, and Cephalocarida. To date the most thorough and comprehensive higher-level phylogenetic analyses within Crustacea using molecular evidence focus on Branchiopoda (Braband et al., 2002; deWaard et al., 2006; Stenderup et al., 2006; Richter et al., 2007) and Malacostraca (Spears et al., 2005; Meland and Willassen, 2007; Jenner et al., 2009). These and larger-scale studies support the monophyly of Branchiopoda and Malacostraca. Remipedia and Cephalocarida are both considered monophyletic (Martin and Davis, 2001; Koenemann et al., 2007); however, their phylogenetic positions remain unknown (Jenner, 2010). Although Ostracoda is traditionally considered monophyletic (Martin and Davis, 2001), consistent with a recent morphological phylogenetic analysis (Horne et al., 2005), molecular evidence instead unites podocopid ostracodes more closely with branchiurans (and possibly pentastomids) than with myodocopids (Spears and Abele, 1997; Regier et al., 2005, 2008). The monophyly of Maxillopoda seems increasingly doubtful. Although maxillopodan monophyly is suggested on the basis of some morphological evidence (Wills, 1997; Ax, 1999), other morphological studies disagree (Schram and Koenemann, 2004b), and molecular evidence contradicts maxillopodan monophyly (Spears and Abele, 1997; Regier et al., 2005, 2008). However, although various studies include samples of maxillopodan taxa, so far no broadly sampled molecular maxillopodan phylogeny is available.

There is evidence that Crustacea s. str. may represent a para- or even polyphyletic assemblage of arthropods, and the concept of a hexapod-crustacean clade, Pancrustacea or Tetraconata, has been proposed independently in a number of studies (e.g., Regier and Shultz, 1997; Spears and Abele, 1997; Zrzavý and Štys, 1997; García-Machado et al., 1999; Lavrov et al., 2004; Schram and Koenemann, 2004b; Cook et al., 2005; Regier et al., 2008; Reumont et al., 2009). With respect to extant taxa this means that hexapods fall within Crustacea, although it remains unclear to which extant crustacean taxon they would be most closely related (Jenner, 2010).

1.4. Strengths and limitations of the present analysis

A conspicuous feature of published molecular phylogenetic analyses of higher-level arthropod relationships, including Crustacea, is that the results are often strongly sensitive to analysis parameters such as the choice of loci and taxa, method of sequence alignment, method of phylogenetic analysis, and choice of evolutionary model in model-based phylogenetic analyses. In our analysis, we include the largest sample of crustacean diversity in a single molecular phylogenetic analysis to date. In order to find a balance between species sampling and data density, we base our analysis on available and newly generated sequence data for three loci (18S rDNA, 16S rDNA, and COI). The results can serve as a baseline for comparisons with future studies, and provide a test of available hypotheses (see Jenner, 2010). Although skepticism exists about the utility for deeper phylogenetic levels of especially the relatively fast-evolving mitochondrial loci, we agree with Cameron et al. (2004) that no convincing arguments exist for *a priori* exclusion of individual mitochondrial loci from phylogenetic analyses of higher-level arthropod relationships. Mitochondrial data by itself may indeed be insufficient (due to saturation and accumulated noise and non-phylogenetic signals) to robustly resolve such relationships,

but in view of positive clade contributions of mitochondrial loci to such higher-level analyses (Cameron et al., 2004; Jenner et al., 2009), there is little reason to exclude them. This is not to say that the inclusion of mitochondrial data is necessarily without problems, as it has been shown for several taxa that mitochondrial evidence may conflict with nuclear and/or morphological data (Cameron et al., 2004; Hassanin, 2006; Kjer and Honeycutt, 2007; Rota-Stabelli and Telford, 2008; Timmermans et al., 2008). Problematical issues concern, for example, the choice of outgroup taxa, heterogeneity of base composition and rates, and patterns of substitution among sites and taxa. Despite the possibility of such problems, our study is aimed to empirically assess the performance of the selected combined loci, and to determine which phylogenetic problems are in most urgent need of new and different data.

We performed a series of phylogenetic analyses, varying the methods of sequence alignment, alignment refinement, exclusion of ambiguously aligned regions or those with randomly similar sequences (alignment masking), phylogenetic inference method, and model settings. Although this represents a far from exhaustive set of possible treatments of our data, these analysis variables are known to be important in determining phylogenetic results. Consequently we believe that the combined results provide an informative summary of which hypotheses are reasonably supported by these data, and what areas are most in need of further attention.

2. Material and methods

2.1. Taxon sampling and choice of molecular markers

We selected representatives of relevant extant groups of crustaceans, insects, myriapods, and chelicerates to evaluate sister group relationships of the major arthropod lineages. Our taxon sample includes 88 terminal taxa representing all major groups of Crustacea (57 taxa), Hexapoda (13 taxa), Myriapoda (5 taxa), Chelicerata (11 taxa), Onychophora (1 taxon), and Tardigrada as an outgroup (Appendix A, Table A1). In view of recent suggestions based on neuroanatomical and phylogenomic evidence that onychophorans may be an arthropod ingroup, we designated only tardigrades as the outgroup, which allows us to test the phylogenetic position of the onychophorans.

One central objective for this analysis was an evaluation of alignment methods, in particular for ribosomal genes. We included both previously published and new sequences for three loci: 18S rDNA, 16S rDNA, and cytochrome *c* oxidase I (COI).

However, our desire for comprehensive taxon sampling across the major arthropod groups necessitated a trade-off regarding the choice of genetic markers. The genes of our preferred choice were not available for all of the taxa we selected. Therefore, we decided to tolerate incomplete gene sequences and even missing markers for some taxa. In order to maximize data density per taxon, we constructed composite (chimerical) higher-level terminal units in several cases by combining gene sequences of closely related taxa (see Appendix A). We argue that this strategy should not distort phylogenetic analyses, provided the composite taxa are monophyletic with respect to the other terminal taxa (Springer et al., 2004). Given the relatively distant relationships between the included terminals, this assumption appears justified. In our phylogenetic trees, chimerical taxa are named after the next available or an unambiguous higher rank, for example, *Hypochilus thorelli* + *H. pococki* = *Hypochilus*. The only exceptions are the two outgroup taxa that were named Onychophora and Tardigrada as a matter of convenience (see Appendix A).

2.2. Laboratory work

New DNA extractions and generation of new sequences were performed both at the Zoologisches Forschungsmuseum Alexander Koenig in Bonn and the University of Veterinary Medicine Hannover. The tissues of collected species were preserved in 94–99% ethanol or RNAlater and stored at −20 °C. DNA extraction of complete specimens or muscle tissue followed the standard protocols of the different manufacturers. For *Pleomothra apletocheles*, DNA was extracted from 15 mg pleonal muscle tissue using the Qiagen Mini Kit. For the other species muscle tissue or complete specimens were extracted using the DNeasy Blood & Tissue Kit (Qiagen) or the NucleoSpin Tissue Kit (Machery-Nagel) following the manufacturer's protocols. For the specimens processed in Bonn the incubation procedure was slightly modified. The samples were incubated overnight; before proceeding with extraction, 8 μl RNAse (10 mg/ml) was added for 10 min. Different primer sets were used for each gene for polymerase chain reaction (PCR) and cycle sequencing (Table 1).

PCR and cycle sequencing conditions differed slightly between the laboratories in Hannover and Bonn. For details see electronic supplementary files. PCR products were purified using the following kits: Nucleospin ExtractionII (Machery Nagel) and QIAquick purification Kit (Qiagen). Cycle sequencing took place on different thermocyclers and sequencers, and some samples were sequenced by Macrogen. Cycle sequencing reactions were carried out on both strands. The resulting electropherograms were checked and assembled using the software module SeqMan (Lasergene, DNA Star).

2.3. Alignments and data evaluation prior to tree reconstruction

Prior to alignment, we carried out BLASTN and MEGABLAST (Altschul et al., 1997) searches for each sequence, including both newly generated and published (GenBank) sequences, to identify possible contamination. Ambiguous sequences were excluded from the analyses. In addition, we verified that the COI data did not contain any nuclear copies of mitochondrial-derived genes (numts; see Buhay, 2009). For two terminal taxa, there were multiple 18S sequences available that differed conspicuously in the standard (MUSCLE) alignment. Since it was not possible to unambiguously identify the "correct" sequence in the standard alignment, we decided to include both 18S sequences for these two taxa, the mystacocarid *Derocheilocaris typica* and the symphylan *Scutigerella causeyae*; both species are represented as doubled terminal taxa (see also below).

One focus of this study was the influence of multiple sequence alignment methods on phylogenetic analysis. Consequently, we conducted a series of analyses to determine the effects of different combinations of these variables on our data set (see Table 2 for an overview). These included:

1 Alternative methods of multiple sequence alignment using either MUSCLE or MAFFT
2 Alignment methods based on secondary structure information
3 Identification and removal of ambiguously aligned and randomly similar regions (alignment masking)
4 RY-coding for the mitochondrial marker COI and the loop regions of 16S rDNA to correct for saturation effects
5 Model settings in MrBayes (Huelsenbeck and Ronquist, 2001)

We conducted an extensive set of pretests on a preliminary data set in order to assess the effects of varying the above analysis parameters. This allowed us to determine which experimental manipulations to perform on our final data set (Table 2), the results

Table 1
List of PCR and cycle sequencing (CS) primers for the three molecular markers used in this study.

Marker	Primer name	Reaction	Primer sequence (in 5′–3′ direction)	Direction	Source
16S rDNA	mt16S-ar	PCR & CS	cgc ctg ttt atc aaa aac at	forward	Palumbi, 1996
16S rDNA	mt16Sbr	PCR & CS	ccg gtc tga act cag atc acg t	reverse	Palumbi, 1996
16S rDNA	16Sa	PCR & CS	cgc ctg ttt atc aaa aac at	forward	Palumbi, 1996
16S rDNA	16Sb	PCR & CS	ccg gtc tga act cag atc acg	reverse	Palumbi, 1996, modified
16S rDNA	LRJ12887	PCR & CS	ccg gtc tga act cag atc acg t	forward	Simon et al., 1994
16S rDNA	LRN13398	PCR & CS	cgc ctg ttt aac aaa aac at	reverse	Simon et al., 1994
18S rDNA	18SfwSS1	PCR & CS	ggt tga tcc tgc cag taa ttg tat gct	forward	Schalla, unpubl.
18S rDNA	329	PCR & CS	taa tga tcc ttc cgc agg ttc acc tac gg	reverse	Trisha Spears, pers. comm.
18S rDNA	18A1	PCR & CS	ctg gtt gat cct gcc agt cat atg c	forward	Dreyer and Wägele, 2001
18S rDNA	1800	PCR & CS	gat cct tcc gca ggt ttca cct acg	reverse	Dreyer and Wägele, 2001
18S rDNA	700 F-MR	CS	gcc gcg gta att cca gc	forward	Raupach, unpubl.
18S rDNA	700R	CS	cgc ggc tgc tgg cac cag ac	reverse	Dreyer and Wägele, 2001
18S rDNA	1000F	CS	cga tca gat acc gcc cta gtt c	forward	Dreyer and Wägele, 2001
18S rDNA	1155R	CS	ccg tca att cct tta agt ttc ag	reverse	Dreyer and Wägele, 2001
18S rDNA	1250 FN-MR	CS	ggc cgt tct tag ttg gtg gag	forward	Raupach, unpubl.
18S rDNA	1500R	CS	cat cta ggg cat cac aga cc	reverse	Wollscheid et al., unpubl.
COI	HCO (+vector)	PCR & CS	taa tac gac tca cta tag ggt aaa ctt cag ggt gac caa aaa atc a	forward	Folmer et al., 1994
COI	LCO (+vector)	PCR & CS	att tag gtg aca cta tag aat ggt caa caa atc ata aag ata ttg	reverse	Folmer et al., 1994
COI	HCO	PCR & CS	taa act tca ggg tga cca aaa aat ca	forward	Folmer et al., 1994
COI	LCO	PCR & CS	ggt caa caa atc ata aag ata ttg g	reverse	Folmer et al., 1994

of which are presented in this article. In our pretests, we tested RY-coding for the COI sequences and for the 16S rDNA loop regions to counteract the effects of saturation and inhomogeneous base composition. RY-coding was originally used to assign third codon positions of protein-coding mitochondrial genes to one of two categories, purines (R) or pyrimidines (Y) (Phillips and Penny, 2003). We applied RY-coding to all alignment positions, and generated an improved model likelihood LnL. RY-coding was used for Runs 1–3.

Sequence pre-alignments were performed with the alignment programs MUSCLE (Edgar, 2004), and MAFFT (Katoh et al., 2002) using the linsi algorithm. Tests of MAFFT have indicated that its algorithms are more reliable for rDNA genes represented by expansion segments and ambiguous regions with variable length polymorphisms (Katoh and Toh, 2008).

The pre-alignments for Runs 1–3 were realigned according to information on the secondary structure of the 18S and 16S sequences using RNAsalsa, a new program for aligning rDNA sequences that implements information of secondary structures. RNAsalsa contains a constraint-guided thermodynamic folding algorithm and comparative evidence methods (for the exact algorithm and software download see the homepage at: http://rnasalsa.zfmk.de). We used constraints for 16S and 18S sequences of *Anopheles*. RNAsalsa automatically generates a secondary consensus structure (in dot bracket form) for the data set that can be used for mixed model analyses.

These optimized alignments were then scanned for the presence of random sequence similarity with the program ALISCORE (Misof and Misof, 2009; http://aliscore.zfmk.de) for Runs 1–3. ALISCORE uses a sliding window approach to generate profiles of randomness. Sequence positions within this sliding window are assumed to be aligned based on random similarity if the observed score is not exceeding 95% of the scores obtained for random sequences generated in a Monte Carlo resampling process. Aliscore generates a listfile including all positions that have been identified as "randomly similar aligned". We used default settings for ALISCORE, the window size was w = 6, gaps were treated as ambiguous characters (-N option), and for Runs 1–3, we chose the -r option for the maximum number of sequence comparisons. Finally, the Perl script ALICUT (written by Patrick Kück, ZFMK Bonn; http://www.zfmk.de/) was used to delete any alignment region of Runs 1–3 that consisted of sequences exhibiting random similarity (alignment masking).

For the masking process in ALICUT, the consensus secondary structure given in RNAsalsa was included into the alignment. Consequently, both the aligned sequences and the consensus sequence were masked; this way, the program allows the user to consider secondary structure information for phylogenetic analysis, for example, by implementing mixed models for RNA molecules. By default, ALICUT excludes stem positions if identified as "randomly similar aligned" and converts the corresponding stem nucleotide into a dot ignoring covariation. However, we think that evolution of stem positions is constrained by secondary structure and covariation patterns. Therefore, we used the -s function in ALICUT to keep all stem positions in the alignment.

In addition to RNAsalsa and ALISCORE, we carried out an alternative approach considering secondary structure information for 18S and 16S sequences. We clipped both terminal regions manually from the pre-aligned data set for Runs 4–8, because these regions appeared to contain erroneous or doubtful sequence fragments for a number of taxa. Subsequently, the pre-alignments of 18S and 16S were realigned manually based on secondary structure information (for methodological suggestions see Kjer, 1995 and Kjer et al., 2007). We used reconstructions of the secondary structure that were available for some of the taxa included in our analysis on the Comparative RNA Web (CRW) site (Cannone et al., 2002) and the European ribosomal RNA database (Wuyts et al., 2004). After a general identification of homologous structures, we were able to reallocate largely misaligned sections of the sequence within the pre-alignment. For example, we found relatively long sections of 18S sequences that were misaligned for five taxa (*Derocheilocaris*, Tanaidacea, *Lightiella*, *Allopaurus* and *Scutigerella*). These sections contained several hundred bps that were entirely misaligned by up to 1500 positions within the 18S alignment. In their new positions, the sections could be unambiguously allocated and realigned according to highly conserved structures. In addition, we realigned numerous smaller sections based on recognizable, unique motifs, so that an estimated 40% of the positions in the standard alignment were resolved and rearranged. The structural optimization also revealed that the two 18S sequences of *Scutigerella*, which differed markedly in the pre-alignment, were highly compatible after rearrangements. Therefore, we excluded the shorter one of the two sequences

Table 2
Different settings and steps of multiple sequence alignment and phylogenetic tree reconstruction explored in this study. All Bayesian analyses were performed using a parallel version of MrBayes (MPI) 3.0 on HP quad core blade systems (32 GB RAM each) of a Linux cluster (ZFMK, Bonn). Maximum likelihood analyses were conducted with the program GARLI and the algorithm FastDNA ML on a dual core G5 Macintosh (UVMH).

Analysis	Run 0	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8
Program/method	MrBayes						GARLI ML		Fast-DNA ML
(Pre-) Alignment software	MUSCLE	MAFFT	MAFFT	MUSCLE	MUSCLE	MUSCLE	MUSCLE	MUSCLE	MUSCLE
Secondary-structure optimization	No	RNAsalsa	RNAsalsa	RNAsalsa	By hand	By hand	By hand	By hand	By hand
Alignment evaluation	No	ALISCORE[a]	ALISCORE[a]	ALISCORE[a]	By hand	By hand	By hand	By hand	By hand
Alignment masking	No	ALICUT	ALICUT	ALICUT	No	By hand	No	By hand	By hand
RY-coding (COI+16S)	No	Yes	Yes	Yes	No	No	No	No	No
Model settings	nst=6 gamma	mixed[b] gamma	mixed[b] invgamma	mixed[b] gamma	nst=2 gamma	nst=2 gamma	nst=2 gamma	nst=2 gamma	n.a.
Partitions	5	5	5	5	3	3	1	1	1
n-Generations	20 mio	40 mio	40 mio	40 mio	40 mio	40 mio	5 mio	5 mio	n.a.
Bootstrap replicates	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	500	273	n.a.
Network of data matrix	No	Yes	No	Yes	Yes	Yes	No	No	No

nst, number of substitution types; n.a., not available; UVMH, University of Veterinary Medicine Hannover; ZFMK, Zoologisches Forschungsmuseum Alexander Koenig.
[a] With -r option.
[b] For details see Section 2.4.

(AF007106) from further analysis (Runs 4–8). Similary misaligned sections were found within the 16S partition. After structural optimization, we prepared two data sets for phylogenetic analysis. Runs 4 and 6 contain the structurally optimized alignment with all original sequence data, including numerous length polymorphisms. This data set is composed of 5218 characters (18S = 3885, 16S = 673, and COI = 658 characters). For Runs 5, 7 and 8, we deleted single uninformative sites (sites containing nucleotides for only one taxon) and highly variable sections that could not be resolved according to secondary structure information. The smaller data set for Runs 5, 7 and 8 has an 18S partition with 2184 characters, while that of 16S has 444 positions.

Furthermore, we reconstructed phylogenetic networks (Huson and Bryant, 2006) to evaluate the data structure and potential conflicts for our different approaches using the neighbor-joining algorithm in Splitstree (Huson, 1998). Without constraining the results to conform to a tree-like diagram, these phylogenetic networks can be used to visualize the presence of conflicting signals in the data. Conflicts are indicated by non-parallel edges that point to conflicting splits of taxa, and show the relative support for splits in the data by the number of parallel edges supporting a certain split, and the length of the edges (as an indicator for the weight of the split, analogous to branch lengths in a tree). For a detailed description of phylogenetic networks see Huson and Bryant (2006) and Wägele and Mayer (2007).

2.4. Phylogenetic tree reconstruction

For the Bayesian analyses, all runs were performed on a parallel version of MrBayes (MPI) 3.0 with 20–40 million generations for each run. We used standard settings and 4 chains for each of the two parallel runs in MrBayes.

In a series of extensive pretests of the partitioned data matrix, we carried out Bayes Factor Test A and B to identify the best model for the final runs following the criteria of Kass and Raftery (1995). For detailed descriptions of the Bayes Factor Test see Nylander et al. (2004) and Kass and Raftery (1995). In addition, we checked the convergence of each parameter for each run both "by hand" and using the software Tracer 1.4 (Drummond and Rambaut, 2007).

Our initial run (Run 0) was conducted with nst = 6. However, Bayes Factor Test A showed significant convergence problems with this setting. Therefore, we chose the second-best model (nst = 2) for the three partitions of Runs 4–5.

For Runs 1–3, we used a mixed-model setting with five partitions as follows: 1 = 18S loop regions, 2 = 18S stems regions, 3 = 16S loop regions, 3 = 16S stem regions, 5 = COI (see Table 3). For partitions 1, 3 and 5, the 4-by-4 (Standard DNA) model was chosen, for partitions 2 and 4, we applied the doublet model (nst = 2) to account for secondary structures and covariation of paired stem positions. For the 18S loop (partition 1), we chose nst = 2, while the loop region of 16S and the COI sequences were RY-coded to compensate for saturation effects. Since RY-coding only allows transversions, the setting nst = 1 was chosen. In addition, we conducted a test for compositional base heterogeneity of our data set using the program PAUP 4.0b10 (Swofford, 2002), and RY-coding was chosen to accommodate heterogeneous base compositions. The rate variation parameter for all partitions was set to gamma; in Run 2, we chose invgamma. An unlinking of partitions showed a better model likelihood in the pretests (BFT A + B); however, parameter convergence of preliminary test runs with unlinked partitions was more problematic than for runs without unlinked partitions. Therefore, we preferred again the more conservative model settings for a better convergence of parameters.

The tree reconstructions in MrBayes 3.0 were compared to maximum likelihood analyses in Runs 6–8. Maximum likelihood (ML) bootstrap analyses were carried out with the software program GARLI 0.96 (Zwickl, 2006). For both GARLI runs, the maximum number of generations was set to 5 mio, ratematrix was 2 (nst = 2), statefrequencies were estimated as well as the proportion of invariant sites, and the rate heterogeneity model was gamma. We ran ML bootstrap analyses, with 500 replicates for Run 6, and 273 for Run 7. In addition, the algorithm FastDNA ML, implemented in the program BIOEDIT (Hall, 1999), was used for Run 8.

3. Results

3.1. Phylogenetic topologies

The trees resulting from our nine analyses are shown in Figs. 1–3 and Figs. A1–A7 (Appendix B). All trees show certain higher-level

Table 3
Numbers of positions for the partitioned data sets based on alignments using MAFFT and MUSCLE.

Alignment procedure	18S loops	18S stems	16S loops	16S stems	CO1	Total lengths
Run 3 (MUSCLE)	982	698	218	86	544	2528
Runs 1-2 (MAFFT)	1011	688	208	94	546	2547

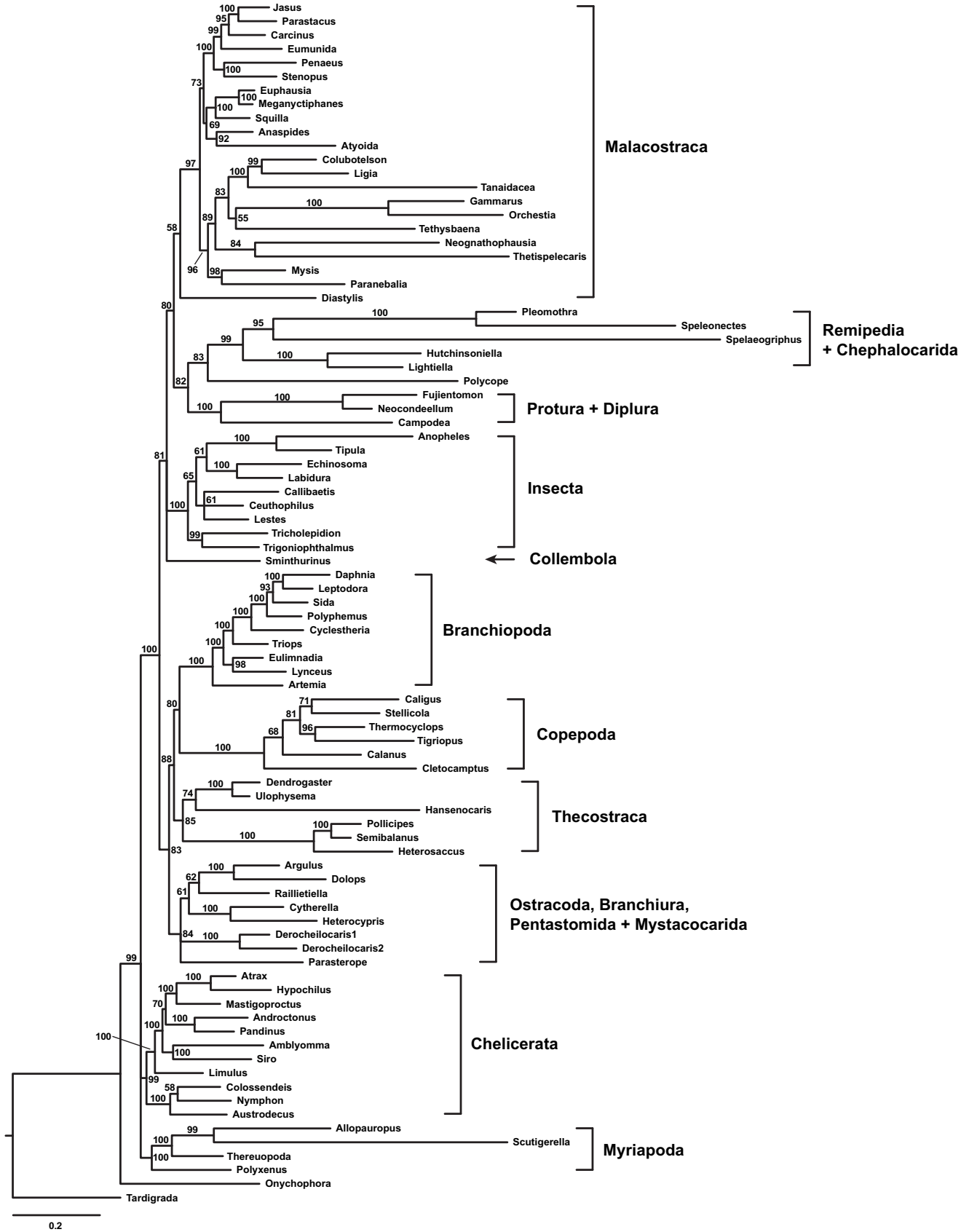

Fig. 1. Bayesian analysis of Run 5, based on a pre-alignment using MUSCLE, and manual secondary-structure optimization with alignment masking. Numbers on nodes represent posterior probability values. See Table 2 for additional settings and parameters.

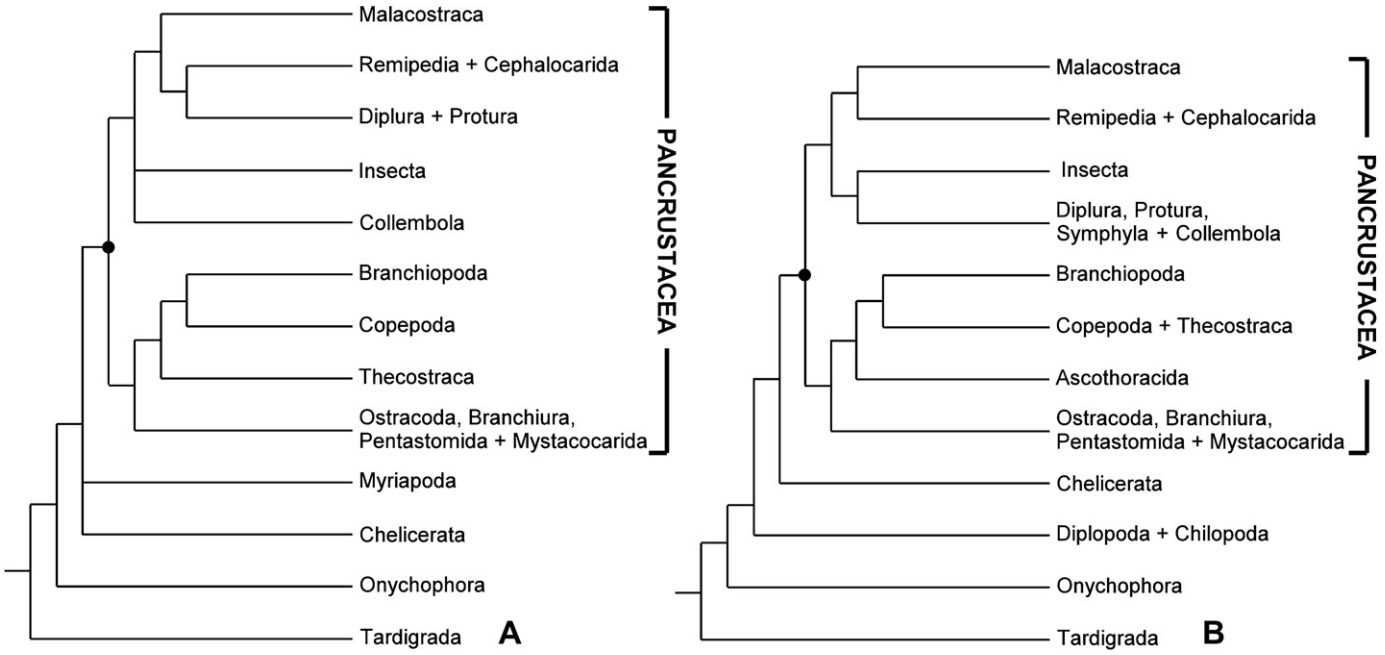

Fig. 2. Condensed versions of phylogenetic trees shown in Fig. 1 (A) and Fig. A7 (B) in Appendix B. The black dot marks the clade Pancrustacea. See the text for the problematic positions of the ostracode *Polycope* and the cumacean *Diastylis* in these trees.

clades, including Branchiopoda, Copepoda, Cirripedia, Pycnogonida, and Cladoceromorpha. However, as shown in the results summary of Fig. 4, the analyses differ substantially in their ability to recover other higher-level groupings, such as Pancrustacea, Hexapoda, Chelicerata, Myriapoda and Malacostraca. In only two of our analyses (Runs 4 and 5), do we find a monophyletic Arthropoda that excludes Onychophora. Moreover, none of our trees supports or resolves hypothesized high-level clades such as Myriochelata and Mandibulata. However, our focus was primarily to investigate crustacean phylogeny, and some of the results are striking. First, we never recover a monophyletic Crustacea that excludes hexapods, which conflicts with the results of several morphological and combined evidence analyses in the literature, but which is in line with several previous molecular studies (see Jenner, 2010). Second, we never recover a monophyletic Ostracoda, confirming increasing suspicions that podocopans and myodocopans are not sister taxa. Third, the trees of Runs 4, 5 and 8 are the most resolved ones and show the same basic topology. They suggest two basic pancrustacean clades. The first clade includes the malacostracans, remipedes, cephalocarids and hexapods, while the other one contains the branchiopods and maxillopodans (copepods, thecostracans, mystacocarids, branchiurans, pentastomids and ostracodes, excluding the myodocopan ostracode *Polycope* (but see *Problematic data* below). In these trees, Maxillopoda is paraphyletic with respect to Branchiopoda.

3.2. Multiple sequence alignments: MUSCLE

The tree obtained from Run 0 (Fig. 3) can be used as a baseline for comparison with the other analyses because it is based on a non-optimized MUSCLE alignment. Although it recovers several clades such as Branchiopoda, Copepoda, Euchelicerata, Chelicerata, and Insecta, it is characterized by an overall lack of resolution on higher taxonomic levels.

3.3. Structural optimization method: RNAsalsa, ALISCORE and ALICUT

Interestingly, the topology of the trees obtained by optimizing the data with RNAsalsa, ALISCORE and ALICUT (Appendix B, Figs. A1–A3) are very similar to the non-optimized tree (Fig. 3), and they appear equally unresolved. As can be seen in Fig. 4, these trees are outliers with respect to the small number of clades that are shared by the results of Runs 4–8. For example, they did recover a monophyletic Pycnogonida and Arachnida but not a monophyletic Chelicerata, which was, however, found in both the non-optimized and manually optimized Bayesian analyses (Figs. 1–3; Appendix B, Fig. A4). Interestingly, comparing the trees based on MAFFT (Runs 1 and 2) and MUSCLE (Run 3) pre-alignments, the latter shows a likely long-branch attraction artifact in uniting unrelated taxa with the longest branches: the symphylan *Scutigerella,* the malacostracan *Spelaeogriphus*, the remipedes, and the cephalocarids.

3.4. Structural optimization method: manual alignment

The phylogenetic reconstructions obtained from manual alignments based on secondary structure information and Bayesian analyses (Run 5: Figs. 1, 2; Run 4: Fig. A4) clearly deviate from the results of the other runs by being more resolved on deeper levels, including a monophyletic Arthropoda (excluding Onychophora). They yield trees with well-supported monophyletic Myriapoda, Chelicerata, and Pancrustacea. (Figs. 1, 2, A4); the pancrustaceans fall into two broad clades, one of which includes the branchiopodans and maxillopodans. In these trees, as in the one derived from the likelihood analysis of Run 8 (Fig. A7), Maxillopoda is paraphyletic with respect to Branchiopoda. These trees agree in dividing the maxillopodans across three clades, (1) Copepoda, (2) Cirripedia, and (3) Mystacocarida + Branchiura + Pentastomida + Ostracoda (except *Polycope*). Thecostraca is only a clade in the trees of Run 5 (Figs. 1, 2). The other major clade includes hexapods, remipedes, cephalocarids, and malacostracans. However, although Insecta are monophyletic in these trees, Hexapoda are not (see below under *Problematic data*).

Bootstrap analyses based on maximum likelihood (ML) for the data sets of Runs 6 and 7 produced poorly resolved trees, with paraphyletic Myriapoda and Chelicerata (Appendix B, Figs. A6, A7). Although Pancrustacea is supported in both trees, the only monophyla within this clade were Branchiopoda and Malacostraca

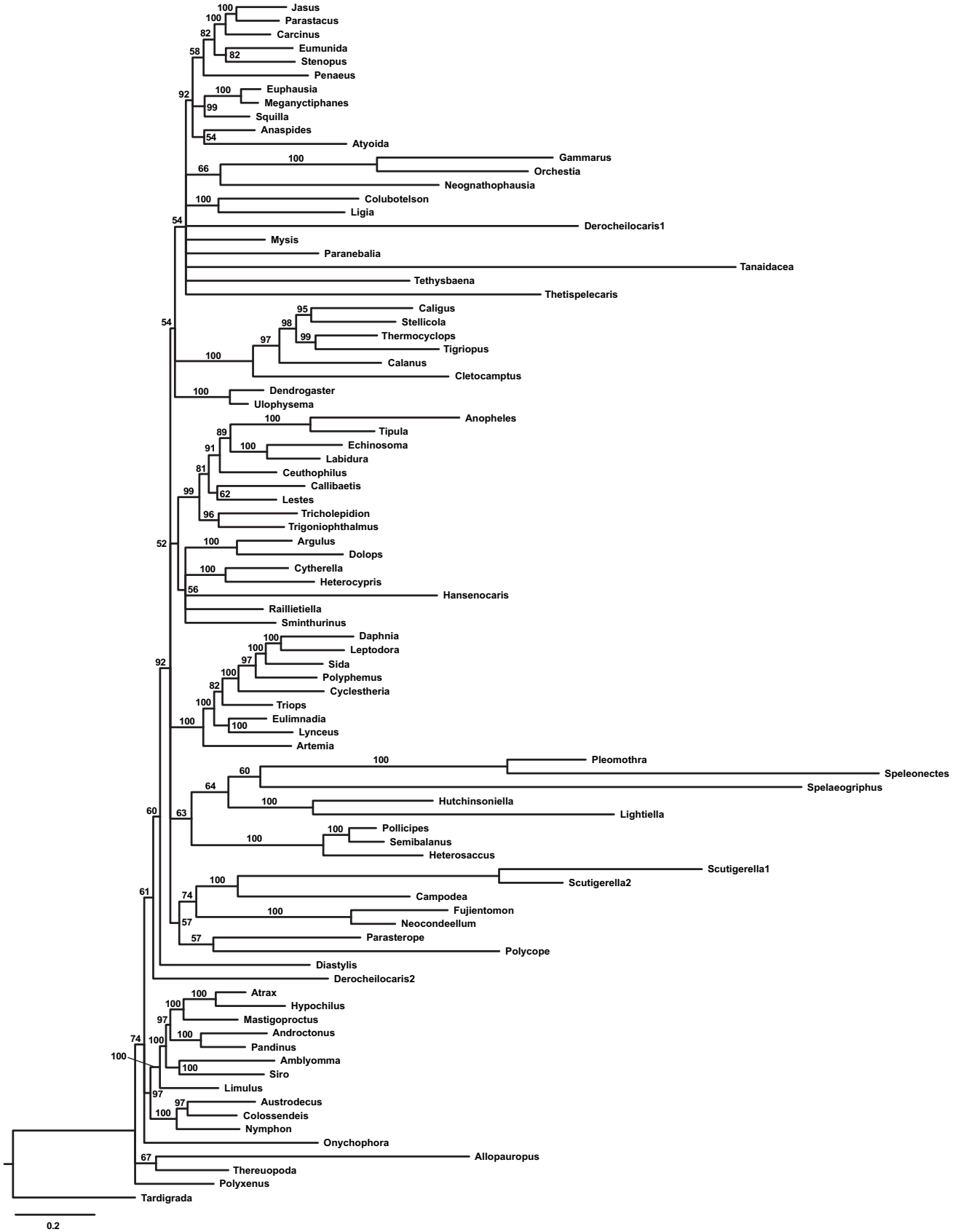

Fig. 3. Bayesian analysis of Run 0, based on multiple sequence alignment using MUSCLE without secondary structure optimization and alignment masking. Numbers on nodes represent posterior probability values. See Table 2 for additional settings and parameters.

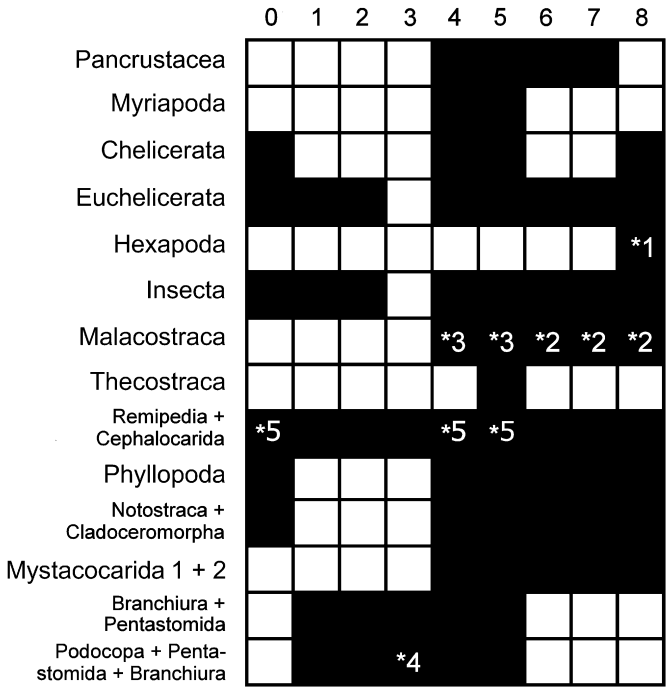

Fig. 4. Navajo rugs showing the distribution of arthropod clades for all nine analyses presented in this study (top row; see Table 2 for an overview). Black squares indicate monophyly, white squares non-monophyly. Numbered asterisks within the diagram are denoted as follows: *1 = including *Scutigerella*; *2 = excluding *Diastylis*; *3 = excluding *Spelaeogriphus*; *4 = including Mystacocarida 1; *5 = including *Spelaeogriphus*.

(excluding *Diastylis*; see *Problematic data*). As in most other analyses, Remipedia + Cephalocarida are sister taxa.

3.5. Problematic data

Three taxa are problematic in that their position is unstable between analyses, or they are placed in unexpected positions. These taxa are the myodocopan ostracode *Polycope*, the cumacean *Diastylis*, and the mystacocarid *Derocheilocaris*2. For example, in the tree obtained from Run 5, the ostracode *Polycope* is nested within a clade composed of Remipedia + Cephalocarida and Diplura + Protura (Figs. 1, 2). In Run 8, *Polycope* appears at the base of the tree on a small clade with the onychophoran (Fig. A7). A BLAST search with the 18S sequence of *Polycope* indicates a high similarity with collembolan and other non-ostracode sequences, indicating possible contamination. Yamaguchi and Endo (2003), who included the 18S sequence of *Polycope* in a molecular analysis of Ostracoda, noted that the unusual length of their alignment was probably "owing to numerous inferred insertion and/or deletion events, especially in the sequences of *Polycope japonica*". They supposed that the position of *Polycope* in their tree was the result of long-branch attraction. Similarly, all the highest BLAST hits for the 18S sequence of our *Diastylis* sequence are echinoderms, which would explain its basal position in several of the trees, far apart from other malacostracans. The highest BLAST hits for the mystacocarid *Derocheilocaris*2 18S sequence are mites, and this is consistent with the finding of the mite *Acarus* being the sister taxon of *Derocheilocaris* in the 18S phylogeny of Wheeler et al. (2004). Intriguingly, however, this old mystacocarid sequence is the sister taxon to the newly sequenced mystacocarid in the results of Runs 4–8 with high support, and this clade groups within one of the maxillopodan clades. This may imply that the old sequence is a chimerical sequence composed partly of mite and crustacean bits.

The phylogenetic positions of some taxa in our trees may have been affected by long-branch attraction, including, possibly among others, the remipedes, the peracaridan *Spelaeogriphus*, the cephalocarids and the myriapod *Scutigerella*. In the tree obtained by Run 3, these arthropods group in a conspicuous cluster of likely unrelated long-branch taxa (which may, however, in part be the result of largely misaligned sections).

During the manual alignment, the unusual length and structure of the 18S sequence of *Speleonectes tulumensis* caught our eye. A comparison with new 18S sequence data, which were generated parallel to this study from several species of Remipedia, suggests that the unusual sequence of *S. tulumensis* may represent a pseudogene, as already assumed by Spears and Abele (1997). Thus, although we find the intriguing clade of Remipedia and Cephalocarida in the majority of our analyses, this may well be the result of long-branch attraction.

3.6. Data evaluation by network reconstruction and base frequency tests

To visualize the presence and nature of potentially conflicting signals in our data we present four phylogenetic networks (Fig. 5; Figures S1–S3 in electronic supplementary files). The networks clearly show that conflicting signals prevail in our data set, indicated by the preponderance of non-parallel edges that represent conflicting splits of groups of taxa. This lack of a strong tree-like signal is additionally reflected by the presence of many unresolved areas, and low clade support values in some of our trees. Certain clades are apparent in the networks, such as Cirripedia, Copepoda, and Branchiopoda. Accordingly, we recovered these clades in all our phylogenetic analyses.

The networks show that optimization of alignments, either by hand or by software can improve the structure of the data to some extent by removing conflict. This is illustrated by the pycnogonids. The networks in Figs. 5 and S1 are based on the same alignment, with the difference that network 1 (Fig. 5) adopts RY-coding of the nucleotides (coding As and Gs as R, and Cs and Ts as Y). Comparison of networks 1 and 2 (Figs. 5, S1) shows that the three included pycnogonids, an expected clade, only group together in network 1. The same improvement is seen in networks 3 and 4 (Figs. S2, S3 in electronic supplementary files), when compared to network 1 (Fig. 5).

In addition, we checked our data set for heterogeneous base composition for the MAFFT, RY-coded data matrix of Run 1 (Appendix C, Table A2). We found that base compositional homogeneity could be rejected for the total data set (including all groups) and also for the set of all crustaceans ($p = 0.000000$). In contrast, base composition homogeneity could not be rejected for more restricted branchiopod and maxillopodan groups.

4. Discussion

Although the exact divergence times of the major arthropod clades remain to be established, current molecular clock estimates and fossil evidence agree that they, including the main crustacean lineages, started to diverge at least early in the Paleozoic (Walossek, 1999; Regier et al., 2005; Zhang et al., 2007; Peterson et al., 2008). Moreover, preliminary analyses (Regier et al., 2005) suggested that many of the major crustacean lineages may have originated and started to diversify during a relatively brief period of time. This makes the accurate reconstruction of their branching order very challenging, a conclusion confirmed by our analyses.

Although certain relationships are robust in our trees, notably the monophyly of generally accepted clades such as Branchiopoda,

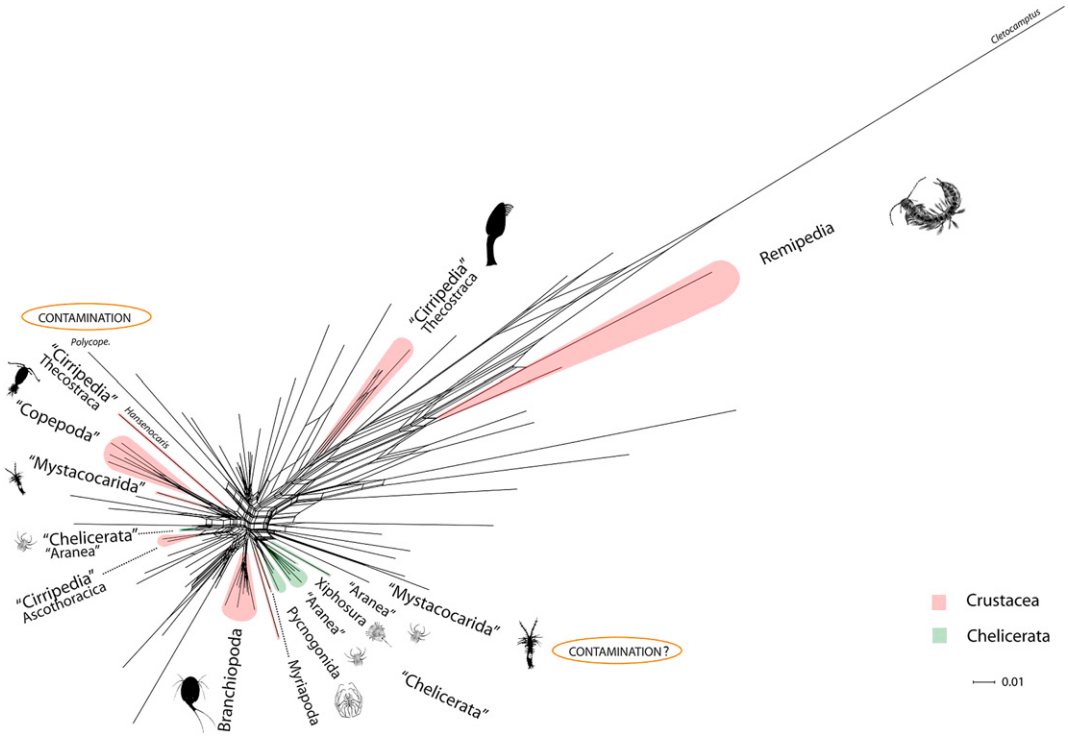

Fig. 5. Neighbor-network of Run 1, based on p-distances constructed in Splitstree (V4.10). The complete (concatenated) RY-coded alignment was used after identification and exclusion of randomly similar sections by the software tools ALISCORE and ALICUT. The colored areas represent crustaceans (red) and chelicerates (green). Quotation marks indicate groupings that are not supported as monophyletic in this network. Presumed contaminated sequences are highlighted by orange circles. See also Section 2.3.

Insecta, and Euchelicerata, many higher-level relationships remain unresolved. The best-resolved trees do suggest a basic phylogenetic split within Pancrustacea into two major clades (Fig. 2), but their monophyly and the relationships within them generally lack their monophyly and the relationships within them are not always fully supported. In view of the methodological variations encompassed by our study, this clearly suggests the need for more and/or different data. These could include complete ribosomal sequences, a denser taxon sampling, incorporation of new loci, and the exploration of alternative outgroups that are separated from the ingroup taxa by a shorter branch. Interestingly, whereas the analysis of quite limited samples of 18S rDNA sequences was sufficient to convincingly sketch the outlines of what was to become the widely accepted new animal phylogeny (Halanych, 2004), more work needs to be done in order to draw a comparable outline of a newly emerging consensus of arthropod and crustacean relationships.

Nevertheless, future analyses of our and other markers can be informed by the variations in results that we obtained in the different runs. The results clearly reveal the importance and impact of different alignment strategies. None of the algorithms we used was able to detect and correct relatively long misaligned sections within the 18S and 16S partitions. These large misaligned sections were identified by eye and realigned manually based on secondary structure information. A commonly held opinion is that manual alignments are subjective and (thus) not repeatable, implying that automated, computerized alignments are objective. However, our automated pre-alignments of ribosomal genes contain obvious errors, such as gaps (ranging from one to several hundred nucleotides) that are correctly aligned for most taxa, but which are obviously misaligned for individual taxa (see Section 3). The resulting misalignment of conserved regions can easily be corrected manually. Therefore, we agree with Kjer at al. (2007), who argued that ignoring apparently falsely aligned, non-homologous positions is in fact also a subjective decision that is likely to affect the resulting phylogenies.

Workers disagree about the relative merit of different strategies, notably manual versus automated sequence alignment (e.g., Kjer, 2004 vs. Ogden et al., 2005). The theoretical preferences of the authors of the current article are spread along this continuum, and as a result it becomes impossible to reach consensus about which particular result is the best. Although a thorough discussion of the theoretical pros and cons of the different approaches that we have adopted is beyond the scope of this article, arguments can be made to prefer one or the other sets of results. For example, the automated pre-alignment of our data (underlying Run 0) contains obvious shortcomings. Hence, both automated and manual refinements may improve phylogenetic signal, as shown both in the improved structure (and reduced conflict) of the phylogenetic networks and/or the better resolution of the trees. For example, the results based on manual alignment are much better resolved than the other analyses, and it is only these analyses that show expected clades such as a monophyletic Arthropoda and Myriapoda. However, a preference for rigorously repeatable automated methods is widespread among systematists, although it is common knowledge that such programs are never perfect. In our study it became clear that the sequence alignment and masking programs can have difficulty with dealing with shorter sequence fragments, so that manual improvements seemed an obvious solution. It is our view that in such a situation of alternative, but non-perfect, methods, one performs various analyses and compares the results with an open mind. Please note that comparison between different runs is complicated due to their requirement of different evolutionary models, so that variation in the results cannot unambiguously be ascribed to a single analysis variable.

A caveat of the current study, as of any published study, is that time limitations have prevented us from doing additional analyses that may have improved our results. For example, we would have

liked to compare analyses based on time-homogeneous and time-heterogeneous models (see Reumont et al., 2009), and assess in more detail the potential of long-branch attraction in our data set. Yet, it is clear that our data set by itself is insufficient to resolve this challenging phylogenetic problem. Future high-level phylogenetic studies will have to explore additional markers, principally nuclear protein-coding genes, which have shown great promise (Regier et al., 2005, 2008). Additionally, we need to further develop tools and methods, such as the programs used in our study, to evaluate the structure of the data (phylogenetic networks) and to refine sequence alignments, in order to separate true phylogenetic signal from confounding non-phylogenetic signals and noise. Last but not least, we should strive to keep an open mind in the evaluation of results generated under different assumptions. Consensus can only arise through inclusion, not exclusion.

Acknowledgments

We thank Karl Kjer, Bernhard Misof, Karen Meusemann, Harald Letsch, Patrick Kueck, Tim Littlewood, and Fred Schram for constructive comments on data, analyses and taxon sampling. We are grateful to the following colleagues for their help in collecting and providing specimens: Tom Iliffe (remipede specimens of *Pleomothra apletocheles*), Jörg Ott (*Lightiella incisa*), Gabi Strieso (*Argulus foliacaetus*), Michael Raupach (*Colossendeis* sp.), and Peter Frenzel (determination of ostracodes). Laboratory work in Bonn was conducted with the help of Berit Ullrich and Steffi Bäcker. This study was in part supported by a grant from the German Research Foundation to S. Koenemann (DFG KO 3483/1-1) and to W. Waegele (DFG WA 530/34); M. Hoenemann was supported by a grant from Boehringer Ingelheim Fonds.

Appendix A

Table A1 List of genetic markers, specimens and taxa used for the phylogenetic analyses. Sequences obtained from GenBank are shown by accession numbers (Acc. no.); new sequences are in bold, sequences from the Hannover lab are marked 1, those from the Bonn lab are marked 2. Gene sequences that we were unable to acquire are marked as "n/a". Classification of Crustacea according to Martin and Davis (2001). See Section 2.1 for additional information.

		Acc. no. COI	Acc. no. 16S	Acc. no. 18S
CRUSTACEA				
Remipedia				
Speleonectidae	*Speleonectes tulumensis*	NC_005938	NC_005938	L81936
Godzilliidae	*Pleomothra apletocheles*	**GU067682**	**GU067680**	**GU067681**
Branchiopoda				
Anostraca	*Artemia franciscana*	NC_001620	NC_001620	AJ238061
Notostraca	*Triops cancriformis*	**GQ328960**[2]	**GQ328946**[2]	EU370422
Diplostraca				
Laevicaudata	*Lynceus brachyurus* +*L. macleyanus*	DQ467706	**GQ328954**[2]	**GQ328957**[2]
Spinicaudata	*Eulimnadia braueriana*	EF189667	EF189604	EF189621
Cyclestherida	*Cyclestheria hislopi*	DQ889093	EF189603	AF144209
Cladocera				
Anomopoda	*Daphnia magna*+*D.* cf. *magna*	AY803061	**GQ328951**[2]	EU370423
Ctenopoda	*Sida crystallina*	AF277889	DQ470594	AM490294
Onychopoda	*Polyphemus pediculus*	**GQ328966**[2]	**GQ328955**[2]	EF189633
Haplopoda	*Leptodora kindtii*	DQ310659	**GQ328950**[2a]	AF144214
Malacostraca				
Stomatopoda	*Squilla mantis*	**GQ328967**[2]	**GQ328956**[2]	**GQ328957**[2]
Leptostraca	*Paranebalia longipes*	n/a	AY744909	EF189630
Syncarida				
Anaspidacea	*Anaspides tasmaniae*	DQ889076	AF133685	L81948
Eucarida				
Euphausiacea	*Euphausia pacifica*	AF177184	AF177176	AY141010
	Meganyctiphanes norvegica	AY601091	AY744910	DQ900731
Decapoda				
Dendrobranchiata	*Penaeus monodon*+*P. semisulcatus*	NC_002184	NC_002184	DQ079766
Pleocyemata				
Stenopodidea	*Stenopus hispidus*	AF125441	AY583884	AY743957
Anomura	*Eumunida sternomaculata*	EU243561	AY351260	AF436011
Palinura	*Jasus verreauxi*	AF192883	AF192874	AF498665
Astacidea	*Parastacus pugnax*	EF599157	AF175239	AF235969
Brachyura	*Carcinus maenas*	FJ159028	AJ130811	AY583974
Caridea	*Atyoida bisulcata*	n/a	EF489995	DQ079738
Peracarida				
Mysida	*Mysis oculata*	EF609269	DQ189194	AM422510
Lophogastrida	*Neognathophausia ingens*	DQ889115	n/a	AM422475
Mictacea	*Thetispelecaris remex*	n/a	n/a	AY781416
Amphipoda	*Orchestia cavimana*	EF989708	AY744911	AY826953
	Gammarus pulex	EF570334	AJ269626	EF582923
Isopoda	*Ligia oceanica*	NC_008412	NC_008412	AF255698
	Colubotelson thomsoni	AF255775	AF259531	AF255703
Cumacea	*Diastylis sculpta*+*D.* sp.	AF137510	U81512	Z22519
Tanaidacea	Tanaidacea sp.	AF520452	n/a	AY743939
Spelaeogriphacea	*Spelaeogriphus lepidops*	n/a	n/a	AY781414
Thermosbaenacea	*Tethysbaena argentarii*	n/a	DQ470612	AY781415

(*continued on next page*)

Appendix A (*continued*)

		Acc. no. COI	Acc. no. 16S	Acc. no. 18S
Maxillopoda				
Mystacocarida	*Derocheilocaris typicus* 1	**GQ328961**[2]	n/a	EU370429
	Derocheilocaris typicus 2	**GQ328961**[2]	n/a	L81937
Copepoda				
Calanoida	*Calanus pacificus*	AF315013	AF315006	L81939
Harpacticoida	*Cletocamptus deitersi*	AF315010	AF315003	n/a
	Tigriopus fulvus+*T. californicus*	DQ913891	DQ913891	EU370430
Cyclopoida	*Stellicola* sp.	DQ889130	n/a	AY627004
Cyclopoida	*Thermocyclops inversus*+*T.* sp.	EU770558	n/a	DQ107580
Siphonostomatoida	*Caligus elongatus*	EF452647	AY660020	AY627020
Ostracoda				
Myodocopa				
Myodocopida	*Parasterope gamurru*	n/a	EU587255	EU591819
Halocyprida	*Polycope japonica*	n/a	n/a	AB076657
Podocopa				
Platycopida	*Cytherella leizhouensis*	n/a	n/a	AB076611
Podocopida				
Cypridocopina	*Heterocypris* sp.	n/a	**GQ328947**[2]	EU370424
Branchiura				
Arguloida	*Argulus nobilis*+*A. americanus*	AY456187	AY456187	M27187
Arguloida	*Dolops ranarum*+*D.* sp.	DQ889096	n/a	DQ813453
Pentastomida				
Cephalobaenida	*Raillitiellia* sp.	n/a	n/a	EU370434
Thecostraca				
Facetotecta	*Hansenocaris itoi*	n/a	n/a	AF439393
Ascothoracida	*Dendrogaster asterinae*	n/a	n/a	AF057560
	Ulophysema oeresundense	n/a	n/a	L26521
Cirripedia				
Sessilia	*Semibalanus balanoides*	**GQ328964**[2]	**GQ328952**[2]	EU370426
Pedunculata	*Pollicipes pollicipes*	**GQ328962**[2]	**GQ328948**[2]	EU370427
Kentrogonida	*Heterosaccus californicus*	n/a	AY520756	AY265359
Cephalocarida				
	Hutchinsoniella macracantha	AY456189	AY456189	L81935
	Lightiella incisa	**GQ328968**[2]	n/a	**GQ328959**[2]
HEXAPODA				
Protura				
Sinentomata	*Fujientomon dicestum*	n/a	n/a	AY596359
Acerentomata	*Neocondeellum dolichotarsum*	n/a	n/a	AY037170
Diplura	*Campodea fragilis*+*C. tillyardi*	DQ529236	NC_008233	AF173234
Collembola	*Sminthurinus bimaculatus*	AY555545	AY555555	AY555522
Insecta				
Archaeognatha	*Trigoniophthalmus alternatus*	NC_010532	NC_010532	U65106
Zygentoma	*Tricholepidion gertschi*	AY191994	AY191994	AF370789
Pterygota	*Callibaetis ferrugineus*	AY326804	AF370873	AF370791
	Lestes rectangularis	n/a	EF044271	FJ010011
Neoptera	*Echinosoma yorkense*	n/a	AY144636	AY144626
	Labidura riparia[a]	AB435163	AY144640	U65114, AY707333, AY707356
	Ceuthophilus gracilipes+*C. uthaensis*	AY793593	AY793561	AY521870
	Tipula sp.	AY165639	EU005437	X89496
	Anopheles gambiae+*A. albimanus*	DQ465336	L20934	L78065
MYRIAPODA				
Chilopoda	*Thereuopoda clunifera*	AY288739	AY288716	AF119088
Symphyla	*Scutigerella causeyae* 1	DQ666065	DQ666065	AY336742
	Scutigerella causeyae 2	DQ666065	DQ666065	AF007106
Pauropoda	*Allopauropus* sp.	n/a	n/a	DQ399857
Diplopoda	*Polyxenus lagurus*	AF370840	n/a	X90667
CHELICERATA				
Pycnogonida	*Austrodecus glaciale*	DQ390048	DQ389994	DQ389890
	Nymphon sp.	**GQ328963**[2]	**GQ328949**[2]	EU420136
	Colossendeis sp.	**GQ328965**[2]	**GQ328953**[2]	EU420135
Xiphosura	*Limulus polyphemus*	AF216203	AF373606	L81949
Arachnida				
Araneae	*Atrax* sp.	n/a	AF370857	AF370784
	Hypochilus thorelli+*H. pococki*	NC_010777	NC_010777	AF062951
Scorpiones	*Pandinus imperator*	AY156582	AY156567	AY210831
	Androctonus australis	AJ506919	AJ506868	X77908
Acari	*Amblyomma triguttatum*	AB113317	AB113317	AF018641
Opiliones	*Siro valleorum*	AY639580	AY639552	AY639492
Uropygi	*Mastigoproctus giganteus*	NC_010430	NC_010430	AF005446
ONYCHOPHORA				
Peripatidae	*Epiperipatus biolleyi*+*Euperipatoides leuckarti*	NC_009082	NC_009082	U49910
TARDIGRADA				
Heterotardigrada	*Echiniscus testudo*	EF620375	n/a	DQ839607

[a] 18S sequences for *Labidura riparia* consisted of three, non-contiguous fragments.

Appendix B

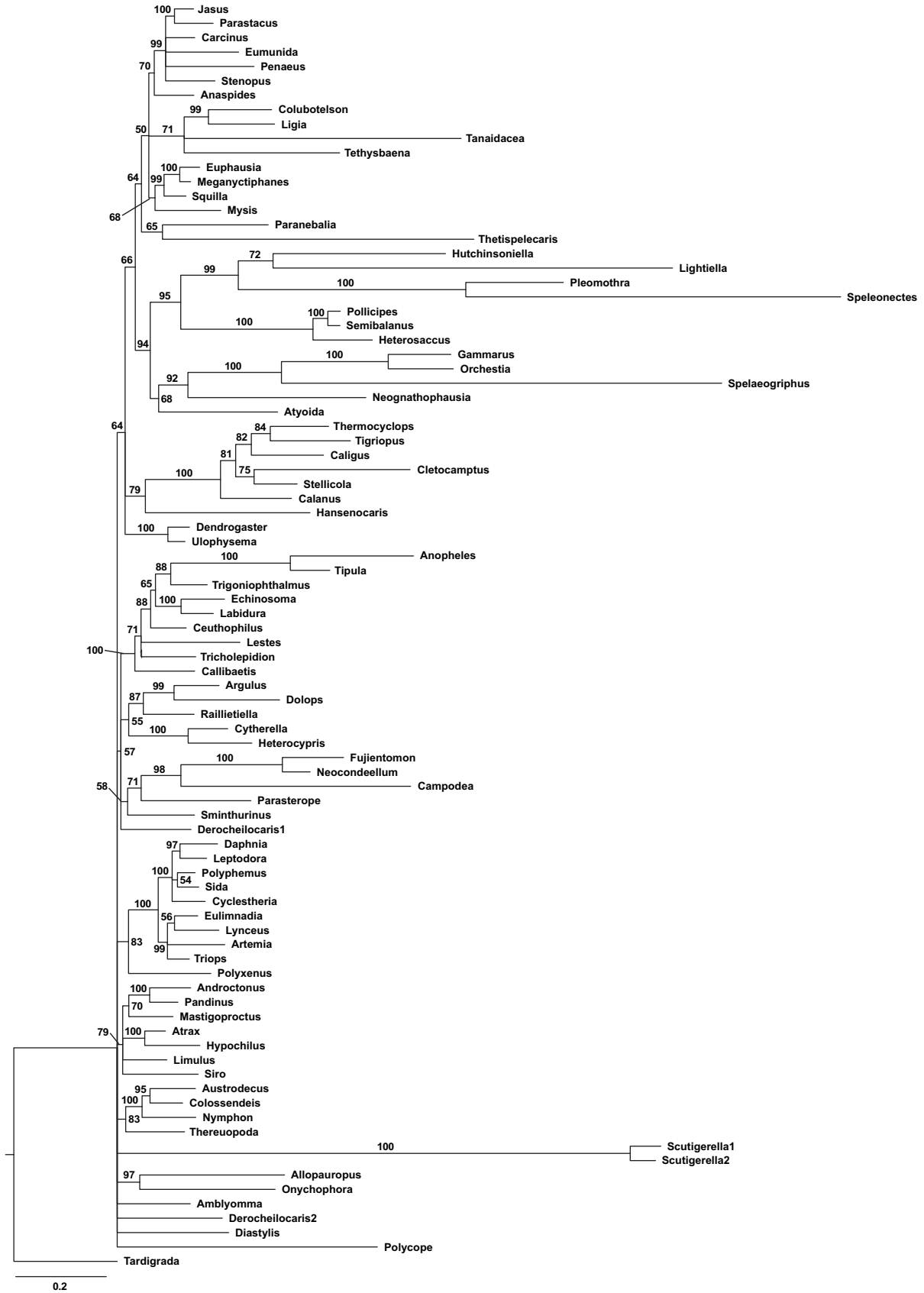

Fig. A1. Bayesian analysis of Run 1, based on pre-alignment (MAFFT), secondary-structure optimization (RNAsalsa), alignment evaluation (ALISCORE) and masking (ALICUT). Numbers on nodes represent posterior probability values. See Table 2 for additional settings and parameters.

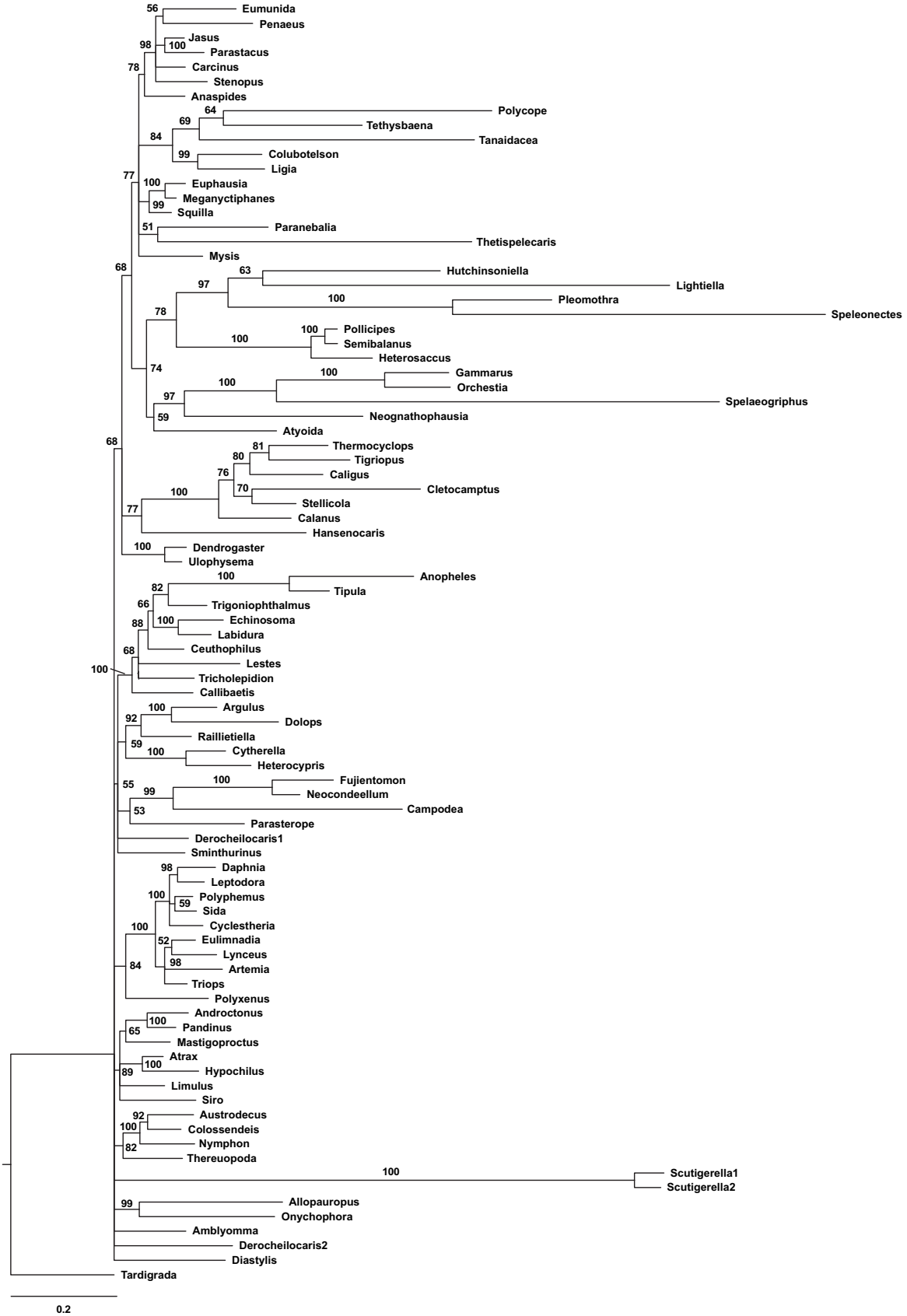

Fig. A2. Bayesian analysis of Run 2, based on a pre-alignment using MAFFT, secondary-structure optimization (RNAsalsa), alignment evaluation (ALISCORE) and masking (ALICUT). Numbers on nodes represent posterior probability values. See Table 2 for additional settings and parameters.

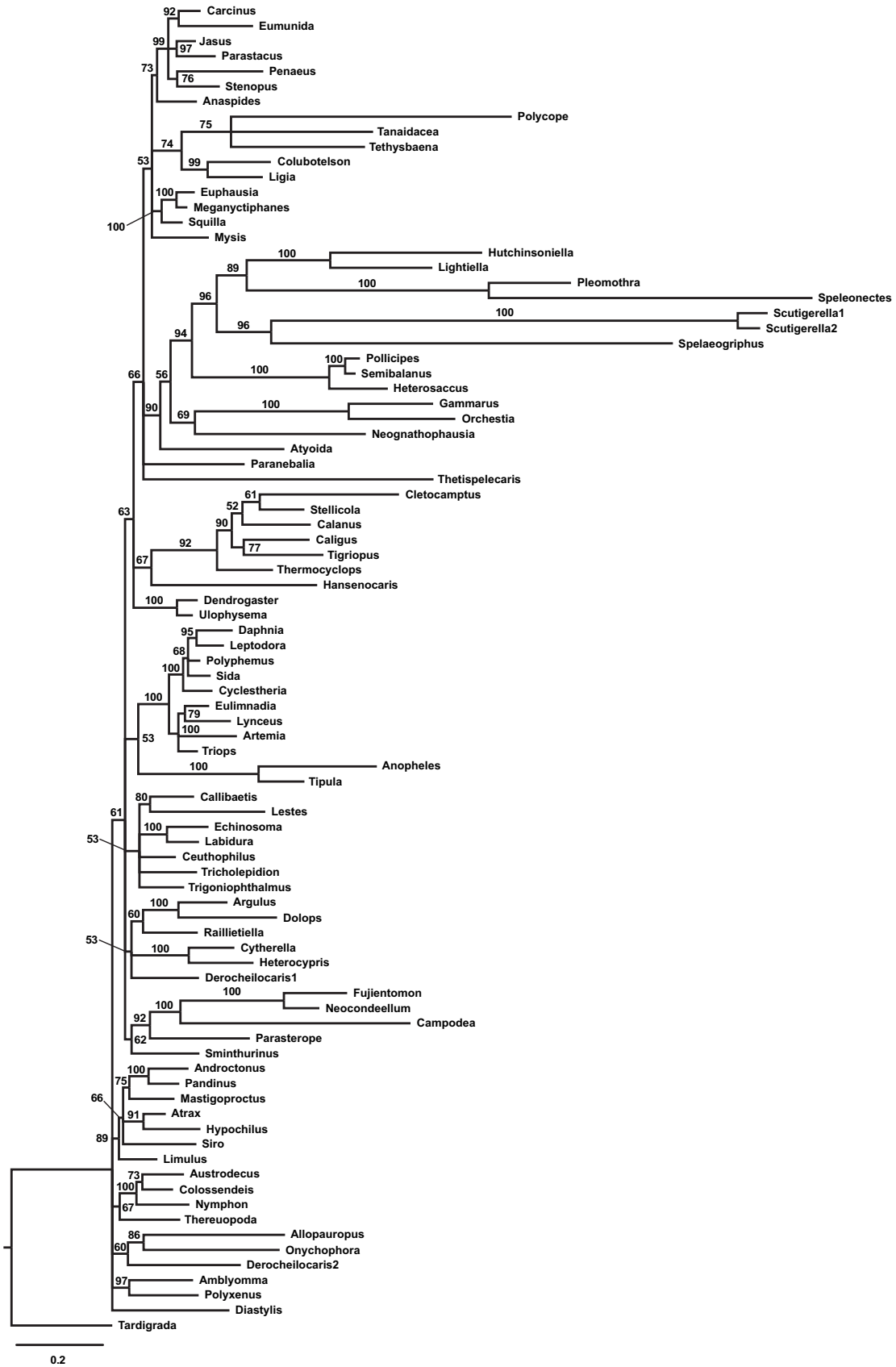

Fig. A3. Bayesian analysis of Run 3, based on a pre-alignment using MUSCLE; secondary-structure optimization carried out with RNAsalsa, alignment evaluation and masking with ALISCORE and ALICUT, respectively. Numbers on nodes represent posterior probability values. See Table 2 for additional settings and parameters.

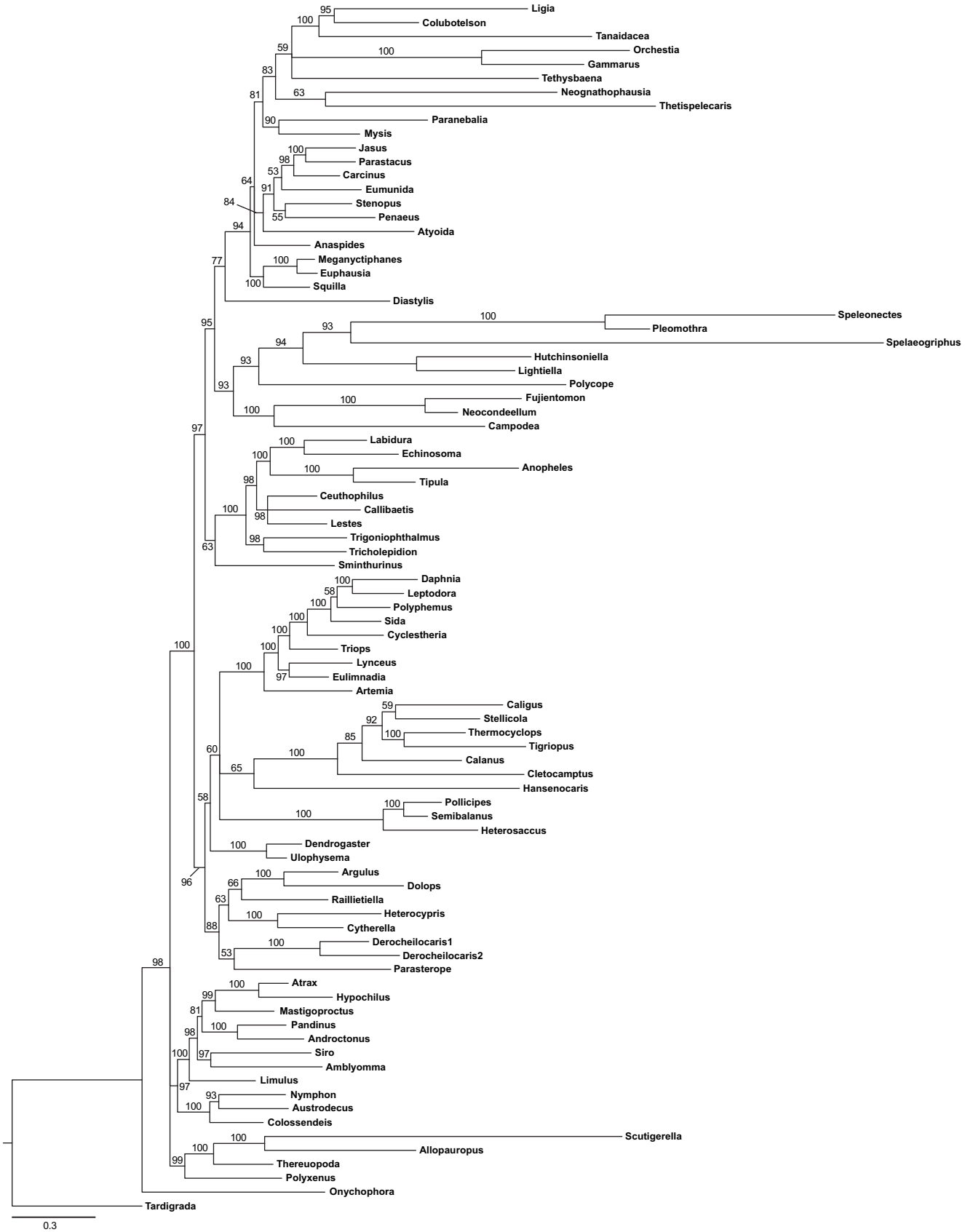

Fig. A4. Bayesian analysis of Run 4, based on a pre-alignment using MUSCLE, and manual secondary-structure optimization without alignment masking. Numbers on nodes represent posterior probability values. See Table 2 for additional settings and parameters.

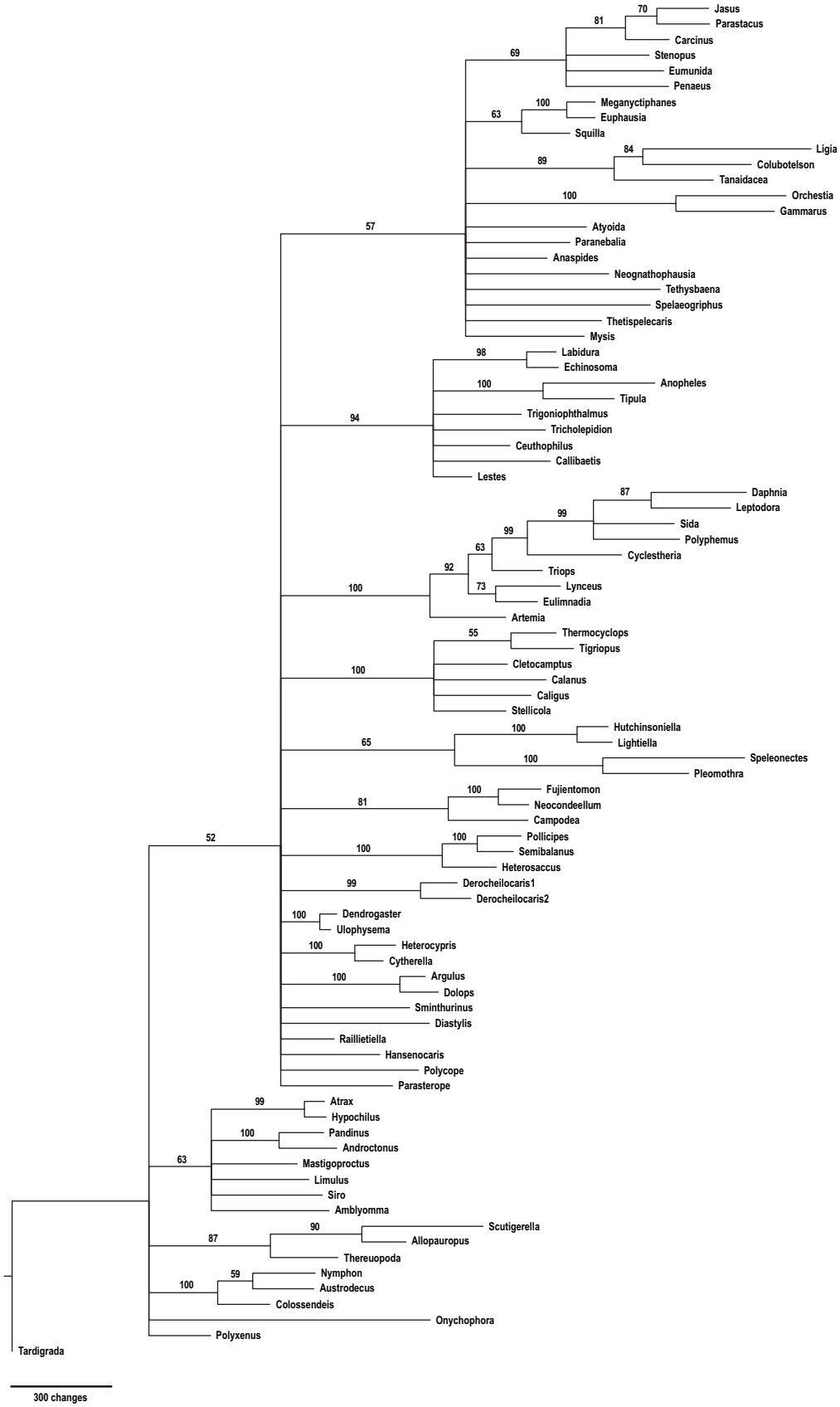

Fig. A5. Maximum likelihood analysis of Run 6 conducted with the program GARLI. Analysis based on same data matrix as in Run 4 (pre-alignment using MUSCLE, manual secondary-structure optimization without alignment masking). Numbers on nodes represent bootstrap values calculated from 500 replicates. See Table 2 for additional settings and parameters.

Fig. A6. Maximum likelihood analysis of Run 7 conducted with the program GARLI. Analysis based on same data matrix as in Run 5 (pre-alignment using MUSCLE, manual secondary-structure optimization with alignment masking). Numbers on nodes represent bootstrap values calculated from 273 replicates. See Table 2 for additional settings and parameters.

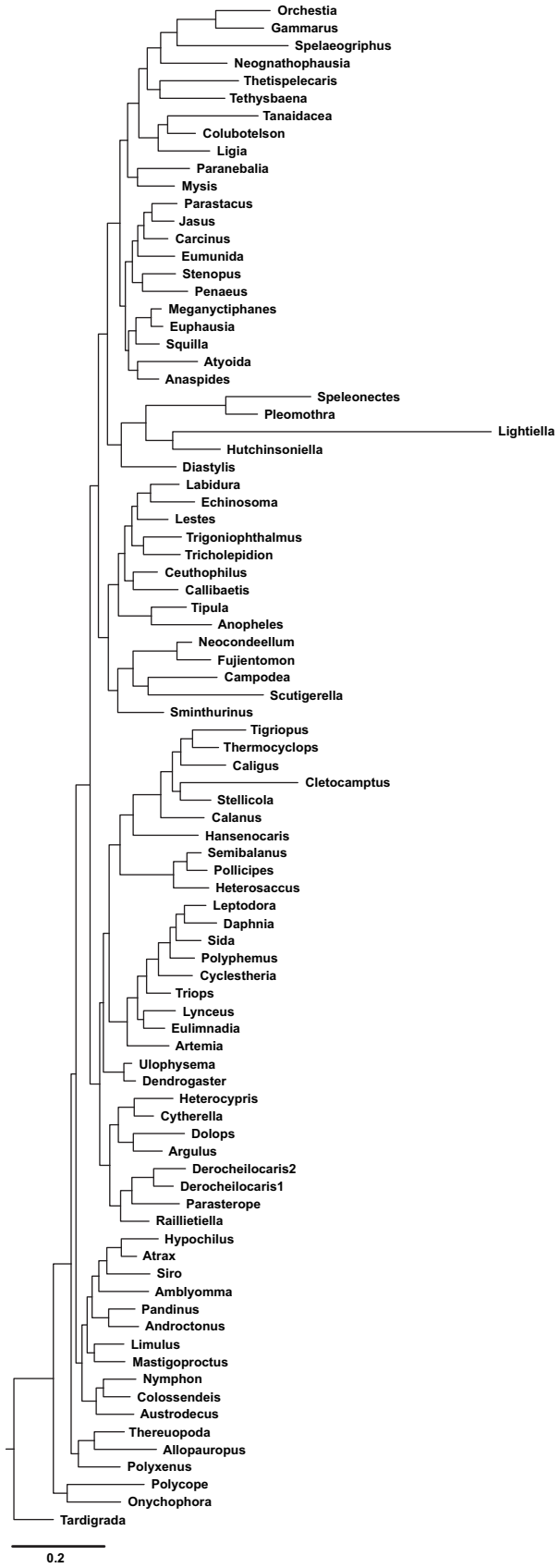

Fig. A7. Single phylogenetic tree obtained by maximum likelihood analysis of Run 8, conducted with the algorithm FastDNA ML (implemented in the program BIOEDIT). Analysis based on same data matrix as in Run 5 (pre-alignment using MUSCLE, manual secondary-structure optimization with alignment masking). See Table 2 for additional settings and parameters, and Fig. 3 for a condensed version of this tree.

Appendix C

Table A2 Base frequency testing for the data set of Run 1. Homogeneous base composition is rejected for p-values under 0.05. Taxa groups excluded from the test are marked by a minus (−) before the parentheses.

	Taxa	p-value
Major groups in the dataset		
All	88	0.00000000
All (−) (Myriapoda)	83	0.00000000
All (−) (Myriapoda, Araneae)	75	0.00000000
All (−) (Myriapoda, Chelicerata)	72	0.00000000
All (−) (Myriapoda, Chelicerata, Tardigrada)	72	0.00000000
All (−) (Myriapoda, Chelicerata, Tardigrada, Onychophora)	71	0.00000000
All (−) (Myriapoda, Chelicerata, Tardigrada, Onychophora, Pterygota)	64	0.00000000
Crustacea	57	0.00000000
Crustacea (−) (Cephalocarida)	55	0.00000000
Crustacea (−) (Remipedia)	55	0.00000169
Crustacea (−) (Remipedia, *Cletocamptus*)	54	0.00006150
Crustacea (−) (Remipedia, Cephalocarida, Pentastomida, Mystacocarida)	51	0.00018063
Crustacea (−) (Remipedia, Cephalocarida, Pentastomida, Mystacocarida, *Cletocamptus*)	50	0.00018063
Crustacea (−) (Remipedia, *Cletocamptus*, Cephalocarida, Tanaidacea, *Thetispelecaris*	50	0.01087916
Crustacea (−) (Remipedia, *Cletocamptus*, Cephalocarida, Tanaidacea, *Thetispelecaris*, *Tethysbaena*)	49	0.02002612
Crustacea (−) (Remipedia, *Cletocamptus*, Cephalocarida Tanaidacea, *Thetispelecaris*, *Tethysbaena*, *Spelaeogriphus*)	48	0.05957794
Branchiopod and maxillopodan groups		
Branchiopoda	9	0.99999856
Copepoda	6	0.16156874
Copepoda (−) (*Cletocamptus*)	5	0.79434022
Ostracoda	4	0.17542550
Ostracoda (−) (*Heterocypris*)	3	0.28077139
Cirripedia	6	0.23684645
Branchiopoda+Mystacocardia	10	0.99826169
Branchiopoda+Mystacocardia+Copepoda	16	0.30029120
Branchiopoda+Mystacocardia+Copepoda (−) (*Cletocamptus*)	15	0.82277081
Branchiopoda+Ostracoda (−) (*Heterocypris*)+Copepoda (−) (*Cletocamptus*)	17	0.35031847
Branchiopoda+Ostracoda (−) (*Heterocypris*)+Copepoda (−) (*Cletocamptus*)+Cirripedia	24	0.37823571

Appendix. Supplementary data

The supplementary data associated with this article can be found in the on-line version at doi: doi:10.1016/j.asd.2009.10.003.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389–3402.

Ax, P., 1999. Das System der Metazoa II. Ein Lehrbuch der phylogenetischen Systematik. Gustav Fischer Verlag, Stuttgart.

Bourlat, S.J., Nielsen, C., Economou, A., Telford, M.J., 2008. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. Molecular Phylogenetics and Evolution 49, 23–31.

Braband, A., Richter, S., Hiesel, R., Scholtz, G., 2002. Phylogenetic relationships within the Phyllopoda (Crustacea, Branchiopoda) based on mitochondrial and nuclear markers. Molecular Phylogenetics and Evolution 25, 229–244.

Budd, G.E., 1996. The morphology of *Opabinia regalis* and the reconstruction of the arthropod stem-group. Lethaia 29, 1–14.

Budd, G., Telford, M., 2009. The origin and evolution of arthropods. Nature 457, 812–817.

Buhay, J., 2009. "COI-like" Sequences are becoming problematic in molecular systematic and DNA barcoding studies. Journal of Crustacean Biology 29, 96–110.

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

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., Pande, N., Shang, Z., Yu, N., Gutell, R.R., 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3, 2.

Carapelli, A., Nardi, F., Dallai, R., Boore, J.L., Liò, P., Frati, F., 2005. Relationships between hexapods and crustaceans based on four mitochondrial genes. In: Koenemann, S., Jenner, R.A. (Eds.), Crustacea and Arthropod Relationships. CRC Press, Boca Raton, pp. 295–306.

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 Evolutionary Biology 7, S8.

Cisne, J.L., 1974. Evolution of the world fauna of aquatic free-living arthropods. Evolution 28, 337–366.

Cobbett, A., Wilkinson, M., Wills, M.A., 2007. Fossils impact as hard as living taxa in parsimony analyses of morphology. Systematic Biology 56, 753–766.

Colgan, D.J., Hutchings, P.A., Beacham, E., 2008. Multi-gene analyses of the phylogenetic relationships among the Mollusca, Annelida, and Arthropoda. Zoological Studies 47, 338–351.

Cook, C.E., Yue, Q., Akam, M., 2005. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proceedings of the Royal Society of London B 272, 1295–1304.

Dahl, E., 1963. Main evolutionary lines among recent Crustacea. In: Whittington, H.B., Rolfe, W.D.I. (Eds.), Phylogeny and Evolution of Crustacea. Museum of Comparative Zoology, Cambridge, pp. 1–15.

deWaard, J.R., Sacherova, V., Cristescu, M.E.A., Remigio, E.A., Crease, T.J., Hebert, P.D.N., 2006. Probing the relationships of the branchiopod crustaceans. Molecular Phylogenetics and Evolution 39, 491–502.

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+Insecta. In: Deuve, T. (Ed.), Origin of the Hexapoda. Annales de la Société Entomologique de France 37, pp. 85–103.

Dreyer, H., Wägele, J.W., 2001. Parasites of crustaceans (Isopoda: Bopyridae) evolved from fish parasites: Molecular and morphological evidence. Zoology-Analysis of Complex Systems 103, 157–178.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214.

Dunlop, J.A., Arango, C.P., 2005. Pycnogonid affinities: a review. Journal of Zoological Systematics and Evolutionary Research 43, 8–21.

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., Sørensen, 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 the resolution of the animal tree of life. Nature 452, 745–749.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.
Edgecombe, G.D., 2004. Morphological data, extant Myriapoda, and the myriapod stem-group. Contributions to Zoology 73, 207–252.
Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome C Oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294–299.
Fortey, R.A., Thomas, R.H. (Eds.), 1997. Arthropod Relationships. Chapman & Hall, London.
García-Machado, E., Pempera, M., Dennebouy, N., Oliva-Suarez, M., Mounolou, J.C., Monnerot, M., 1999. Mitochondrial genes collectively suggest the paraphyly of Crustacea with respect to Insecta. Journal of Molecular Evolution 49, 142–149.
Giribet, G., Carranza, S., Baguna, J., Riutort, M., Ribera, C., 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. Molecular Biolgy and Evolution 13, 76–84.
Giribet, G., Edgecombe, G.D., Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157–161.
Giribet, G., Richter, S., Edgecombe, G.D., Wheeler, Q.D., 2005. The position of crustaceans within the Arthropoda - evidence from nine molecular loci and morphology. In: Koenemann, S., Jenner, R.A. (Eds.), Crustacea and Arthropod Relationships. CRC Press, Boca Raton, pp. 307–352.
Glenner, H., Thomsen, P.F., Hebsgaard, M.B., Sorensen, M.V., Willerslev, E., 2006. The origin of insects. Science 314, 1883–1884.
Haeckel, E.H.P.A., 1866. Generelle Morphologie der Organismen: allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie. Bd. 2, Allgemeine Entwickelungsgeschichte der Organismen. Verlag von Georg Reimer, Berlin.
Halanych, K.M., 2004. The new view of animal phylogeny. Annual Review of Ecology. Evolution and Systematics 35, 229–256.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.
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. Molecular Phylogenetics and Evolution 38, 100–116.
Heymons, R., 1901. Die Entwicklungsgeschichte der Scolopender. Zoologica 33, 1–244.
Horne, D.J., Schön, I., Smith, R.J., Martens, C., 2005. What are Ostracoda? A cladistic analysis of the extant superfamilies of the subclasses Myodocopa and Podocopa (Crustacea: Ostracoda). In: Koeneman, S., Jenner, R.A. (Eds.), Crustacea and Arthropod Relationships. CRC Press, Boca Raton, pp. 249–273.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23, 254–267.
Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J., Kim, W., 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413, 154–157.
Jenner, R.A., 2010. Higher-level crustacean phylogeny: consensus and conflicting hypotheses. Arthropod Structure & Development 39 (2–3), 143–153.
Jenner, R.A., Littlewood, D.T.J., 2008. Problematica old and new. Philosophical Transactions of the Royal Society B 363, 1503–1512.
Jenner, R.A., Ní Dhubhghaill, C., Ferla, M., Wills, M.A., 2009. Eumalacostracan phylogeny and total evidence: limitations of the usual suspects. BMC Evolutionary Biology 9, 21.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.
Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinformatics 9, 286–298.
Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066.
Kjer, K.M., 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Molecular Phylogenetics and Evolution 4, 314–330.
Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Systematic Biology 53, 506–514.
Kjer, K.M., Honeycutt, R.L., 2007. Site specific rates of mitochondrial genomes and the phylogeny of Eutheria. BMC Evolutionary Biology 7, 8.
Kjer, K.M., Gillespie, J.J., Ober, K.A., 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. Systematic Biology 56, 133–146.
Koenemann, S., Schram, F.R., Hoenemann, M., Iliffe, T.M., 2007. Phylogenetic analysis of Remipedia (Crustacea). Organisms Diversity and Evolution 7, 33–51.
Lankester, E.R., 1904. The structure and classification of the Arthropoda. Microscopical Society (London) Quarterly, Journal n. ser 47, 523–582.
Latreille, P.A., 1817. In: Cuvier, G.L.C.F.D. (Ed.), Le Régne animal distribué d'après son Organisation, pour servir de Base à l'Histoire naturelle des Animaux et d'Introduction à l'Anatomie comparée. Les crustacés, les arachnides et les insectes, Vol. 3. Deterville, Paris, pp. 1–653.
Lavrov, D.V., Brown, W.M., Boore, J.L., 2004. Phylogenetic position of the Pentastomida and (pan) crustacean relationships. Proceedings of the Royal Society of London B 271, 537–544.
Mallatt, J., Giribet, G., 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Molecular Phylogenetics and Evolution 40, 772–794.
Manton, S.M., 1973. Arthropod phylogeny—a modern synthesis. Journal of the Zoological Society of London 171, 111–130.
Marletaz, F., Giles, A., Caubit, X., Perez, Y., Dossat, C., Samain, S., Gyapay, G., Wincker, P., Le Parco, Y., 2008. Chaetognath transcriptome reveals ancestral and unique features among bilaterians. Genome Biology 9, R94.
Martin, J.W., Davis, G.E., 2001. An updated classification of the Recent Crustacea. Natural History Museum of Los Angeles County, Science Series 39, 1–124.
Meland, K., Willassen, E., 2007. The disunity of "Mysidacea" (Crustacea). Molecular Phylogenetics and Evolution 44, 1083–1104.
Misof, B., Misof, K., 2009. A Monte Carlo approach successfully identifies randomness of multiple sequence alignments. Systematic Biology 58, 21–34.
Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., Frati, F., 2003. Hexapod origins: monophyletic or paraphyletic? Science 299, 1887–1889.
Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53, 47–67.
Ogden, T.H., Whiting, M.F., Wheeler, W.C., 2005. Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: a reply to Kjer. Cladistics 21, 295–302.
Palumbi, S., 1996. Nucleid Acids II: The polymerase chain reaction. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer Associates, Sunderland, pp. 205–247.
Paps, J., Baguna, J., Riutort, M., 2009. Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. Proceedings of the Royal Society of London B 276, 1245–1254.
Park, S.J., Lee, Y.S., Hwang, U.W., 2007. The complete mitochondrial genome of the sea spider *Achelia bituberculata* (Pycnogonida, Ammotheidae): arthropod ground pattern of gene arrangement. BMC Genomics 8, 343.
Peterson, K.J., Cotton, J.A., Gehling, J.G., Pisani, D., 2008. The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. Philosophical Transactions of the Royal Society B 363, 1435–1443.
Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Molecular Phylogenetics and Evolution 28, 171–185.
Pocock, R.I., 1893a. On the classification of the tracheate Arthropoda. Zoologischer Anzeiger 16, 271–275.
Pocock, R.I., 1893b. On the classification of the tracheate Arthropoda—A correction. Nature 49, 124.
Podsiadlowski, L., Braband, A., Mayer, G., 2008. The complete mitochondrial genome of the onychophoran *Epiperipatus biolleyi* reveals a unique transfer RNA set and provides further support for the Ecdysozoa hypothesis. Molecular Biology and Evolution 25, 42–51.
Regier, J.C., Shultz, J.W., 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Molecular Biology and Evolution 14, 902–913.
Regier, J.C., Shultz, J.W., Kambic, R.E., 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proceedings of the Royal Society of London B 272, 395–401.
Regier, J.C., Shultz, J.W., Ganley, A.R., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J.E., Cummings, M.P., Martin, J.W., Cunningham, C.W., 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Sytematic Biology 57, 920–938.
Reumont, B.M.v., Meusemann, K., Szucsich, N.U., Dell'Ampio, E., Gowri-Shankar, V., Bartel, D., Simon, S., Letsch, H.O., Stocsits, R.R., Luan, Y.X., Waegele, J.W., Pass, G., Hadrys, H., Misof, B., 2009. Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. BMC Evolutionary Biology 9, 119.
Richter, S., Olesen, J., Wheeler, W.C., 2007. Phylogeny of Branchiopoda (Crustacea) based on a combined analysis of morphological data and six molecular loci. Cladistics 23, 301–336.
Roeding, F., Hagner-Holler, S., Ruhberg, H., Ebersberger, I., Von Haeseler, A., Kube, M., Reinhardt, R., Burmester, T., 2007. EST sequencing of Onychophora and phylogenomic analysis of Metazoa. Molecular Phylogenetics and Evolution 45, 942–951.
Rota-Stabelli, O., Telford, M.J., 2008. A multi criterion approach for the selection of optimal outgroups in phylogeny: recovering some support for Mandibulata over Myriochelata using mitogenomics. Molecular Phylogenetics and Evolution 48, 103–111.
Schram, F.R., 1986. Crustacea. Oxford University Press, Oxford.
Schram, F.R., Hof, C.H.J., 1998. Fossils and interrelationships of major crustacean groups. In: Edgecombe, G.D. (Ed.), Arthropod Fossils and Phylogeny. Columbia University Press, New York, pp. 233–302.
Schram, F.R., Koenemann, S., 2004a. Are the crustaceans monophyletic? In: Cracraft, J., Donoghue, M.J. (Eds.), Assembling the Tree of Life. Oxford University Press, New York, pp. 319–329.
Schram, F.R., Koenemann, S., 2004b. Developmental genetics and arthropod evolution: on body regions of Crustacea. In: Scholtz, G. (Ed.), Evolutionary Developmental Biology of Crustacea, Crustacean Issues 15. Balkema, Lisse, pp. 75–92.
Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and

a compilation of conserved polymerase chain-reaction primers. Annals of the Entomological Society of America 87, 651–701.
Snodgrass, R.E., 1935. The principles of Insect Morphology. McGraw-Hill, New York.
Snodgrass, R.E., 1938. Evolution of Annelida, Onychophora and Arthropoda. Smithsonian Miscellaneous Collections 138, 1–77.
Spears, T., Abele, L.G., 1997. Crustacean phylogeny inferred from 18S rDNA. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Chapman & Hall, London, pp. 169–187.
Spears, T., DeBry, R.W., Abele, L.G., Chodyla, K., 2005. Peracarid monophyly and interordinal phylogeny inferred from nuclear small-subunit rinosomal DNA sequences (Crustacea: Malacostraca: Peracarida). Proceedings of the Biological Society of Washington 118, 117–157.
Springer, M.S., Scally, M., Madsen, O., de Jong, W.W., Douady, C.J., Stanhope, M.J., 2004. The use of composite taxa in supermatrices. Molecular Phylogenetics and Evolution 30, 883–884.
Stenderup, J.T., Olesen, J., Glenner, H., 2006. Molecular phylogeny of the Branchiopoda (Crustacea) - Multiple approaches suggest a "diplostracan" ancestry of the Notostraca. Molecular Phylogenetics and Evolution 41, 182–194.
Strausfeld, N.J., Strausfeld, C.M., Loesel, R., Rowell, D., Stowe, S., 2006. Arthropod phylogeny: onychophoran brain organization suggests an archaic relationship with a chelicerate stem lineage. Proceedings of the Royal Society of London B 273, 1857–1866.
Swofford, D.L., 2002. PAUP*4.0b10. Phylogenetic Analysis using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.
Timmermans, M.J.T.N., Roelofs, D., Marien, J., Van Straalen, N.M., 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 Evolutionary Biology 8, 83.
Vaccari, N.E., Edgecombe, G.D., Escudero, C., 2004. Cambrian origins and affinities of an enigmatic fossil group of arthropods. Nature 430, 554–557.
Wägele, J.W., 1993. Rejection of the "Uniramia" hypothesis and implications for the Mandibulata concept. Zoologische Jahrbücher. Abteilung für Systematik 120, 253–288.
Wägele, J.W., Mayer, C., 2007. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. BMC Evolutionary Biology 7, 147.
Walossek, D., 1993. The Upper Cambrian *Rehbachiella* and the phylogeny of Branchiopoda and Crustacea. Fossils and Strata 32, 1–202.
Walossek, D., 1999. On the Cambrian diversity of Crustacea. In: Schram, F.R., von Vaupel Klein, J.C. (Eds.), Crustaceans and the Biodiversity Crisis. Proceedings of the Fourth International Crustacean Congress, Amsterdam. Brill, Leiden, pp. 3–27.
Walossek, D., Müller, K.J., 1990. Upper Cambrian stem-lineage crustaceans and their bearing upon the monophyletic origin of Crustacea and the position of *Agnostus*. Lethaia 23, 409–427.
Walossek, D., Müller, K.J., 1998. Early arthropod phylogeny in light of the Cambrian "Orsten" fossils. In: Edgecombe, G.D. (Ed.), Arthropod Fossils and Phylogeny. Columbia University Press, New York, pp. 185–231.
Wheeler, W.C., 1997. Sampling, groundplans, total evidence and the systematics of arthropods. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Columbia University Press, New York, pp. 185–231.
Wheeler, W.C., Giribet, G., Edgecombe, G.D., 2004. Arthropod systematics. The comparative study of genomic, anatomical, and paleontological information. In: Cracraft, J., Donoghue, M.J. (Eds.), Assembling the Tree of Life. Oxford University Press, New York, pp. 281–295.
Wills, M.A., 1997. A phylogeny of recent and fossil Crustacea derived from morphological characters. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships, The Systematics Association Special Volume Series 55. Chapman & Hall, London, pp. 189–209.
Wilson, G.D.F., 1992. Computerized analysis of crustacean relationships. Acta Zoologica 73, 383–389.
Wuyts, J., Perrière, G., Van De Peer, Y., 2004. The European ribosomal RNA database. Nucleic Acids Research 32, D101–D103.
Yamaguchi, S., Endo, K., 2003. Molecular phylogeny of Ostracoda (Crustacea) inferred from 18S ribosomal DNA sequences: implication for its origin and diversification. Marine Biology 143, 23–38.
Zhang, X.-G., Siveter, D.J., Waloszek, D., Maas, A., 2007. An epipodite-bearing crown-group crustacean from the Lower Cambrian. Nature 449, 595–598.
Zrzavý, J., Štys, P., 1997. The basic body plan of arthropods: insights from evolutionary morphology and developmental biology. Journal of Evolutionary Biology 10, 353–367.
Zrzavý, J., Hypsa, V., Vlásková, M., 1997. Arthropod phylogeny: taxonomic congruence, total evidence and conditional combination approaches to morphological and molecular data sets. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Columbia University Press, New York, pp. 97–107.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis. The University of Texas at Austin.