9-12-2012

Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement, and Pancrustacean Phylogeny

Todd H. Oakley
University of California - Santa Barbara

Joanna M. Wolfe
Yale University

Annie R. Lindgren
Portland State University

Alexander K. Zaharoff
University of California - Santa Cruz

Citation Details

Oakley, Todd H., et al. "Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement, and Pancrustacean Phylogeny. Accepted for publication in: Molecular biology and evolution 30.1 (2013): 215-233.

MBE Advance Access published September 12, 2012

Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement and Pancrustacean Phylogeny

Todd H. Oakley[1], Joanna M. Wolfe[2], Annie R. Lindgren[1,3], and Alexander K. Zaharoff[1]

[1]Ecology Evolution and Marine Biology

University of California, Santa Barbara

Santa Barbara, CA 93106

[2] Department of Geology and Geophysics

Yale University

New Haven, CT 06520

[3] Present Address: Department of Biology

Portland State University

Portland, OR 97201

Abstract – An ambitious, yet fundamental goal for comparative biology is to understand the evolutionary relationships for all of life. Yet many important taxonomic groups have remained recalcitrant to inclusion into broader scale studies. Here, we focus on collection of 9 new 454 transcriptome data sets from Ostracoda, an ancient and diverse group with a dense fossil record, which is often under-sampled in broader studies. We combine the new transcriptomes with a new morphological matrix (including fossils) and existing Expressed Sequence Tag

(EST), mitochondrial genome, nuclear genome and rDNA data. Our analyses lead to new insights into ostracod and pancrustacean phylogeny. We obtained support for three epic pancrustacean clades that likely originated in the Cambrian: Oligostraca (Ostracoda, Mystacocarida, Branchiura, Pentastomida); Multicrustacea (Copepoda, Malacostraca, Thecostraca); and a clade we refer to as Allotriocarida (Hexapoda, Remipedia, Cephalocarida, Branchiopoda). Within the Oligostraca clade, our results support the unresolved question of ostracod monophyly. Within Multicrustacea, we find support for Thecostraca plus Copepoda, for which we suggest the name Hexanauplia. Within Allotriocarida, some analyses support the hypothesis that Remipedia is the sister taxon to Hexapoda, but others support Brachiopoda+Cephalocarida as the sister group of hexapods. In multiple different analyses, we see better support for equivocal nodes using slow-evolving genes or when excluding distant outgroups, highlighting the increased importance of conditional data combination in this age of abundant, often anonymous data. Yet, when we analyze the same set of species and ignore rate of gene evolution, we find higher support when including all data, more in line with a 'total evidence' philosophy. By concatenating molecular and morphological data, we place pancrustacean fossils in the phylogeny, which can be used for studies of divergence times in Pancrustacea, Arthropoda, or Metazoa. Our results and new data will allow for attributes of Ostracoda, such as its amazing fossil record and diverse biology, to be leveraged in broader scale comparative studies. Further, we illustrate how adding extensive next-generation sequence data from understudied groups can yield important new phylogenetic insights into long-standing questions, especially when carefully analyzed in combination with other data.

Keywords: Arthropoda, Phylogeny, Pancrustacea, Ostracoda, Oligostraca, Transcriptomics, Concatenated analyses

INTRODUCTION

The ever-intensifying deluge of molecular sequence information presents both opportunities and challenges for the reconstruction of the history and timing of life on earth. One major challenge is that the sheer volume of data can quickly outstrip the computational power available to conduct cutting edge, statistically rigorous methods, especially during exploratory phases of analysis. While complex model-based phylogenetic techniques recently have made enormous strides in speed (e.g. Guindon, Gascuel 2003; Stamatakis 2006; Ayres et al. 2012), multi-gene datasets large enough to overload any supercomputer are now commonplace, owing to EST and next-generation sequencing technologies . Yet the magnitude of available data and broad applicability of new sequencing technologies also afford opportunities. For example, large-scale transcriptome information can be collected from species without prior genetic knowledge, unlike PCR-based studies that require gene-specific primers. As such, groups highly diverged from model systems can now be studied in unprecedented detail using next-generation sequencing. Another opportunity is that when data are cheap and abundant, the best data for the question at hand can be discovered and retained and data inappropriate for the question can be culled or down-weighted (e.g. Jeffroy et al. 2006; Lartillot, Philippe 2008; von Reumont et al. 2012). While culling approaches are likely to rekindle philosophical debates on the merits of 'total evidence' (Kluge 1989) versus 'conditional combination' of data (Bull et al. 1993), sound definitions of appropriate data, coupled with the pragmatic necessity for computational tractability make attractive the conditional analysis of data. Here, we capitalize on the power of next-generation sequencing technology to investigate the understudied Ostracoda and their position within Pancrustacea, and we show that taxon sampling and attributes of gene families, namely rate of evolution and outgroup selection, can have a strong influence on final results.

Ostracods are small (usually 1–2 mm) crustaceans, which today live in virtually all aquatic habitats, including deep and shallow seas, and small temporary to large freshwater bodies worldwide. Most ostracods fossilize well (except many Myodocopa) because they often live in ocean sediments, and they possess a calcified, usually bivalved carapace, which fully encloses their body. As a result, ostracods have a prolific and complete (Foote, Sepkoski 1999) fossil record that could be used to study divergence times across Pancrustacea, Arthropoda, or Metazoa, which generally have a less complete rock record. Ostracods are also of biological interest, for example exhibiting great variation in eye type (Oakley, Cunningham 2002; Tanaka 2005). Despite interesting paleontological and biological features, ostracods have remained largely refractory to inclusion in larger scale phylogenetic studies. A primary reason for this is that ostracods are ancient and diverse. The root of crown Ostracoda is some 500 million years old (Tinn, Oakley 2008), so ostracods are not only distantly related to any model organism, but are also often distantly related even to each other.

Despite their early origin, diverse biology, and importance in the fossil record, the Ostracoda are not well represented in broader studies , so fundamental questions and opportunities remain. Of the estimated >20,000 living ostracod species from 5 ancient orders (Horne, 2002), very few have been included in broader pancrustacean or arthropod studies. Several recent studies have neglected Ostracoda completely (Timmermans et al. 2008; Andrew 2011; Rota-Stabelli et al. 2011; Strausfeld, Andrew 2011), while others have included only 1-3 species from 1 or 2 suborders (Regier et al. 2008; Regier et al. 2010; von Reumont et al. 2012). Therefore, we still lack fundamental knowledge about the group, such as whether or not Ostracoda are monophyletic. Counter to monophyly, there is weak support for polyphyly in rDNA studies (Spears, Abele 1998; Oakley, Cunningham 2002), which would have important implications including the possible convergent origins of biomineralization and carapace development (Wakayama 2007). In contrast, monophyly is suggested by morphological

phylogenetic analyses (Horne et al. 2005), although multiple putative near outgroups were not analyzed. Recent analysis of 62 protein-coding genes was also consistent with ostracod monophyly, but included only three ostracod species from two suborders, and yielded low support values (Regier et al. 2010). Nevertheless, these limited studies indicate that Ostracoda have an important position within Pancrustacea as a whole because Ostracoda may be a member of Oligostraca (Zrzavy et al. 1998; Regier, Shultz, Kambic 2005; Mallatt, Giribet 2006; Regier et al. 2008), which may form the sister-group to the rest of the Pancrustacea (Regier et al. 2010).

As a riotously speciose and evolutionarily and ecologically important animal clade, the phylogeny and taxonomy of Pancrustacea (Hexapoda + Crustacea) has received considerable attention for decades. Although some progress has been made toward consensus opinions on formerly contentious hypotheses, including support for the monophyly of Pancrustacea and the polyphyly of Maxillopoda (Boxshall 1983; Abele et al. 1992; Friedrich, Tautz 1995; Zrzavy, Stys 1997; Boore, Lavrov, Brown 1998; Shultz, Regier 2000; Dohle 2001; Giribet, Edgecombe, Wheeler 2001; Richter 2002; Delsuc, Phillips, Penny 2003; Nardi et al. 2003; Regier, Shultz, Kambic 2005) - a number of phylogenetic questions still remain. In addition to the question of ostracod monophyly, the sister group to Thecostraca (a group including barnacles) may be Malacostraca (Mallatt, Giribet 2006; Regier et al. 2008; Meusemann et al. 2010; Regier et al. 2010) or Copepoda (Wills et al. 1998; von Reumont et al. 2012). Another outstanding question is the sister group to Hexapoda, which may be the Xenocarida (Remipedia, Cephalocarida) (Giribet, Edgecombe, Wheeler 2001; Regier, Shultz, Kambic 2005; Regier et al. 2010) or perhaps Remipedia (Ertas et al. 2009; von Reumont et al. 2012). A third open question is the phylogenetic position of Branchiopoda (a group including water fleas like *Daphnia*), which may be the sister-group to Hexapoda (Babbitt, Patel 2005; Jenner 2010), or may be the sister group of Multicrustacea (Regier et al. 2010).

Here, we incorporate diverse new transcriptome data from four of five orders, and six of nine suborders of Ostracoda with newly integrated morphological (including fossils), existing EST, mitochondrial-genome, nuclear-genome and rDNA data. While incorporating the understudied Ostracoda, we find good support for several contentious hypotheses, especially when excluding fast-evolving gene families, when excluding distant outgroups from focal hypotheses, or when increasing taxon representation with single genes. We find good support for three large pancrustacean clades with likely origins in the Cambrian: Oligostraca are the sister group to the rest of Pancrustacea, and Oligostraca is further divided into two clades, Multicrustacea, and a clade we call Allotriocarida (Allotrios = 'strange', carida= 'shrimp'). Within the Oligostraca clade, we find support for monophyletic Ostracoda. Within Multicrustacea, our analyses support Hexanauplia (Thecostraca, Copepoda; epithet refers to plesiomorphy of six naupliar molts). Within Allotriocarida, our analyses are equivocal; some support Remipedia as the sister group of Hexapoda, and the dissolution of Xenocarida, with Cepalocarida as the sister group of Branchiopoda. Other analyses show support for remipedes as the sister-group to the rest of Allotriocarida. More broadly, our analyses indicate that previously understudied clades can now be efficient targets of large-scale genetic data, and including these clades using next-generation technologies may often lead to new insights on long-standing phylogenetic controversies.

METHODS

Data

Specimen collection RNA extraction and cDNA synthesis. — We used 454 pyrosequencing methods to collect new transcriptome data from 9 ostracod species from 6 different suborders plus one other oligostracan (*Argulus*) (Table 1). Pyrosequencing yields longer read lengths than

some competing next-generation sequencing technologies, which allowed more robust assembly of transcriptomes in the absence of genomic sequences. Because our future studies will analyze genes expressed in ostracod eyes, we obtained tissue for cDNA from whole bodies, bodies minus eyes and/or eyes alone of pooled individuals for each species (see Table S1 for details). We usually extracted RNA using the organic solvent TRIzol (Invitrogen) according to manufacturer's protocol and treating with TurboDNAse (Applied Biosystems). For *C. californica* and *A. jonesi,* we used the Nucleospin RNA XS isolation kit (Macherey-Nagal). Purified RNA was quantified on a Qubit Flurometer (Invitrogen). We generated cDNA using the SMART or SMARTer cDNA synthesis kit (Clontech). To reduce sequencing artifacts due to poly-T tracts, we used modified 3' primers for first strand synthesis: (SMART) 5'- AAG CAG TGG TAT CAA CGC AGA GTG GCC GAG GCG GGC CTTTTTTTTTTCTTTTTTTTTT – 3' and (SMARTer) 5'- AAG CAG TGG TAT CAA CGC AGA GTA CTTTTTTCTTTTTT -3'. We conducted second strand synthesis using the amplification protocol outlined in the SMART/SMARTer cDNA kits, varying cycle number from 18-22 depending on initial RNA concentration (Table S1). Amplified cDNA was purified using phenol:chloforom:isoamyl protocol and quantified on a Qubit fluorometer (Invitrogen). We pooled separate second strand reactions for each species and tissue type to reach a concentration of 5-7 ug for each cDNA pool. The resulting cDNA samples were shipped either to Duke University or Brigham Young University for titanium pyrosequencing using the Roche 454 platform, according to manufacturer's instructions, employing partial runs with either a manifold or barcodes (Table S1).

Additional molecular data. — We analyzed additional, mostly previously published, molecular data, focusing first on major pancrustacean clades and species included in multiple previous data sets, and second on including exemplars of ostracod families with rDNA data. In particular, we analyzed data from 62 single-copy nuclear protein-coding genes of 27 species (including 3 ostracods) (Regier et al. 2010), plus Expressed Sequence Tag (EST) data from 7 species, all 13 protein coding genes from 15 species' mitochondrial genomes, 6 species' entire genome

sequence (predicted proteomes), 18s rDNA data from 79 species (including 18 new sequences), and 28S rDNA data from 30 species (including 19 new sequences). The sources of these data are detailed in Supplemental Tables 2,3. We included two outgroups from outside Pancrustacea, the myriapod *Scutigera coleoptrata* and the chelicerate *Limulus polyphemus* for several reasons. First, they represent each of the two major arthropod clades outside of Pancrustacea. Second, they have much data represented in our ingroup taxa. Third, they are relatively short-branch taxa in previous studies (e.g. Regier et al. 2010) and short branch outgroups may retain stronger phylogenetic signal in ingroup comparisons (e.g. Lyons-Weiler, Hoelzer, Tausch 1998).

Novel Morphological Matrix. — We scored 183 morphological characters, mainly from literature sources, for 93 extant and 16 fossil pancrustaceans. Characters came primarily from three previous publications. We used all 29 characters scored by Horne et al (2005) for ostracod superfamiles; we did not score additional morphological characters to differentiate species below the superfamily level. Next we used 36 of 97 arthropod-wide characters from Hou et al (2010), which is based on the dataset of Wills (1998). We excluded those characters constant within Pancrustacea and those redundant with Horne et al (2005). In addition, we analyzed 89 characters from Rota-Stabelli, et al. (2011). Twenty-nine additional characters came from other morphological studies (Huys, Boxshall 1991; Wheeler et al. 2001; Høeg, Kolbasov 2002; Pérez-Losada, Høeg, Crandall 2004; Olesen 2009; Syme, Oakley 2012). We used MorphoBank (O'Leary, Kaufman 2011) to concatenate morphological data sets and to score all taxa for as many characters as possible. We incorporated many new character codings for fossils based on personal observations. Our morphological matrix and full character descriptions are available on MorphoBank (morphobank.org) (Project 689).

We included sixteen fossil pancrustaceans in our matrix, which can be used for divergence time studies. Due to our particular aim to resolve the placement of ostracods within Pancrustacea, we followed Hou et al (2010) and included Bradorriids and phosphatocopines, which have been allied with ostracods in the past. We also included five crown-group ostracods with well-preserved 'soft-parts' from the Silurian, Triassic, and Cretaceous, which have been hypothesized as members of Myodocopa and Podocopa. This is especially important, as incongruence in ostracod divergence times estimated from molecular vs. fossil data by Tinn and Oakley (2008) may have been driven by problems with fossil placement. In particular, characteristics of the carapace may be homoplastic (Siveter, Sutton, Briggs 2007; Tinn, Oakley 2008). To combat this, our matrix focused on soft part (including appendage) characters.

Transcriptome Analyses

Assembly. — We assembled new transcriptome data with GS De novo Assembler v2.3 ('newbler'; 454 Life Sciences/Roche) to create a cDNA de novo assembly with default threshold options. We used LUCY (Chou, Holmes 2001) to trim low quality nucleotide reads and deleted any assembled contig below 100 nucleotides in length. Assembled EST's from public databases were provided by Roeding (2009). We obtained data from Regier et al (2010) from GenBank and treated those protein coding genes like EST/transcriptome data in our analyses.

Ortholog determination. — We used HaMStR (Ebersberger, Strauss, von Haeseler 2009) to determine orthologs. HaMStR first employs genewise (Birney, Clamp, Durbin 2004) to translate cDNA sequences in all reading frames. HaMStR then uses profile Hidden Markov Models (HMMs) and hmmr (Eddy 1998) to search all translations for matching genes. For the hmm gene models, we used the 'arthropoda_hmmr3' set of core orthologs, provided with HaMStR. 33 of the 62 proteins analyzed in Regier et al (2010) were not present in these core orthologs, so we trained new HMMs for those proteins using hmmr3 and alignments of each gene from 5

species that cover the phylogenetic breadth of our final analysis: *Skogsbergia lerneri*, *Cypridopsis vidua*, *Speleonectes tulumensis*, *Triops longicaudatus*, *Limulus polyphemus*, and *Scutigera coleoptrata*. After finding candidate orthologs with hmmr, HaMStR next uses blast (Altschul 1997) to search a reference genome, for which we used *Drosophila melanogaster*. If the putative ortholog did not find the fly ortholog as the most similar hit, the gene in question was not retained for phylogenetic analysis. As a result, genes containing in-paralogs (sensu Sonnhammer, Koonin 2002), including for example the common phylogenetic marker EF1- α (e.g. Regier et al. 2008), are not always retained as orthologs by HaMStR.

Alignment and Alignment Masking. — We next aligned each gene family using MUSCLE (Edgar, 2004), and estimated the ML tree topology and branch lengths assuming a WAG model, implemented in RAxML (Stamatakis 2006). We used BioPerl to determine the average length of all branches within a gene family, and then excluded any genes on a terminal branch that was more than 4 times the average. We found this approach removed sequence artifacts, mainly poorly translated sequences. Finally, we reduced noise in the data by identifying and removing aligned regions that did not show more similarity than random. Here, we used ALISCORE and ALICUT (Misof, Misof 2009; Kück et al. 2010) including the '-e' option recommended for EST data. We placed all data in a local MySQL database and wrote custom perl and bash scripts to allow easy generation of data subsets based on criteria such as data types, species, and estimated rate of evolution of the gene family. We coded wrappers (available from T.H.O. upon request) for most of these bioinformatics tools for use in the Galaxy bioinformatics platform (Giardine et al. 2005).

Rate of Evolution – We utilized estimates of rates of evolution for each gene family to select which data to include for different analyses based on rate. In order to compare rates of evolution directly between gene families, we required a gene to be present in all species examined, but

for EST's few genes are present for every species. As such, we compared genes from species for which full genome sequences are available ('proteome-species'). We estimated a phylogeny of proteome-species by aligning and concatenating all orthologous genes as above, and then we used RAxML to estimate branch lengths for each gene family on the overall most likely tree. We used the sum of all branch lengths for each gene family as a measure of its rate of evolution. These measures were used to select genes based on rate in subsequent analyses.

Phylogenetic Analyses

Maximum Likelihood. — Analyses with RAxML 7.2.8 using HPC options (Stamatakis 2006) allowed us to concatenate all data types together, including morphological (binary and multistate), rDNA, EST, and mitochondrial proteins. We analyzed various subsets of the full dataset (explained in results), and each time partitioned data by type. We divided morphological data into two partitions (binary and multi) to allow different models to be applied to each. For the multi-state data, we report analyses using the MK model, as preliminary analyses of the multi-state partition with the GTR model gave non-sensical results. For each tree search, we employed the combined bootstrap and best-scoring ML tree search (option "-f a"), which implements 5 separate Slow ML searches to find the best ML tree. We did not attempt the computationally intensive enterprise of determining separate best-fit models for each of 1000 different genes. Instead, we assumed a GTR model for the rDNA, which is best-fit for multiple similar datasets (Oakley, Cunningham 2002; Oakley 2005; Tinn, Oakley 2008). For EST's we employed the WAG model in all cases, and for mitochondrial proteins, we employed the arthropod mitochondrial (mtART) model (Abascal, Posada, Zardoya 2007). To compare alternative topological hypotheses, we implemented SH (Shimodaira-Hasegawa) tests (Shimodaira, Hasegawa 1999), implemented in RAxML by comparing the best tree found under a constraint to the overall best tree. We investigated the effects of missing data in the

Oligostraca clade (with *Limulus* and *Scutigera* as outgroups) by analyzing data subsets. We created 3 data subsets by only retaining genes present in >0, >5, or >10 species. We also created 5 other data subsets by retaining species possessing >0, >25, >50, >100, >200 data partitions. For each data subset, we investigated bootstrap support with 50 pseudoreplicates for clades of interest.

Fossil Placement and Divergence Times. — An often overlooked element of divergence time estimation is analysis of the phylogenetic relationships of fossils, which can have strong influence on final results (Tinn, Oakley 2008). Instead, fossil placement is often assumed based solely on taxonomic authority (but see Ware, Grimaldi, Engel 2010; Pyron 2011). We used two different methods to determine the phylogenetic placement of fossils. First, we used a Maximum Likelihood (ML) fossil placement algorithm developed by Berger and Stamatakis (2010). This method assumes a molecular phylogeny for a set of extant taxa, and then generates weights for each morphological character based on congruence with the molecular phylogeny. Next, the method attaches the fossils to every possible branch of the molecular tree, and in each case calculates the likelihood of observing the weighted morphological data. The placement of each fossil in the molecular tree is the placement with the maximum likelihood estimate. For easier discussion, we term this method 'weighted fossil placement'. This method is currently only implemented with binary characters in RAxML 7.2.8, and so we could not include our multistate characters in this analysis without developing new software. Second, we examined the placement of fossils in what we term 'concatenated fossils' analyses. Here, we concatenated molecular and morphological data and analyzed the matrices in RAxML 7.2.8. Because we obtained higher support values when analyzing major clades separately (see results), and because analysis of the entire matrix including fossils is very computer time intensive, we performed 'concatenated fossils' analyses on the three separate major pancrustacean clades.

We conducted divergence time analyses utilizing PhyloBayes 3.3b (Lartillot, Lepage, Blanquart 2009), which utilizes Bayesian MCMC sampling to estimate divergence times of a fixed topology. We assumed the topology depicted in Figure 1, and utilized all nuclear protein coding data (ie all 454, EST, and Regier genes), as PhyloBayes does not allow for analysis of mixed data types, precluding the combination of morphological and rDNA data. We report analyses from a relaxed molecular clock, assuming the 'uncorrelated gamma multipliers' model, and uniform priors on three fossil constraints. We also placed a gamma prior on the root, with a mean divergence time of 542 MY (the base of the Cambrian) and a standard deviation of 10 million years. With available computational resources we were able to run the MCMC chain for 1300 steps, and we discarded the first 500 as burnin. We also explored penalized likelihood with an autocorrelated relaxed clock model implemented in r8s (Sanderson 2003).

RESULTS

Data. — Our final data set contained 109 species (93 extant, 16 fossils) and 273785 aligned characters (not all characters present for all species, for example, we included 27 ostracods that only have available morphology and rDNA test ostracod monophyly). Our final data set contained 136 binary and 46 multi-state morphological characters. The final aligned and screened rDNA data (28S plus 18S) comprised 7748 nucleotide characters. The nuclear protein coding genes numbered 1001 genes and 263,306 amino acid characters. The mitochondrial genome proteins totaled 2547 aligned amino acid characters. We analyzed numerous different subsets of this full data set (Table 2).

Phylogenetic analysis

Extant Species Topology.— We obtained support for three epic pancrustacean clades: Oligostraca (Ostracoda, Mystacocarida, Branchiura, Pentastomida) (Zrzavy et al. 1998);

Multicrustacea (Copepoda, Malaxostraca, Thecostraca) (Regier et al. 2010; von Reumont et al. 2012); and Allotriocarida (Hexapoda, Remipedia, Cephalocarida, Brachiopoda). We are the first to propose the name Allotriocarida (which is also Clade #33 of Regier et al., (Regier, Shultz, Kambic)), and our support for this clade, and each of the epic clades, is consistent across our analyses of different data subsets.

The analysis of all extant species with six or more character partitions (there are 6 full genomes, so this minimum usually requires a gene to be present in at least 1 species without a genome) and all character partitions present in four or more extant species, analyzed by data type in RAxML, resulted in strong bootstrap support (100%) for most nodes (Figure 1). We call this data set 'Extant Total' (Table 1). In the Extant Total analysis, the three epic clades are well-supported by bootstrap analysis, Oliogstraca at 100%, Multicrustacea at 95%, and Allotriocarida at 81%. In this analysis, monophyly of classes, including Thecostraca (although represented in our data only by Cirripedia), Copepoda, Malacostraca, Hexapoda and Branchiopoda is supported, each with 100% bootstrap value. Although nearly every node in this most inclusive analysis had very high bootstrap support, four important nodes did not. First, within Oligostraca, the ML tree showed non-monophyly of Ostracoda, with Podocopa grouping with Ichthyostraca (Pentastomida, Branchiura, and Mystacocarida), with only 58% support. Second, within Multicrustacea, Hexanauplia (Thecostraca, Copepoda) was supported with only 39% support. The last two equivocal nodes are within Allotriocarida. The remipede *Spelonectes tulumensis* is sister to Hexapoda with only 67%, and the cephalocarid *Hutchinsoniella macracantha* is the sister group to Branchiopoda with 75% support.

To further test the epic clades, and better understand the four equivocal nodes, we performed multiple additional analyses (Table 3). In particular, we examined nuclear protein data alone to test whether mitochondrial proteins, rDNA, and morphology had a strong impact on our results. We still found strong support for the three epic clades, especially Oligostraca (94%) and Multicrustacea (91%). Although support for Allotriocarida dropped somewhat to 75%

in this analysis, it was still retained in the ML topology. This analysis also failed to support ostracod monophyly.

In additional analyses, two of the highly uncertain nodes were clarified, but two remained equivocal. Monophyly of Ostracoda and Hexanauplia were both better supported in additional analyses, sometimes with very high values. When including only more slowly evolving genes plus rDNA and morphology (Slow 2.5 data set), ostracod monophyly is recovered in the maximum likelihood tree with bootstrap support of 17%. This support increases to 35% when analyzing only the slowest genes plus rDNA and morphology (Slow 2.0 data set). When analyzing the Oligostraca alone (Oligostraca Restricted data set), ostracod monophyly is supported by 85% of bootstrap replicates (Figure 2A). When adding exemplars of ostracod families with rDNA and morphological data, bootstrap support for ostracod monophyly is very high at 96% (Fig. 3). Hexanauplia sometimes has stronger support in additional analyses. With the slowest genes, support goes up to 50%. By studying Multicrustacea taxa alone, Hexanauplia is supported at 85% (Fig. 2B). Despite this reasonably high bootstrap support, a SH-test implemented in RAxML indicates that Hexanauplia is not significantly better at p=0.05 than a tree constrained to fit the Communostraca hypothesis [D(LH): -32.53 SD: 20.69].

Two nodes within Allotriocarida were not well supported, and additional analyses did not improve support. First, we find the remipede *Speleonectes tulumensis* to be the sister taxon to Hexapoda. The highest support of 67% is in our Extant Full analysis, and excluding more rapidly evolving genes yields decreased support at 13% and 11%. In our analysis of Allotriocarida alone, the remipede was not the sister taxon of Hexapoda, but rather it was the sister group to all other Allotriocarida (88%). Because this could be caused by a simple change to the root placement within Allotriocarida, we performed another analysis using outgroups from Multicrustacea, and we obtained the same ingroup topology with higher support (100%) (Fig. S1). Similar (and causally related) to the placement of the remipede, placement of the cephalocarid *Hutchinsoniella* is somewhat equivocal. In our Extant Full analysis, *Hutchinsoniella*

is the sister group of *Branchiopoda* with 75% support. When excluding rapidly evolving genes, support is lower at 57% and 51%. In the analyses of only Allotriocarida, *Hutchinsoniella* is reasonably supported as the sister group to Branchiopoda at 89% and 100% with multicrustacean outgroups (Fig. S1).

Fossil Placement and Divergence Times.—Within Oligostraca, we placed five different fossils within the Ostracoda using two different phylogenetic methods that utilize morphological characters. Three fossils (*Colymbosathon ecplecticos*, *Nasunaris flata*, and *Triadocypris spitzbergensis*) are most closely related to the cylindroleberid ostracod *Actinoseta jonesi* in both concatenated analysis and in site-weighted fossil placement analysis (Fig. 4). Two other ostracod fossils differed in placement depending on analysis. In the concatenated analysis, the Silurian species *Nymphatelina gravida* is the sister group to a clade containing *Actinoseta jonesi* plus the three fossils above, but is a stem-group myodocopid in the site-weighted fossil placement analysis (Fig. 4). The other volatile fossil is *Pattersoncypris*, which is the sister-group of all Myodocopa in the site-weighted placement, but groups with two Cyprididae in the concatenated analysis.

Two bivalved arthropod groups have in the past been allied with Ostracoda. First, we included two bradoriids, which consistently placed outside Pancrustacea in both our concatenated fossil analysis, and our site-weighted fossil placement. Second, Phosphatocopina are bivalved arthropods that were once considered a group of Ostracoda until the discovery of soft parts showed major differences, notably the undifferentiated fourth and fifth cephalic appendage (maxillae in all extant ostracods). In our phylogenetic analyses, these species (*Klausmuelleria* and *Vestrogothia*) proved very volatile. With site-weighted fossil placement, they grouped with Thecostraca. When analyzed with other Multicrustacea in a concatenated analysis, we found a similar placement (Fig. 4). However, we also included phosphatocopines in concatenated analyses with Oligostraca, because of their possible affinity with Ostracoda.

Here, we obtained volatile results, with phosphatocopines as a long-branch sister-group to the ostracod *Puriana* in the concatenated 'Restricted Ostracods' analysis and as a long-branch clade with bradoriids that together are most closely related to *Manawa staceyi* in a concatenated 'Extended Ostracoda' analysis. Due to the volatility and low support when including phosphatocopines with Oligostraca, we do not present these analyses in detail (Fig. S2).

Within Multicrustacea, we placed three fossils. *Waptia fieldensis*, an enigmatic species, was a sister group of Malacostraca under concatenated analysis, and a sister group to Multicrustacea under site-weighted fossil placement. Two other fossil species were allied with the leptostracan *Nebalia hessleri* under concatenated analysis: *Cinerocaris magnifica* and *Nahecaris stuertzi.* However, *Nahecaris* was the sister group of Malacostraca under site-weighted fossil placement. Within Allotriocarida, we placed four fossils. *Lepidocaris rhyniensis* was most closely related to the anostracan *Streptocephalus seali*. Surprisingly, three species were related in a paraphyletic grade at the base of Branchiopoda (in order of closeness) under site-weighted fossil placement: the Orsten fossil *Bredocaris admirabilis, Rehbachiella kinnekullensis*, and *Yicaris dianensis*. The relationships were similar in the concatenated analysis, except the Orsten fossils, which formed a paraphyletic sister group to the cephalocarid *Hutchinsoniella*.

We obtained divergence times with fairly tight confidence intervals for nodes toward the root of the phylogeny, but divergence times with very broad confidence intervals toward the tips of the tree (Fig. 5). Our three major clades are estimated to have diverged very early in arthropod history, perhaps in the Cambrian. The Oligostraca are estimated at 513 million years old (95%CI=535-490), Multicrustacea are estimated at 495 MY (520-469), and Allotriocarida at 498 (521-474). Other nodes of interest include Ostracoda (500; 524-476 MY), Hexapoda (394; 476-270 MY), Copepoda (322; 410-226 MY), and Cirripedia (124; 296-39 MY). Results from Penalized Likelihood implemented in r8s are similar to PhyloBayes results and (Fig S3).

DISCUSSION

One of the next frontiers in Tree of Life studies will be to increase taxon sampling, especially targeting previously understudied groups. This trail can be blazed with next generation sequencing technologies, which allow for anonymous sequencing that does not rely on prior knowledge of closely related genomes. We illustrate with Ostracoda how we now can quickly add extensive data from understudied groups to existing data from better-studied clades, potentially leading to new insights about the understudied clades themselves, and the broader groups to which they belong. Our results and analyses lead us to join a chorus of researchers indicating that conditional combination of data may be a sensible approach when dealing with large, often anonymous, data sets (Rodríguez-Ezpeleta et al. 2007; Lartillot, Philippe 2008; Roeding et al. 2009; Meusemann et al. 2010; von Reumont et al. 2012). In multiple cases where our most inclusive data set yielded equivocal support, we found that excluding rapidly evolving gene families or excluding more distant outgroups led to increased support. As such, our analyses add to a groundswell of recommendations to filter large-scale anonymous data by reasonable criteria. While we used a simplistic approach of filtering by a crude estimate of rate of evolution and by separately reanalyzing strongly supported major clades, other studies have also used more sophisticated approaches to similar effect, such as matrix reduction (e.g. Roeding et al. 2009; Meusemann et al. 2010; von Reumont et al. 2012) and site-heterogeneous mixture models (Rodríguez-Ezpeleta et al. 2007; Lartillot, Philippe 2008). All these approaches are reminiscent of conditional combination approaches espoused at the dawn of the availability of multiple distinct data types (Bull et al. 1993). Our approach led us to several insights into contentious issues in pancrustacean phylogeny.

Oligostraca

Extant Topology. This work adds to a growing consensus that Oligostraca - comprised of Ostracoda, Mystacocarida, Branchiura, and Pentastomida - form the sister-group to the rest of the Pancrustacea. A relationship between the two parasitic taxa, Branchiura and Pentastomida, was first proposed based on sperm morphology (Wingstrand 1972), and later on other morphology (Zrzavy et al. 1998) and molecular data, which also added Ostracoda and the interstitial Mystacocarida (Mallatt, Garey, Shultz 2004; Regier, Shultz, Kambic 2005; Regier et al. 2010). Our analyses show very strong bootstrap support for this clade (99-100%), that is of particular interest for its ancient fossil history. Ostracods are already diverse in the Ordovician (Tinn, Meidla 2001) and may be present in the Cambrian (Harvey, Vélez, Butterfield 2012) and stem-group pentastomids may also be present in the Cambrian (Walossek, Müller 1994; Sanders, Lee 2010; Castellani et al. 2011). Based on the phylogenetic position and ancient divergence from the rest of Pancrustacea, it is clear that Oligostraca should be coveted targets of arthropod phylogenetic studies. While the parasitic pentastomids and the interstitial Mystacocarida can be challenging to collect, Branchiura are common fish parasites and diverse species of Ostracoda are ubiquitous in aquatic environments, so these should be included in future arthropod investigations.

The bulk of our analyses indicate that Ostracoda is a monophyletic clade within Oligostraca. Ours is by far the most comprehensive test of ostracod monophyly to date, as previous studies have had limited taxon or character sampling (Horne et al. 2005; Regier et al. 2008; Koenemann et al. 2010; Regier et al. 2010). . Our most inclusive and taxonomically broad analysis failed to support ostracod monophyly. We suspect that rapidly evolving genes may introduce noise into the most inclusive analysis, supported by the fact that analyzing only rapid genes yields incongruent results, namely ostracod polyphyly (Figure S4). Multiple subsequent analyses using slower genes and focusing only on Oligostraca did support monophyly. Perhaps our most important test of ostracod monophyly was the ‘Ostracod-Extended’ analysis, where we added rDNA and morphological data for exemplars of ostracod families. Importantly, this

dataset contains rDNA (Oakley, Cunningham 2002) and morphological (Horne et al. 2005) data from *Manawa staceyi*, the sole living species in the ostracod Order Palaeocopida, such that all five Orders are represented. In fact this analysis includes representatives of 9 of 10 Suborders, missing only the very rare Sigilloidea, which has no molecular data available. This analysis yielded very strong support for ostracod monophyly (96%, Fig 3). Although less than parsimonious histories are always possible, monophyly fails to support the hypothesis that calcified carapaces evolved convergently in Podocopa and Myodocopa (Wakayama 2007).

Fossil Placement. Three of five fossil ostracods had consistent placement within our pancrustacean phylogeny. First, we found support for the hypotheses of Siveter et al (2003; 2010) and Weitschat (1983a) that the Silurian ostracods *Colymbosathon* and *Nasunaris* and the Triassic *Triadocypris* are related to the extant family Cylindroleberididae. Our present analysis cannot distinguish if these fossils are stem or crown-group cylindroleberidids because we only included one extant exemplar for the family and we did not score morphological characters to differentiate finer than superfamily level. Still, our analyses provide strong confirmation for these fossils as crown-group myodocopids. Therefore, the root of Myodocopida (the common ancestor of *A. jonesi* and *E. morini* in this study) is a reliable calibration point for divergence time studies in Pancrustacea and Arthropoda, with a minimum divergence time as the age of the Herefordshire, 425 MYBP. In addition, we propose that a maximum for Myodocopida is the Burgess Shale (505 MYBP), a Lagerstätte that should have preserved myodocopids had they been present (as many other calcified, bivalved arthropods were preserved).

The two other fossil ostracods had placements that differed depending on the analysis. The Cretaceous *Pattersoncypris* was described as a member of the extant podocope family Cyprididae (Bate 1971; Smith 2000). In our concatenated analysis this is confirmed, but the site-weighted placement method (Berger, Stamatakis 2010) contradicts this entirely and places the fossil on the stem lineage of the Myodocopa. This difference is likely because the site-weighted

method can only use binary traits at present, and many critical characters differentiating ostracods in our matrix are multistate. Testing this explanation awaits methods development. In the mean time, we agree that the Cyprididae placement is more likely, as *Pattersoncypris* possesses very similar limbs to modern representatives (especially fifth, sixth and seventh, as noted by (Smith 2000)). The ostracod *Nymphatelina* was described by Siveter et al (2007) and suggested to be a myodocopid. The alternate positions in our analyses of stem myodocopid (site-weighted placement) or related to the cylindroleberidid *Actinoseta* agree with that suggestion. We also analyzed with the Oligostraca two bradoriid fossils, which have in the past been allied with ostracods based on presence of a bivalved carapace (Sylvester-Bradley 1961), As in Hou et al. (2010), we find the bradoriids to fall outside of Pancrustacea. This is not surprising, as they lack differentiated tritocerebral appendages (mandibles), instead bearing biramous trunk limbs. *Kunyangella* also has only four cephalic limbs (Hou et al. 2010), and five cephalic limbs are a key synapomorphy of Pancrustacea (Rota-Stabelli et al. 2011).

Multicrustacea

Extant Toplogy. Regier et al (2010) coined the term Multicrustaca for the clade including Thecostraca, Copepoda and Malacostraca, for which we find strong support (94-95%). Perhaps the most significant implication of Multicrustacea, is the phylogenetic position of Malacostraca, which has been refractory to consensus (von Reumont et al. 2009; Jenner 2010; Koenemann et al. 2010; Meusemann et al. 2010; Regier et al. 2010; Andrew 2011).. Despite other possibilities, a recurring result is (Malacostraca,(Thecostraca,Copepoda)), which we also recover here. In particular, we explored Hexanauplia (Thecostraca, Copepoda) and found reasonable – although not statistically significant - support in some cases, congruent with some morphological hypotheses (Wills et al. 1998; Martin, Davis 2001). Von Reumont et al. (2012) also advocated this result, and they recovered Hexanauplia after matrix reduction aimed at increasing phylogenetic signal. It seems that the competing result

(Thecostraca,Malacostraca)=Communostraca, which we also obtain in one analysis (Slow 2.5), could be an artifact, as discussed by von Reumont et al. (2012).

Fossil Placement. Based on our analyses incorporating morphological data, we placed five fossils within the Multicrustacea clade. Two fossil placements differ depending on analysis. First, *Waptia* is one of the most enigmatic Burgess Shale arthropods, and we found alternate positions as either a sister group to Malacostraca or to Multicrustacea as a whole. The possible relationship to Malacostraca is supported mainly by eye morphology, which can be homoplastic in Pancrustacea (e.g. Oakley 2003). The ambiguity of its phylogenetic placement makes *Waptia* a poor choice for divergence time constraints. Second, the Devonian fossil *Nahecaris* has been regarded as a stem-group leptostracan, an idea supported by our concatenated analysis. Interestingly, the site-weighed method places *Nahecaris* on the stem lineage of the Malacostraca. This seems to occur due to the lack of leptostracan epipod morphology.

In addition, one fossil placement was consistent between analyses. We find *Cinerocaris* to be the sister taxon of *Nebalia*. This supports the hypothesis of Briggs et al (2004) that it is a stem-group leptostracan, based especially on morphology of the trunk epipods.. As such, *Cinerocaris* provides a valuable calibration point as a member of crown Malacostraca. The root of Malacostraca, the common ancestor of Leptostraca and Eumalacostraca (*Nebalia* and *Libinia* in our analysis) is minimally the age of the Herefordshire Lagerstätte (425 MYBP).

Although the phosphatocopines are traditionally assumed to be related to ostracods (e.g. Müller 1964; Williams et al. 2008; Hou et al. 2010) we unexpectedly and equivocally find them to be allied with Thecostraca. Four morphological characters are implicated in relating phosphatocopines with Thecostraca: an all-encompassing ventral carapace, nauplius larval stage, lack of a differentiated limbless abdomen, and inwardly directed spines on the antennal exopods. This placement is surprising, as recent analyses by Hou et al (2010) placed phosphatocopines as either sister to ostracods or sister to all Crustacea except remipedes.

Clearly the affinities of this group are still under debate, and so using them in divergence time studies would be premature.

Allotriocarida

One of the most compelling questions in pancrustacean phylogeny is what is the sister group of Hexapoda, the riotously speciose clade that includes insects. Similar to “Clade 33” of Regier et al (2005), we find reasonable support (75-85%) for a clade including Hexapoda, Branchiopoda, Remipedia, and Cephalocarida that we call Allotriocarida (allotrios = ‘strange’, carida=‘shrimp’). This clade is satisfying in that it incorporates groups that are under major consideration as the sister taxon to Hexapoda (Spears, Abele 1998; Shultz, Regier 2000; Giribet, Edgecombe, Wheeler 2001; Babbitt, Patel 2005; Regier, Shultz, Kambic 2005; Glenner et al. 2006; Roeding et al. 2009; von Reumont et al. 2009; Meusemann et al. 2010; Regier et al. 2010; Andrew 2011). (Babbitt, Patel 2005; Glenner et al. 2006; Roeding et al. 2009; Meusemann et al. 2010; Andrew 2011), Von Reumont (2012) recently found very strong and consistent support for remipedes as the sister taxon to Hexapoda, and consistent with an Allotriocarida clade, they found Branchiopoda as the sister group to remipedes+hexapods, but they did not analyze any data from Cephalocarida. Those results and ours contrast the 62-protein analysis that found Branchiopoda together with Multicrustacea in a clade named Vericrustacea (Regier et al. 2010). We were tempted to conclude that mitochondrial, rDNA, and/or morphological data were causing our support of Branchiopoda in Allotriocarida rather than Vericrustacea. However, our analysis of nuclear proteins alone (454, EST, and Regier genes) still supports Allotriocarida over Vericrustacea, a result that is statistically significant in an SH test [p<0.01; D(LH)= -491.83, SD= 81.95]. Therefore, our inclusion of six full proteomes and additional transcriptomic datasets likely contributes to our support for Allotriocarida,compared to Regier et al (2010).

Although our support for Allotriocarida is reasonably strong, the sister group to Hexapoda is equivocal in our analyses. Our best candidate is the remipede *Spelonectes tulumensis*. In our most inclusive analysis, we obtained the highest support (67%) for Hexapoda+Remipedia, the clade strongly supported by von Reumont with new transcriptome data that were not included here. Adding those data to our analysis would be an interesting avenue of future research. Unlike ostracod monophyly, support for remipedes+hexapods eroded in additional analyses beyond the most inclusive analysis. Possible reasons are discussed below (see conditional data combination). Also somewhat equivocal is our placement of the cephalocarid *Hutchinsoniella* with Branchiopoda, a relationship proposed in the past (Hessler, Newman 1975; Schram, Hof 1998; von Reumont et al. 2009). We included Cephalocarida, and although we did not add new data, we analyzed more types of data together than previous authors, namely we concatenated morphological data with the nuclear gene data (Regier et al. 2010), rDNA (Giribet, Edgecombe, Wheeler 2001), and complete mitochondrial genome data (Lavrov, Brown, Boore 2004). Although our most inclusive data set supported Branchiopoda+Cephalocarida at 75%, and the analysis of Allotriocarida alone supported this node at 89%, our analyses excluding rapidly evolving genes were not well supported (48-57%). These rapidly evolving genes include almost all mitochondrial genes, and the cephalocarid+Branchiopoda relationship was not recovered in our analysis of nuclear proteins alone, which recovered Xenocarida. Therefore, Cephalocarida+Branchiopoda is being driven by mitochondrial, rDNA and/or morphological data, but is not supported by available nuclear proteins. In summary, we find the inconsistent support for Branchiopoda+Cephalocarida to be intriguing, but adding transcriptome data for *Hutchinsoniella* is necessary before we make strong conclusions.

Fossil Placement. We placed four fossils within the Allotriocarida clade. Interestingly, three 'Orsten-type' fossils (*Bredocaris*, *Rehbachiella*, *Yicaris*) cluster together as stem-group Branchiopods. Orsten fossils (such as *Rehbachiella*) are marine, while nearly all living

branchiopods live in freshwater. As such the phylogenetic position of these fossils could have an impact on differing theories regarding the origin of terrestrial hexapods from a freshwater ancestor, although this hypothesis assumes a sister group relationship between Hexapoda and Branchiopoda, which is not supported by our topology. The Orsten fossils are unique in that they are known mainly from larval stages, with adults presumably not preserved (for an interpretation of the adult Bredocaris as a highly neotenic meiofaunal species, see Müller, Walossek 1988; Boxshall 2007). A number of limb morphology characters and presence of the neck organ seem to drive the placement of Orsten fossils, but codings herein do not account for differences in morphology through ontogeny beyond presence/absence in nauplius larvae (for taxa that hatch as nauplii). Coding of characters for each larval stage is beyond the scope of this paper, but could drastically improve the accuracy of phylogenetic placement of Orsten species In contrast, the Devonian fossil, *Lepidocaris*, was much easier to place. With both analyses, it was a crown-group anostracan. This is consistent with previous discussions. It would be a good calibration point from the Rhynie Chert (410-396 MY), providing a minimum age of 396 MYBP for both Branchiopoda and Anostraca (Table 4,5).

Divergence Time Estimates

Our divergence time estimates highlight a tension between molecular and fossil data. The fossil record yields no unambiguous pancrustacean, much less euarthropod fossils from before the Cambrian, 542 MYBP. At the same time, the amount of molecular divergence coupled with ancient fossils similar to modern families (like the cylindroleberidid ostracod *Colymbosathon ecplecticos* from 425 MYBP) imply a much deeper origin for Pancrustacea. These seemingly contradictory signals have been discussed extensively (e.g. Wray, Levinton, Shapiro 1996; Conway Morris 2000; Blair, Hedges 2005; Erwin et al. 2011) and lead to some of the results depicted here. Namely, our divergence time analyses constrain the root of the phylogeny, for if not, it is estimated to be unreasonably deep, even older than the universe under some models

(analyses not shown). At the same time, very old fossil constraints push some nodes to be old, with necessarily smaller confidence intervals as they push up against the root constraint. These constrained ages also imply very rapid rates of molecular evolution, which could have been possible during a Cambrian explosion. More recent nodes then have very large confidence intervals, as rates of molecular evolution may have changed drastically and are therefore difficult to infer. Despite these large confidence intervals, some known fossils still fall outside our estimated ranges for their crown-group, further highlighting the discord between molecular and fossil data. For example, Briggs et al (2005) described a barnacle from the Herefordshire Lagerstätte dated 425 MYA, yet without fossil constraints near this clade in our analyses, the earliest estimates with our 95% CI are only 296 MYA. Similarly, Cambrian stem-group pentastomid fossils are significantly older than the maximum estimate here of 424 MY for the common ancestor of the pentastomid *Armillifer* and the branchiuran *Argulus*. Despite these contradictions, some divergence estimates are broadly consistent with known fossils. We estimate the poorly fossil–represented Copepoda to be 322 MYA, corresponding well to a recently discovered Carboniferous fossil, 303 MY (Selden et al. 2010). In addition, ostracod mandibles may be present in the Cambrian Deadwood Formation, 510-488 MY, consistent with our estimates here of 500 MY. Given this variation in divergence time estimates, it seems the best way forward for those interested in pancrustacean divergences is to incorporate as much fossil information as possible, preferably by explicit phylogenetic analyses of fossil morphology. Again, this highlights the importance of groups like Ostracoda and Thecostraca that have abundant fossils.

Conditional Data Combination

Our analyses and results add to a rising chorus that decisions about which data to include in analyses can have dramatic effects on the final results, a fact that becomes especially important with large, phylogenomic data sets (Rota-Stabelli et al. 2011; von Reumont et al. 2012). First,

we analyzed data subsets that excluded more rapidly evolving genes. This approach influenced our results in different ways, depending on the clade. For the question of ostracod monophyly, bootstrap support increased incrementally. When only the slowest genes were retained, we obtained the highest bootstrap support for ostracod monophyly. In contrast, when including a similar number of fastest evolving genes, ostracods (and other major clades) were highly polyphyletic, suggesting that rate of evolution of gene families is related to reliability for testing the phylogenetic hypotheses at hand. Fast evolving characters are known to be prone to homoplasy, obscuring phylogenetic signal (Felsenstein 1978). In contrast, when investigating relationships within Allotriocarida, excluding rapidly evolving genes had the opposite effect; bootstrap values were lower for (Remipedia+Hexapoda) and for (Cephalocarida+Branchiopoda) when analyzing only slower evolving genes. Part of this can be explained by mitochondrial data, which are among the fastest evolving genes, and are therefore excluded from the slower evolving gene sets. Within Allotriocarida, there is not yet consensus on the relationships of these taxa, so we cannot say if excluding fast evolving genes is lowering support for the true tree or not. One way forward on this question may be to incorporate more data from remipedes and cephalocarids (von Reumont et al. 2012). Although Cephalocarida are classed in one family, such that adding more species may not add much diversity of taxon sampling, Remipedia are classed in 3 families, such that adding additional diverse species could improve consistency of results for this obviously very challenging question.

We also analyzed each major clade separately, and again found this to impact our results. For both ostracod- and Hexanauplia-monophyly hypotheses, we found strong support when analyzing only the major clade to which they belong. Analyzing one clade at a time could reduce heterotachy, which results from changes in rates of evolution over time. Given the drastic morphological and other differences between major clades, it seems likely that molecular evolution could be similarly disparate, such that analyzing all clades together under a single model of molecular evolution could lead to artifacts, as has been found in simulation

(Kolaczkowski, Thornton 2004; Kolaczkowski, Thornton 2008). We also suspected that that analyzing taxonomic subsets of the full dataset led to more fully sampled matrices and therefore higher support, but this was not borne out by additional analyses within Oligostraca (Table 6). Instead, when analyzing Oligostracan species, we found that support for multiple clades (including Ostracoda) was higher in larger, yet sparser data sets. In contrast, bootstrap support was slightly lower for important clades when including species with sparsely sampled gene sets, but removing these sparsely sampled speciescomes at the considerable expense of reduced taxon sampling (Lindgren et al. 2012). Taken together, these results concur with other authors who indicate that sparse data matrices are not necessarily a problem for phylogenetic analyses (e.g. Driskell et al. 2004).

Conclusion

1) We find that important yet previously understudied taxa, like Ostracoda, can be incorporated with broad-scale studies using next-generation sequencing technology.
2) We find good support for three major pancrustacean clades: i) Oligostraca (Ostracoda, Mystacocarida, Branchiura, Pentastomida), which forms the sister group of the rest of Pancrustacea ii) Multicrustacea (Malacostraca, Cirripedia, Copepoda) iii) Allotriocarida (Branchiopoda, Remipedia, Cephalocarida, Hexapoda).
3) We find for the first time good support for monophyletic Ostracoda, with their closest relatives as Ichthyostraca (Mystacocarida, Branchiura, Pentastomida).
4) We find reasonable support for Hexanauplia (Cirripedia + Copepoda) and variable support for Remipedia+Hexapoda and Cephalocarida+Branchiopoda.
5) We were able to reliably place several fossils within the Pancrustacea, which can be used for calibration points in divergence time studies (Table 5).

6) We find that analyzing data subsets can have a major impact on final results. In particular, excluding rapid genes increased support for ostracod monophyly, but had opposite effect within Allotriocarida. Analyzing major clades separately - reducing heterotachy and/or increasing the density of the data matrix - led to strong support for monophyletic Ostracoda and Hexanauplia.
7) Sparse data matrices, such as those produced by anonymous transcriptome sequencing, can produce phylogenetic results with high bootstrap values.

Supplementary Material

Supplemental Table S1. Details of tissue preparation for pyrosequencing.

Supplemental Table S2. Sources of protein coding genes.

Supplemental Table S3. Sources of rDNA data.

Supplemental Figure S1. Analysis using an alternative multicrustacean outgroup for Allotriocarida does not alter ingroup.

Supplemental Figure S2. Phosphatocopines, when analyzed with Oligostraca, fall within crown group ostracods, but on a very long branch with very low support.

Supplemental Figure S3. Penalized likelihood divergence time estimates using the same constraints as the Bayesian analysis of Figure 5 show similar results.

Supplemental Figure S4. Phylogeny using fast-evolving genes, showing rampant polyphyly, even of often supported clades.

Acknowledgments

Thanks to S. Yamaguchi and G. Tom, for help with data collection and analysis. Thanks to T. Near, G. Edgecombe, T. Hegna, for comments and discussion. This work was supported by grant DEB- 0643840 from the National Science Foundation to THO. Thanks to T Cronin, R

Smith, M Angelos, A Syme, and D Horne for providing identification assistance and/or specimens. Thanks to J Weaver for providing SEMs of specimens. Collection assistance came from I. Oakley, R. Lampe, A. Oakley, TA Oakley, C. Fong. Thanks also go to the Duke and BYU sequencing centers. N. Shaner and S. Haddock (supported by NIH GM087198) kindly provided unpublished 454 data for the halocyprid ostracod. B. Juarez assisted with collecting, specimen identification and some morphological data collection. T. Burmeister shared published EST data. Thanks to C. Cunningham, S. Haddock and Oakley lab for critical comments on earlier drafts. Thanks to M. Alexandrou and M. Caskey for suggesting 'allotrios'. We also acknowledge support from the Center for Scientific Computing at the CNSI and MRL: an NSF MRSEC (DMR-1121053) and NSF CNS-0960316. Author contributions: THO and AKZ collected and identified ostracod specimens. ARL and AKZ developed 454 protocols for Ostracoda with help from THO. ARL and THO collected ostracod rDNA data. THO performed bioinformatic analyses, with assistance from ARL, and AKZ. JMW collected and scored morphological characters, with assistance from THO, and ARL. All authors contributed to writing and all approved the final manuscript.

References

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

Abele LG, Spears T, Kim W, Applegate M. 1992. Phylogeny of Selected Maxillopodan and Other Crustacean Taxa Based on 18S Ribosomal Nucleotide Sequences: A Preliminary Analysis. Acta Zool. 73:373-382.

Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Anderson LI, Trewin NH. 2003. An Early Devonian arthropod fauna from the Windyfield cherts, Aberdeenshire, Scotland. Palaeontology 46:467-509.

Andrew DR. 2011. A new view of insect-crustacean relationships II. Inferences from expressed sequence tags and comparisons with neural cladistics. Arthropod Struct. Dev. 40:289-302.

Ayres DL, Darling A, Zwickl DJ, et al. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst. Biol. 61:170-173.

Babbitt C, Patel NH. 2005. Relationships within the Pancrustacea: examining the influence of additional malacostracan 18S and 28S rDNA. In: S Koenemann, RA Jenner, editors. Crustacea and arthropod relationships: CRC Press.

Bate RH. 1971. Phosphatized ostracods from the Cretaceous of Brazil. Nature 230:397-298.

Bate RH. 1972. Phosphatized ostracods with appendages from the Lower Cretaceous of Brazil. Palaeontology 15:379A-393.

Bate RH. 1973. On Pattersoncypris mocropapillosa Bate. Stereo Atlas of Ostracod Shells 1:101-108.

Berger SA, Stamatakis A. 2010. Accuracy of morphology-based phylogenetic fossil placement under Maximum Likelihood. Computer Systems and Applications (AICCSA), 2010 IEEE/ACS International Conference on. p. 1-9.

Bergström J, Briggs DEG, Dahl E, Rolfe WDI, Stürmer W. 1987. Nahecaris stuertzi, a phyllocarid crustacean from the Lower Devonian Hunsrück Slate. Paläontologische Zeitschrift 61:273-298.

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

Blair JE, Hedges SB. 2005. Molecular clocks do not support the Cambrian explosion. Mol. Biol. Evol. 22:387-390.

Boore JL, Lavrov DV, Brown WM. 1998. Gene translocation links insects and crustaceans. Nature 392:667-668.

Boxshall G. 1983. A comparative functional analysis of the major maxillopodan groups. In: S FR., editor. Crustacean Issues. Rotterdam: Balkema. p. 121-143.

Boxshall GA. 2007. Crustacean classification: on-going controversies and unresolved problems. Zootaxa 1668:313-325.

Briggs DEG, Sutton MD, Siveter DJ. 2004. A new phyllocarid (Crustacea: Malacostraca) from the Silurian Fossil-Lagerstätte of Herefordshire, UK. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 271:131-138.

Briggs DEG, Sutton MD, Siveter DJ, Siveter DJ. 2005. Metamorphosis in a Silurian barnacle. Proceedings. Biological sciences / The Royal Society 272:2365-2369.

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ. 1993. Partitioning and Combining Data in Phylogenetic Analysis. Syst. Biol. 42:384-397.

Castellani C, Maas A, Waloszek D, Haug JT. 2011. New pentastomids from the Late Cambrian of Sweden - deeper insight of the ontogeny of fossil tongue worms. Palaeontographica Abteilung A 293:95-145.

Chou H-H, Holmes MH. 2001. DNA sequence quality trimming and vector removal. Bioinformatics 17:1093-1104.

Conway Morris S. 2000. Special Feature: The Cambrian "explosion": Slow-fuse or megatonnage? Proc. Natl. Acad. Sci. USA 97:4426-4429.

Delsuc F, Phillips MJ, Penny D. 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?". Science 301:1482; author reply 1482.

Dohle W. 2001. Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name 'Tetraconata' for the monophyletic unit Crustacea + Hexapoda. Annales de la Société entomologique de France 37:85-103.

Driskell AC, Ané C, Burleigh JG, McMahon MM, O'meara BC, Sanderson MJ. 2004. Prospects for building the tree of life from large sequence databases. Science 306:1172-1174.

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

Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14:755-763.

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

Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science 334:1091-1097.

Felsenstein J. 1978. Cases in which Parsimony or Compatibility Methods will be Positively Misleading. Syst Biol 27:401-410.

Foote M, Sepkoski JJ. 1999. Absolute measures of the completeness of the fossil record. Nature 398:415-417.

Friedrich M, Tautz D. 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature 376:165-167.

Giardine B, Riemer C, Hardison RC, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 15:1451-1455.

Giribet G, Edgecombe GD, Wheeler WC. 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413:157-161.

Glenner H, Thomsen PF, Hebsgaard MB, Sørensen MV, Willerslev E. 2006. Evolution. The origin of insects. Science (New York, N.Y.) 314:1883-1884.

Guindon S, Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst. Biol. 52:696-704.

Harvey THP, Vélez MI, Butterfield NJ. 2012. Exceptionally preserved crustaceans from western Canada reveal a cryptic Cambrian radiation. Proc. Natl. Acad. Sci. USA 109:1589-1594.

Hessler RR, Newman WA. 1975. A trilobitomorph origin for the Crustacea. Fossils Strata 4:437-459.

Høeg JT, Kolbasov GA. 2002. Lattice organs in y-cyprids of the Facetotecta and their significance in the phylogeny of the Crustacea Thecostraca. Acta Zool. 83:67-79.

Horne DJ, Schon I, Smith RJ, Martens K. 2005. What are Ostracoda? A cladistic analysis of the extant super-families of the subclasses Myodocopa and Podocopa (Crustacea:Ostracoda). In: S Koenemann, RA Jenner, editors. Crustacea and arthropod relationships: CRC Press.

Hou X, Williams M, Siveter DJ, Siveter DJ, Aldridge RJ, Sansom RS. 2010. Soft-part anatomy of the Early Cambrian bivalved arthropods Kunyangella and Kunmingella: significance for the phylogenetic relationships of Bradoriida. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 277:1835-1841.

Hou XG, Siveter DJ, Williams M, Walossek D. 1996. Appendages of the arthropod Kunmingella from the early Cambrian of China: its bearing on the systematic position of the Bradoriida and the fossil record of the Ostracoda. Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 351:1131-1145.

Huys R, Boxshall GA. 1991. Copepod Evolution. London: Ray Society.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225-231.

Jenner RA. 2010. Higher-level crustacean phylogeny: consensus and conflicting hypotheses. Arthropod Struct. Dev. 39:143-153.

Kluge AG. 1989. A Concern for Evidence and a Phylogenetic Hypothesis of Relationships among Epicrates (Boidae, Serpentes). Syst Biol 38:7-25.

Koenemann S, Jenner RA, Hoenemann M, Stemme T, von Reumont BM. 2010. Arthropod phylogeny revisited, with a focus on crustacean relationships. Arthropod Struct. Dev. 39:88-110.

Kolaczkowski B, Thornton JW. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431:980-984.

Kolaczkowski B, Thornton JW. 2008. A mixed branch length model of heterotachy improves phylogenetic accuracy. Mol. Biol. Evol. 25:1054-1066.

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

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

Lartillot N, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 363:1463-1472.

Lavrov DV, Brown WM, Boore JL. 2004. Phylogenetic position of the Pentastomida and (pan)crustacean relationships. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 271:537-544.

Lindgren AR, Pankey MS, Hochberg FG, Oakley TH. 2012. A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. BMC Evol Biol 12:129.

Lyons-Weiler J, Hoelzer GA, Tausch RJ. 1998. Optimal outgroup analysis. Biol. J. Linn. Soc. 64:493-511.

Mallatt J, Giribet G. 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Mol. Phylogen. Evol. 40:772-794.

Mallatt JM, Garey JR, Shultz JW. 2004. Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. Mol. Phylogen. Evol. 31:178-191.

Martin JW, Davis G. 2001. An updated classification of the recent Crustacea. Science Series (Los Angeles) 39.

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

Misof B, Misof K. 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Syst. Biol. 58:21-34.

Müller KJ. 1964. Ostracoda (Bradorina) mit phosphatischen Gehäusen aus dem Oberkambrium von Schweden. Neues Jahrbuch für Geologie und Paläontologie (Abhandlungen) 121:1-46.

Müller KJ. 1983. Crustacea with preserved soft parts from the Upper Cambrian of Sweden. Lethaia 16:93-109.

Müller KJ, Walossek D. 1988. External morphology and larval development of the Upper Cambrian Bredocaris admirabilis (Crustacea: Maxillopoda). Fossils Strata 23:1-70.

Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F. 2003. Hexapod origins: monophyletic or paraphyletic? Science 299:1887-1889.

O'Leary MA, Kaufman S. 2011. MorphoBank: phylophenomics in the "cloud". Cladistics 27:529-537.

Oakley TH. 2003. On homology of arthropod compound eyes. Integr. Comp. Biol. 43:522-530.

Oakley TH. 2005. Myodocopa (Crustacea: Ostracoda) as models for evolutionary studies of light and vision: multiple origins of bioluminescence and extreme sexual dimorphism. Hydrobiologia 538:179-192.

Oakley TH, Cunningham CW. 2002. Molecular phylogenetic evidence for the independent evolutionary origin of an arthropod compound eye. Proc. Natl. Acad. Sci. USA 99:1426-1430.

Olesen J. 2009. Phylogeny of Branchiopoda (Crustacea) – Character Evolution and Contribution of Uniquely Preserved Fossils. Arthropod Systematics & Phylogeny 67:3-39.

Pérez-Losada M, Høeg JT, Crandall KA. 2004. Unraveling the evolutionary radiation of the thoracican barnacles using molecular and morphological evidence: a comparison of several divergence time estimation approaches. Syst. Biol. 53:244-264.

Pyron RA. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst. Biol. 60:466-481.

Regier JC, Shultz JW, Ganley ARD, et al. 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Syst. Biol. 57:920-938.

Regier JC, Shultz JW, Kambic RE. 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 272:395-401.

Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463:1079-1083.

Richter S. 2002. The Tetraconata concept: hexapod-crustacean relationships and the phylogeny of Crustacea. Org. Divers. Evol. 2:217-237.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56:389-399.

Roeding F, Borner J, Kube M, Klages S, Reinhardt R, Burmester T. 2009. A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (Pandinus imperator). Mol. Phylogen. Evol. 53:826-834.

Rota-Stabelli O, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, Pisani D, Philippe H, Telford MJ. 2011. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 278:298-306.

Sanders KL, Lee MSY. 2010. Arthropod molecular divergence times and the Cambrian origin of pentastomids. Syst. Biodivers. 8:63-74.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301-302.

Schram FR, Hof CHJ. 1998. Fossils and the Interrelationships of Major Crustacean Groups. In: GD Edgecombe, editor. Arthropod fossils and phylogeny: Columbia University Press.

Scourfield DJ. 1926. On a new type of crustacean from the Old Red Sandstone (Rhynie Chert Bed, Aberdeenshire)-Lepidocaris rhyniensis, gen. et sp. nov. Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 214:153-187.

Scourfield DJ. 1940. Two new and nearly complete specimens of young stages of the Devonian fossil crustacea Lepidocaris rhyniensis. Proceedings of the Linnean Society of London 152:290-298.

Selden PA, Huys R, Stephenson MH, Heward AP, Taylor PN. 2010. Crustaceans from bitumen clast in Carboniferous glacial diamictite extend fossil record of copepods. Nature communications 1:50.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114-1116.

Shultz JW, Regier JC. 2000. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 267:1011-1019.

Siveter DJ, Briggs DEG, Siveter DJ, Sutton MD. 2010. An exceptionally preserved myodocopid ostracod from the Silurian of Herefordshire, UK. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 277:1539-1544.

Siveter DJ, Sutton MD, Briggs DEG. 2007. Brood care in a Silurian ostracod. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 274:465-469.

Siveter DJ, Sutton MD, Briggs DEG, Siveter DJ. 2003. An ostracode crustacean with soft parts from the Lower Silurian. Science 302:1749-1751.

Siveter DJ, Waloszek D, Williams M. 2003. An early Cambrian phosphatocopid crustacean with three-dimensionally preserved soft parts from Shropshire, England. In: PD Lane, DJ Siveter, RA Fortey, editors. Trilobites and Their Relatives: Wiley-Blackwell. p. 9-30.

Siveter DJ, Williams M, Waloszek D. 2001. A phosphatocopid crustacean with appendages from the Lower Cambrian. Science 293:479-281.

Smith RJ. 2000. Morphology and ontogeny of Cretaceous ostracods with preserved appendages from Brazil. Palaeontology 43:63-98.

Sonnhammer ELL, Koonin EV. 2002. Orthology, paralogy and proposed classification for paralog subtypes. Trends Genet. 18:619-620.

Spears T, Abele LG. 1998. Crustacean phylogeny inferred from 18S rDNA. Arthropod relationships, Fortey, R.A.Thomas, R.H. (eds.).- London (United Kingdom): Chapman and Hall Ltd.

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

Strausfeld NJ. 2011. Some observations on the sensory organization of the crustaceomorph Waptia fieldensis Walcott. Palaeontographica Canadiana 31:157-168.

Strausfeld NJ, Andrew DR. 2011. A new view of insect-crustacean relationships I. Inferences from neural cladistics and comparative neuroanatomy. Arthropod Struct. Dev. 40:276-288.

Sylvester-Bradley PC. 1961. Archaeocopida. . In: RC Moore, editor. Treatise on Invertebrate Paleontology, Part Q, Arthropoda 3. : Geological Society of America and University of Kansas Press. p. Q100-105.

Syme AE, Oakley TH. 2012. Dispersal between Shallow and Abyssal Seas and Evolutionary Loss and Regain of Compound Eyes in Cylindroleberidid Ostracods: Conflicting Conclusions from Different Comparative Methods. Syst. Biol. 61:314-336.

Tanaka G. 2005. Morphological design and fossil record of the podocopid ostracod naupliar eye. Hydrobiologia 538:231-242.

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

Tinn O, Meidla T. 2001. Middle Ordovician ostracods from the Lanna and Holen Limestones, south-central Sweden. Gff 123:129-136.

Tinn O, Oakley TH. 2008. Erratic rates of molecular evolution and incongruence of fossil and molecular divergence time estimates in Ostracoda (Crustacea). Mol. Phylogen. Evol. 48:157-167.

von Reumont BM, Jenner RA, Wills MA, et al. 2012. Pancrustacean phylogeny in the light of new phylogenomic data: support for remipedia as the possible sister group of hexapoda. Mol. Biol. Evol. 29:1031-1045.

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

Wakayama N. 2007. Embryonic development clarifies polyphyly in ostracod crustaceans. J. Zool. 273:406-413.

Walcott CD. 1912. Middle Cambrian Branchiopoda, Malacostraca, Trilobita, and Merostomata. Smithsonian Miscellaneous Collections 57:145-228.

Walossek D. 1993. The Upper Cambrian Rehbachiella kinnekullensis Müller, 1983, and the phylogeny of Branchiopoda and Crustacea. Fossils Strata 32:1-202.

Walossek D. 1995. The Upper Cambrian Rehbachiella, its larval development, morphology and significance for the phylogeny of Branchiopoda and Crustacea. Hydrobiologia 298:1-13.

Walossek D, Müller KJ. 1994. Pentastomid parasites from the Lower Palaeozoic of Sweden. Transactions of the Royal Society of Edinburgh: Earth Sciences 85:1-37.

Ware JL, Grimaldi DA, Engel MS. 2010. The effects of fossil placement and calibration on divergence times and rates: an example from the termites (Insecta: Isoptera). Arthropod Struct. Dev. 39:204-219.

Weitschat W. 1983a. On Triadocypris spitzbergensis Weitschat. Stereo-Atlas of Ostracod Shells 10:127-138.

Weitschat W. 1983b. Ostracoden (O. Myodocopida) mit Weichkörper-Erhaltung aus der Unter-Trias von Spitzbergen. Paläontologische Zeitschrift 57:309-323.

Wheeler WC, Whiting M, Wheeler QD, Carpenter JM. 2001. The Phylogeny of the Extant Hexapod Orders. Cladistics 17:113-169.

Wills MA, Briggs DEG, Fortey RA, Wilkinson M, Sneath PHA. 1998. An arthropod phylogeny based on recent and fossil taxa. In: GD Edgecombe, editor. Arthropod fossils and phylogeny: Columbia University Press. p. 347.

Wingstrand KG. 1972. Comparative spermatology of a pentastomid, Raillietiella hemidactyli, and a branchiuran crustacean, Argulus foliaceus, with a discussion of pentastomid relationships. Kongelige Danske Videnskabernes Selskab Biologiske Skrifter 19:1-72.

Wray GA, Levinton JS, Shapiro LH. 1996. Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla. Science 274:568-573.

Zhang H, Dong X. 2009. Two new species of Vestrogothia (Phosphatocopina, Crustacea) of Orsten-type preservation from the Upper Cambrian in western Hunan, South China. Science in China Series D: Earth Sciences 52:784-796.

Zhang X-g, Siveter DJ, Waloszek D, Maas A. 2007. An epipodite-bearing crown-group crustacean from the Lower Cambrian. Nature 449:595-598.

Zrzavy J, Mihulka S, Kepka P, Bezdek A, Tietz D. 1998. Phylogeny of the Metazoa Based on Morphological and 18S Ribosomal DNA Evidence. Cladistics 14:249-285.

Zrzavy J, Stys P. 1997. The basic body plan of arthropods: insights from evolutionary morphology and developmental biology. J. Evol. Biol. 10:353-367.

Table 1. Collection information for material processed for 454 pyrosequencing.

Species (Suborder)	Locality	Method	Latitude	Longitude	Date(s)	Depth
Actinoseta jonesi **(Myodocopina)**	Cayo Enrique, La Parguera, Puerto Rico	net collecting	17°57.335'N	67°03.185'W	September 12th, 2010	2-3m
Argulus **sp.**	Purchased from Gulf Coast Marine Specimens	Picked from live fish in collections	?	?	?	?
Conchoecissa **sp. (Halocypridina)**	Trawl Lower Sur Canyon on R/V Western Flyer	Trawl on Western Flyer	36.06° N	122.29 W	December 10th, 2009	
Cytherelloidea californica **(Platycopina)**	Camino de la Costa Beach Access, La Jolla, San Diego	algae collecting	24º46.9'N	80º54.58'W	May 14th, 2010	intertidal only on very low tide
Vestalenula **sp. (Darwinulocopina)**	Freshwater Puddle, Isla Colon, Bocas del Toro, Panama	net collecting	9º21.17'N	82º15.45'W	July 29th, 2009	10cm
Euphilomedes morini **(Myodocopina)**	Stern's Wharf Pier, Santa Barbara	Eckman grab	34º24.4'N	119º40.5'W	Oct., Nov., 2008	10m
Puriana **sp. (Cytherocopina)**	Isla Colon, Bocas del Toro, Panama	net collecting	9º21'N	82º15.45'W	July 23rd, 2009	1m
Heterocypris **sp. (Cypridocopina)**	Temporary Pool, More Mesa, Santa Barbara, CA	net collecting	34º25.23'N	119º47.29'W	?	10cm
Skogsbergia	Duck Key	bait trap	24º46.9'N	80º54.58'W	July 16th thru	2-3m

lerneri **(Myodocopina)**	Viaduct, FL				July 18th, 2009	
Vargula tsujii **(Myodocopina)**	Fishermen's Cove, Twin Harbors, Catalina Island, CA	bait trap	33º26.66'N	118º29.34'W	July 10th and 11th, 2009	5-10m

Table 2 – Fossils analyzed in this study

Genus (abbreviation)	Clade	Localities	Period	Stage [a]	Age (mya) [a]	References
***Bredocaris admirabilis* (Ba)**	Allotriocarida	Orsten	Cambrian	Paibian	501	(Müller 1983; Müller, Walossek 1988)
***Cinerocaris magnifica* (Cm)**	Multicrustacea	Herefordshire	Silurian	Wenlock	425	(Briggs, Sutton, Siveter 2004)
***Colymbosathon ecplecticos* (Ce)**	Oligostraca (Ostracoda)	Herefordshire	Silurian	Wenlock	425	(Siveter, Waloszek, Williams 2003)
***Klausmuelleria salopensis* (Ks)**	Multicrustacea	Comley	Cambrian	Toyonian	514-511	(Siveter, Williams, Waloszek 2001; Siveter, Waloszek, Williams 2003)
***Kunmingella* sp. (K)**		Chengjiang	Cambrian	Atdabanian	525	(Hou et al. 1996; Hou et al. 2010)
***Kunyangella cheni* (Kc)**		Chengjiang	Cambrian	Atdabanian	525	(Hou et al.)
***Lepidocaris rhyniensis* (Lr)**	Allotriocarida	Rhynie Chert, Windyfield Chert	Devonian	Pragian	410-396	(Scourfield 1926; Scourfield 1940; Anderson, Trewin 2003)
***Nahecaris stuertzi* (Ns)**	Multicrustacea	Hunsrück Slate	Devonian	Emsian	392-388	(Bergström et al. 1987)
***Nasunaris flata* (Nf)**	Oligostraca (Ostracoda)	Herefordshire	Silurian	Wenlock	425	(Siveter et al. 2010)
***Nymphatelina gravida* (Ng)**	Oligostraca (Ostracoda)	Herefordshire	Silurian	Wenlock	425	(Siveter, Sutton, Briggs 2007)
***Pattersoncypris micropapillosa* (Pm)**	Oligostraca (Ostracoda)	Santana	Cretaceous	Aptian/Albian	108-92	(Bate 1971; Bate 1972; Bate 1973; Smith 2000)
***Rehbachiella kinnekullensis* (Rk)**	Allotriocarida	Orsten	Cambrian	Paibian	501	(Müller 1983; Walossek 1993; Walossek 1995)
***Triadocypris spitzbergensis* (Ts)**	Oligostraca (Ostracoda)	Spitzbergen	Triassic	Lower Triassic	251-245	(Weitschat 1983b)
***Vestrogothia* sp. (Vs)**	Multicrustacea	Bitiao Formation,	Cambrian	Paibian	501	(Müller 1964; Zhang, Dong

		Orsten				2009)
Waptia fieldensis **(Wf)**	Multicrustacea	Burgess Shale, Wheeler Shale	Cambrian	Stage 5/Drumian	505	(Walcott 1912; Strausfeld 2011)
Yicaris dianensis **(Yd)**	Allotriocarida	Yunnan	Cambrian	Atdabanian	525-520	(Zhang et al. 2007)

[a] of type locality

Table 3 - Analyses exploring phylogenetic topology of extant species using different subsets of the total data matrix

Dataset Name	N Species	N Characters	N Genes[a]	% Gaps
Extant Total	48	265388	967	80.2
Extant Slow 2.5	48	177332	631	78.4
Extant Slow 2.0	48	135826	490	76.8
Nuclear Proteins Slow 2	47	127506	483	76.8
Multicrustacea	17	60551	255	70.5
Hexapod Sister	18	259661	958	64.1
Oligostraca Ostracod-Restricted	17	58357	265	67.3
Oligostraca Ostracod-Extended	51	59033	267	86.29[b]

[a] To to be included in the data set, we required that a gene be present in 6 or more species (4 or more for Oligostraca Ostracod-Extended), and its alignment contain 50 or more characters. All data sets except 'Nuclear Proteins Slow 2' include morphology and rDNA characters.

[b] We added exemplars of ostracod families that only have 18S and morphological data available, increasing the proportion of missing data in the matrix.

Table 4 - Analyses exploring placement of fossils in pancrustacean phylogeny

Dataset/Analysis Name	N Species (fossils)	N Characters	N Genes*	% Gaps
Hexsister with Fossils	22 (4)	259661	945	70.6
Multicrustacea with Fossils	22 (5)[a]	60563	256	77.2
Oligostraca Restricted with Fossils	26 (9)[a]	58359	261	78.6
Oligostraca Extended with Fossils	60 (9)[a]	58857	261	88.3
Site-Weighted Fossils	64 (16)	N/A	N/A	N/A

[a]We analyzed Phosphatocopina with both multicrustacea, as suggested by Site-Weighted fossil analysis, and with Oligostraca based on their oft-cited affinity with Ostracoda. These numbers include the two Phosphatocopina fossils.

Table 5 – Calibration points from fossils placed consistently in our analyses.

Node	Fossil(s)	Min Age	Max Age	Descendent Clade 1	Molecular Representative 1	Descendent Clade 2	Molecular Representative 2
Myodocopa	*Colymbosathon*, *Nasunaris*, *Nymphatelina*	425 (Herefordshire)	505 (Burgess Shale)	Myodocopida	*Actinoseta jonesi*	Halocyprida	*Conchoecissa* sp.
Myodocopida	*Colymbosathon*, *Nasunaris*	425 (Herefordshire)	505 (Burgess Shale)	Cylindroleberididae	*Actinoseta jonesi*	Sarsielloidea	*Euphilomedes morini*
Malacostraca	*Cinerocaris*	425 (Herefordshire)		Leptostraca	*Nebalia hessleri*	Decapoda	*Libinia emarginata*
Branchiopoda	*Lepidocaris*	396 (Rhynie Chert)	> 510 (Harvey et al 2012)	Cladocera	*Daphnia pulex*	Anostraca	*Streptocephalus seali*
Anostraca	*Lepidocaris*	396 (Rhynie Chert)		Artemiidae	*Artemia salina*	Streptocephalidae	*Streptocephalus seali*

Table 6 - Analyses exploring effects of missing data in Oligostraca

				Clade Support (Bootstrap Proportions)*						
Data Retained	Species	Characters	% Gaps	Ost	Myd	Mya	S+C	C+V	Oli	Ich
All Genes	17	131112	81%	80	100	100	80	34	100	88
Genes present in >5 sp	17	45179	63%	70	100	100	78	40	100	88
Genes present in >10 sp	17	9605	44%	52	100	100	76	4	100	0
All Species	17	131112	81%	64	100	100	76	62	100	78
Species with >25 partitions	15	57939	64%	100	100	100	76	66	100	N/A
Species with >50 partitions	13	54808	61%	100	100	100	N/A	N/A	N/A	N/A
Species with >100 partitions	9	45931	48%	N/A	N/A	100	N/A	N/A	N/A	N/A
Species with >200 partitions	6	34298	37%	N/A	N/A	100	N/A	N/A	N/A	N/A

*Clade Abbreviations: Ost=Ostracoda, Myd=Myodocopida; Mya=Myodocopa; S+C=Sarsielloidea + Cypridinidae; C+V = *Cytherelloidea* + *Vestalenula*; Oli=Oligostraca; Ich=Ichthyostraca

Figure Legends

Figure 1. Maximum likelihood phylogenetic analyses of extant pancrustaceans based on concatenated protein coding, rDNA, and/or morphological data sets. Numbers at nodes represent bootstrap values (based on 100 replicates). Top left is values from the 'Extant Full' data set. Top right is values when excluding the fastest evolving genes (those with a summed branch length in proteome-species of 2.5 or more), we call this the 'slow 2.5' data set. Bottom

left of each node are bootstrap values using the slowest evolving protein coding genes, which we call 'slow 2.0', and we display the topology from this analysis in the figure. All three of those analyses include rDNA and morphological data. On the bottom right of each node are bootstrap values for nuclear proteins only, excluding mitochondrial proteins, rDNA, and morphology (note no nuclear proteins are available for *V. hilgendorfii*, but this was included in the overall analysis as the only ostracod with a fully sequenced mitochondrial genome). All data sets require a gene to be present in more than six species, otherwise that gene is excluded. The circled 4's are placed next to species with new 454 data, a G in a square is placed next to species for which we analyzed predicted proteomes from full Genome sequences.

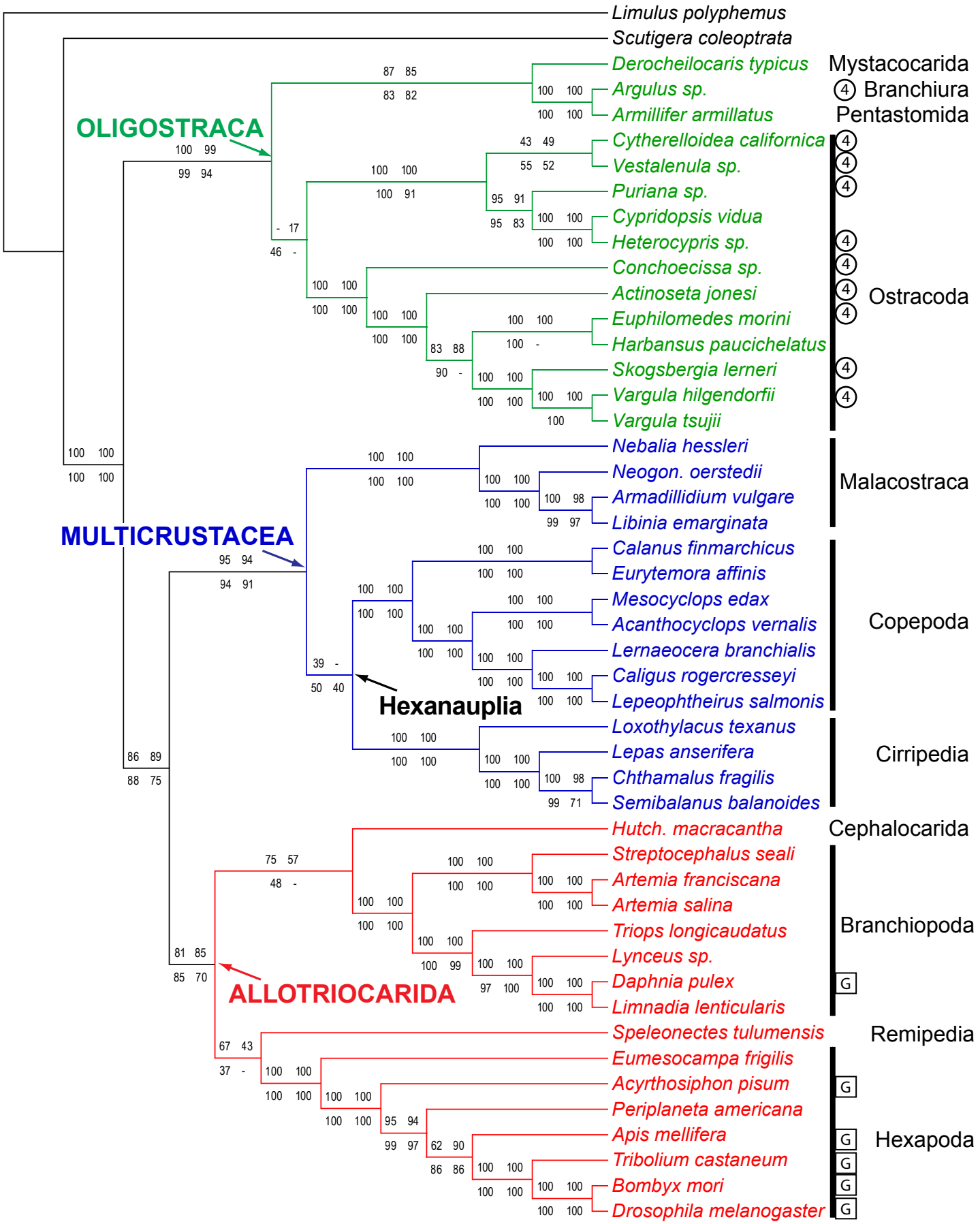
Figure 2. Separate maximum likelihood (ML) phylogenetic analyses of three epic pancrustacean clades based on concatenated protein coding, rDNA, and morphological data: **A**. Oligostraca **B**. Multicrustacea **C**. Allotriocarida. Each analysis was performed using *Limulus polyphemus* and *Scutigera coleoptrata* as outgroups, indicated by an O inside a hexagon at the root. We display the ML topology and branch lengths with support values from 100 bootstrap replicates below each node.

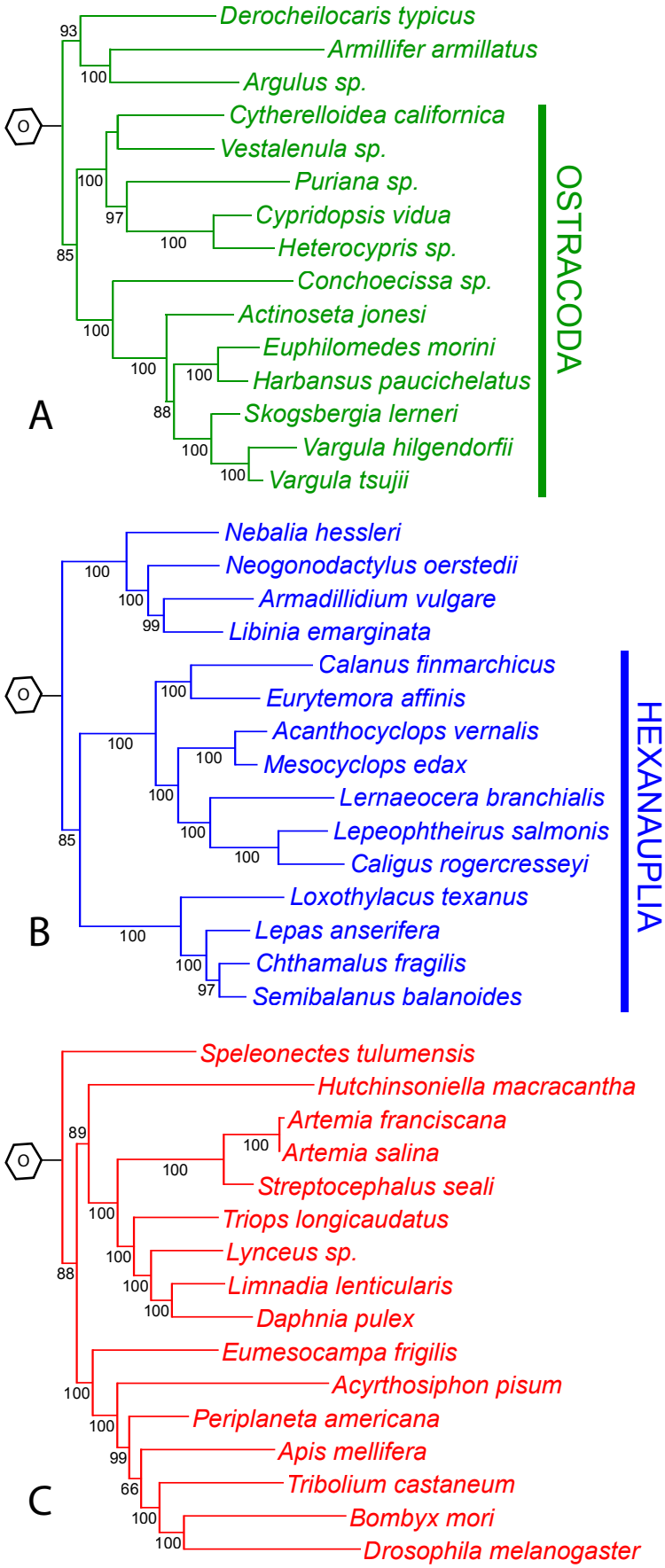
Figure 3. Maximum likelihood phylogeny of Oligostraca based on concatenated protein coding, rDNA, and morphological data, with taxon sampling in Ostracoda extended to included exemplars of families with rDNA. We display the ML topology and branch lengths with support values from 100 bootstrap replicates below each node (support values below 60% are not shown). The circled 4's are placed next to species with new 454 data.

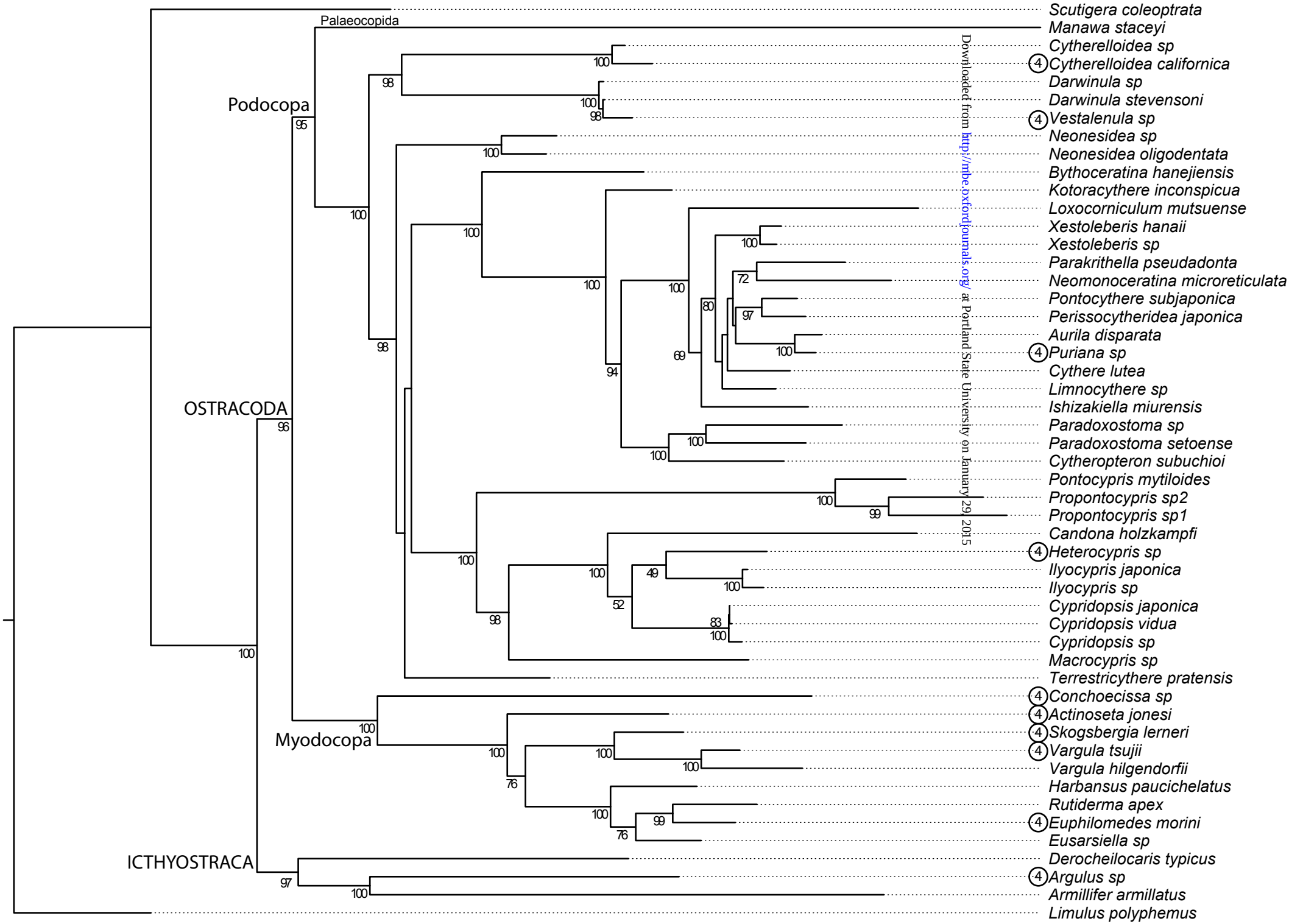
Figure 4. Summary of two different fossil placement analyses. Black squares represent placement of pancrustacean fossils based on concatenated analyses of morphological, protein coding, and rDNA data, with each of the three major clades analyzed separately, using

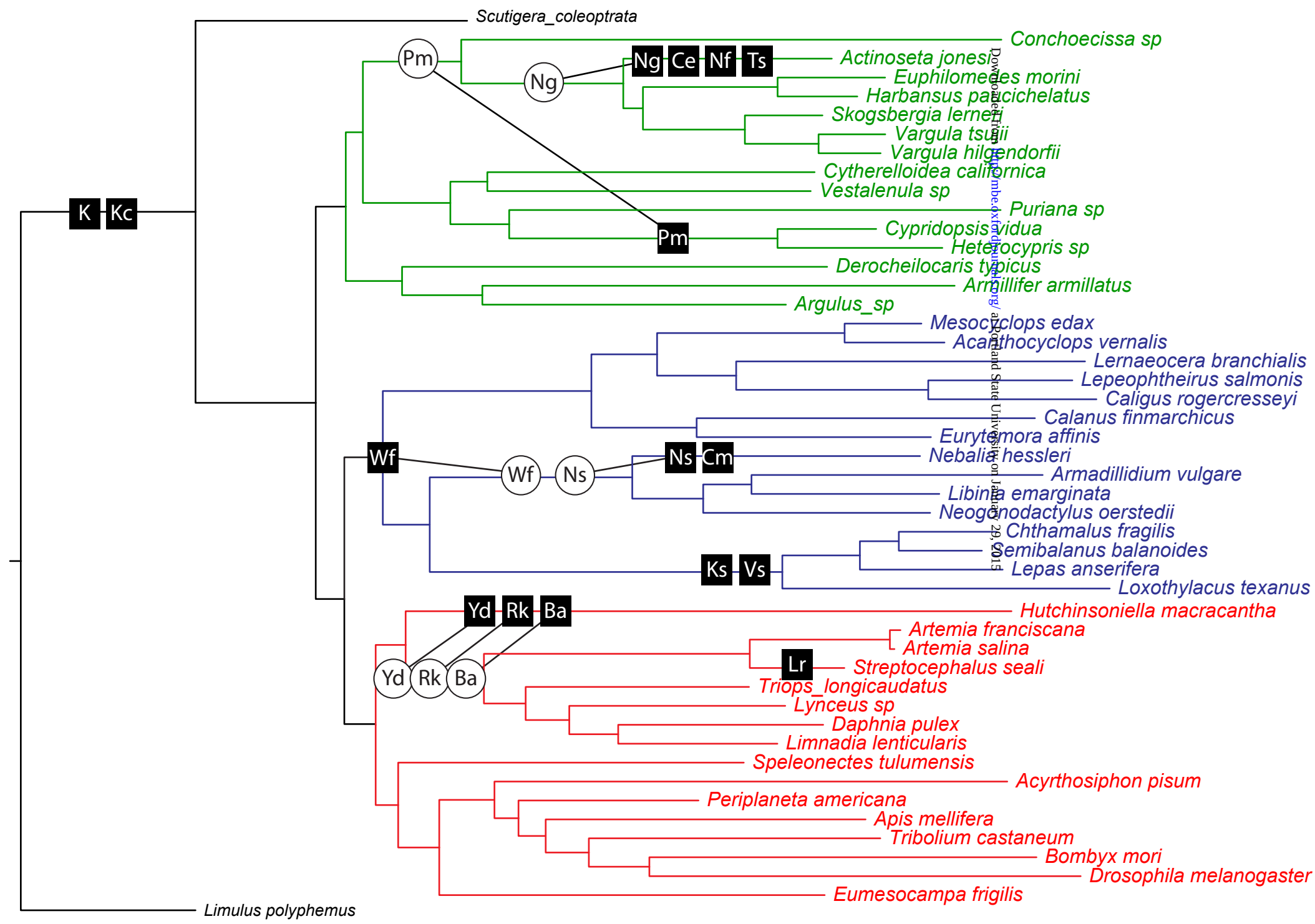
Scutigera and *Limulus* as outgroups. White circles indicate fossils that placed differently in a site-weighted fossil placement analysis (Berger, Stamatakis 2010). For this analysis, we used our ML tree from the 'slow 2.0' analysis. The algorithm determines weights for each binary character based on congruence with the molecular tree, then maximizes the placement of each fossil on the tree using ML. Fossil abbreviations are listed in Table 2.

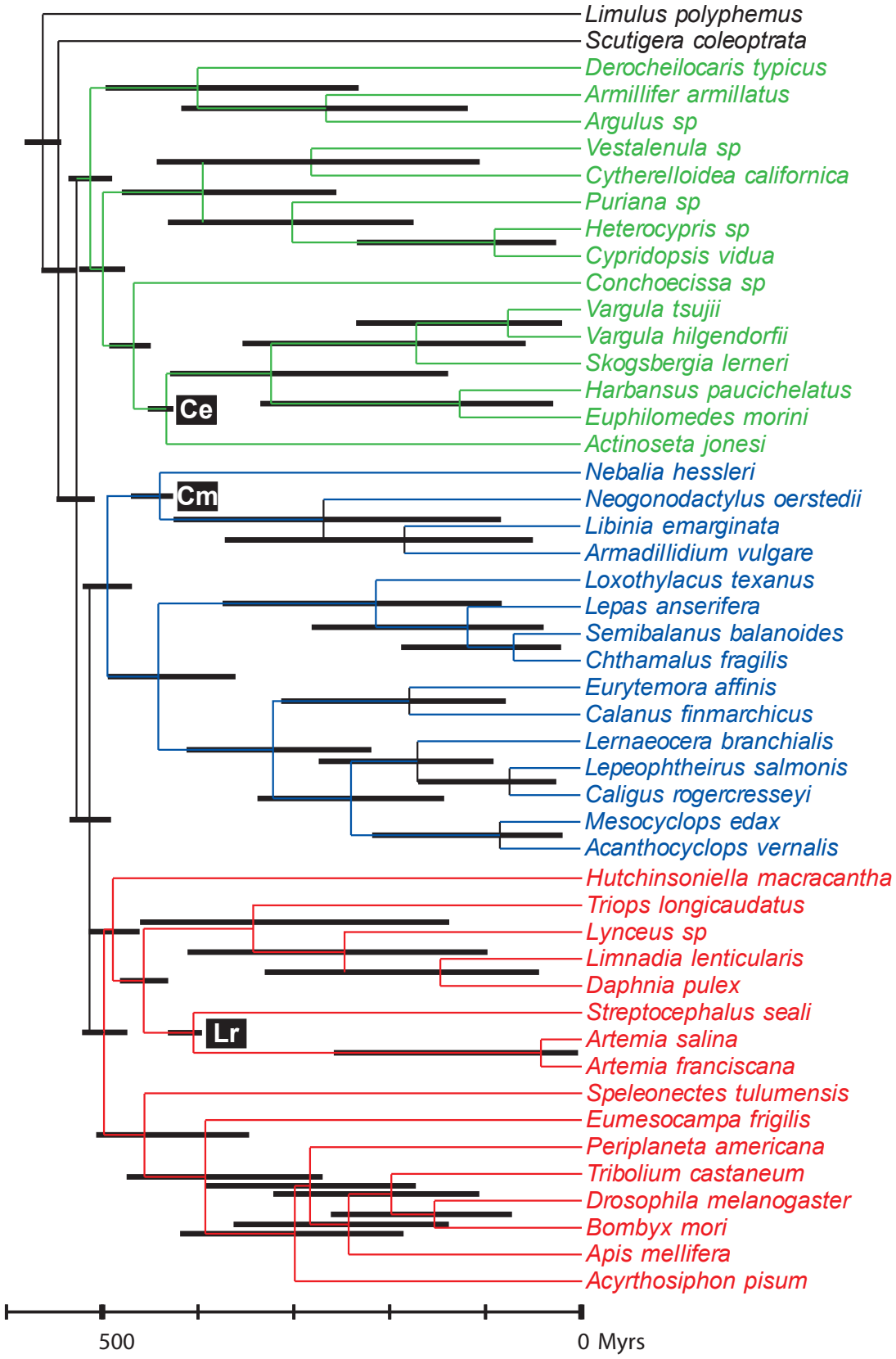
Figure 5. Bayesian analysis of divergence times using Phylobayes (Lartillot, Lepage, Blanquart 2009). We used three fossil calibrations, which were placed reliably with phylogenetic analyses (Fig. 4, Table 5). The three fossils are indicated on the tree with abbreviations, *Cinerocaris magnifica* (Cm), *Colymbosathon ecplecticos* (Ce), and *Lepidocaris rhyniensis* (Lr). We used an uncorrelated gamma model to relax the assumption of a molecular clock, with additional details in Methods. Black bars on nodes represent 95% confidence intervals on divergence times.


Limulus polyphemus
Scutigera coleoptrata
Derocheilocaris typicus
Argulus sp.
Armillifer armillatus
Mystacocarida
4 Branchiura
Pentastomida
OLIGOSTRACA
Cytherelloidea californica
Vestalenula sp.
Puriana sp.
Cypridopsis vidua
Heterocypris sp.
Conchoecissa sp.
Actinoseta jonesi
Euphilomedes morini
Harbansus paucichelatus
Skogsbergia lerneri
Vargula hilgendorfii
Vargula tsujii
Ostracoda
Nebalia hessleri
Neogon. oerstedii
Armadillidium vulgare
Libinia emarginata
Malacostraca
MULTICRUSTACEA
Calanus finmarchicus
Eurytemora affinis
Mesocyclops edax
Acanthocyclops vernalis
Lernaeocera branchialis
Caligus rogercresseyi
Lepeophtheirus salmonis
Copepoda
Hexanauplia
Loxothylacus texanus
Lepas anserifera
Chthamalus fragilis
Semibalanus balanoides
Cirripedia
Hutch. macracantha
Cephalocarida
Streptocephalus seali
Artemia franciscana
Artemia salina
Triops longicaudatus
Lynceus sp.
Daphnia pulex
Limnadia lenticularis
Branchiopoda
ALLOTRIOCARIDA
Speleonectes tulumensis
Remipedia
Eumesocampa frigilis
Acyrthosiphon pisum
Periplaneta americana
Apis mellifera
Tribolium castaneum
Bombyx mori
Drosophila melanogaster
Hexapoda








Limulus polyphemus
Scutigera coleoptrata
Derocheilocaris typicus
Armillifer armillatus
Argulus sp
Vestalenula sp
Cytherelloidea californica
Puriana sp
Heterocypris sp
Cypridopsis vidua
Conchoecissa sp
Vargula tsujii
Vargula hilgendorfii
Skogsbergia lerneri
Harbansus paucichelatus
Euphilomedes morini
Actinoseta jonesi
Ce
Nebalia hessleri
Neogonodactylus oerstedii
Libinia emarginata
Armadillidium vulgare
Cm
Loxothylacus texanus
Lepas anserifera
Semibalanus balanoides
Chthamalus fragilis
Eurytemora affinis
Calanus finmarchicus
Lernaeocera branchialis
Lepeophtheirus salmonis
Caligus rogercresseyi
Mesocyclops edax
Acanthocyclops vernalis
Hutchinsoniella macracantha
Triops longicaudatus
Lynceus sp
Limnadia lenticularis
Daphnia pulex
Streptocephalus seali
Artemia salina
Artemia franciscana
Lr
Speleonectes tulumensis
Eumesocampa frigilis
Periplaneta americana
Tribolium castaneum
Drosophila melanogaster
Bombyx mori
Apis mellifera
Acyrthosiphon pisum
500
0 Myrs