Exploring phylogenomic relationships within Myriapoda: should high matrix occupancy be the goal?

ROSA FERNÁNDEZ[1], GREGORY D. EDGECOMBE[2] AND GONZALO GIRIBET[1]

[1]*Museum of Comparative Zoology & Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA*

[2]*Department of Earth Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*

1

Abstract.—Myriapods are one of the dominant terrestrial arthropod groups including the diverse and familiar centipedes and millipedes. Although molecular evidence has shown that Myriapoda is monophyletic, its internal phylogeny remains contentious and understudied, especially when compared to those of Chelicerata and Hexapoda. Until now, efforts have focused on taxon sampling (e.g., by including a handful of genes in many species) or on maximizing matrix occupancy (e.g., by including hundreds or thousands of genes in just a few species), but a phylogeny maximizing sampling at both levels remains elusive. In this study, we analyzed forty Illumina transcriptomes representing three myriapod classes (Diplopoda, Chilopoda and Symphyla); twenty-five transcriptomes were newly sequenced to maximize representation at the ordinal level in Diplopoda and at the family level in Chilopoda. Eight supermatrices were constructed to explore the effect of several potential phylogenetic biases (e.g., rate of evolution, heterotachy) at three levels of mean gene occupancy per taxon (50%, 75% and 90%). Analyses based on maximum likelihood and Bayesian mixture models retrieved monophyly of each myriapod class, and resulted in two alternative phylogenetic positions for Symphyla, as sister group to Diplopoda + Chilopoda, or closer to Diplopoda, the latter hypothesis having been traditionally supported by morphology. Within centipedes, all orders were well supported, but two nodes remained in conflict in the different analyses despite dense taxon sampling at the family level, situating the order Scolopendromorpha as sister group to a morphologically-anomalous grouping of Lithobiomorpha + Geophilomorpha in a subset of analyses. Interestingly, this anomalous result was obtained for all analyses conducted with the most complete matrix (90% of occupancy), being at odds not only with the sparser but more gene-rich supermatrices (75% and 50% supermatrices) or with the matrices optimizing phylogenegic informativeness and the most conserved genes, but also with previous hypotheses based on morphology, development or other molecular data sets. We discuss the implications of these findings in the context of the ever more prevalent quest for completeness in phylogenomic studies. [Chilopoda; Diplopoda; Symphyla: gene tree; species tree; node calibration; missing data.]

The status and interrelationships of Myriapoda have been major questions in arthropod systematics for more than a century (see a recent review in Edgecombe 2011). Variably argued to be mono-, para- or polyphyletic, molecular systematics has decisively weighed in in favor of myriapod monophyly (Regier et al. 2008; Regier et al. 2010; Rehm et al. 2014). The closest relative of Myriapoda remains a topic of discussion, though most recent analyses of arthropod phylogeny support an insect-crustacean clade named Pancrustacea or Tetraconata as sister group of Myriapoda under the Mandibulata hypothesis (Regier et al. 2010; Rota-Stabelli et al. 2011; Zwick et al. 2012; Rehm et al. 2014), a result also supported in earlier combined analyses of molecules and morphology (e.g., Giribet et al. 2001). The main alternative is an alliance between Myriapoda and Chelicerata (Paradoxopoda or Myriochelata hypotheses), retrieved in many analyses in the 1990s and 2000s. Debate has now largely shifted to the interrelationships between the four main clades of Myriapoda – Chilopoda (centipedes), Diplopoda (millipedes), Pauropoda and Symphyla – and the branching patterns of the two diverse groups, Chilopoda and Diplopoda.

The most widely endorsed phylogenetic framework for myriapods based on morphology and development divides the group into Chilopoda and a putative clade named Progoneata based on its anteriorly situated gonopore. The standard resolution within Progoneata is Symphyla as sister group of Pauropoda and Diplopoda, the latter two grouped as Dignatha (Dohle 1980, 1997; Edgecombe 2011). Progoneata has found some support from molecular phylogenetics (Gai et al. 2008; Regier et al. 2010) but it has also been contradicted by a grouping of Chilopoda and Diplopoda (Rehm et al. 2014). The most novel hypothesis to emerge from molecular data is an unanticipated union of pauropods and symphylans (Dong et al. 2012), a group that has been formalized as Edafopoda (Zwick et al. 2012).

Pauropoda and Symphyla are relatively small groups (835 and 195 species, respectively), with few taxonomic specialists. This contrasts with the more diverse and intensely studied Chilopoda and Diplopoda (Fig. 1). Of these, Chilopoda is the more extensively analyzed phylogenetically, the relationships

within its four diverse orders having been evaluated based on morphology and targeted sequencing of a few markers (e.g., Giribet and Edgecombe 2013; Vahtera et al. 2013; Bonato et al. 2014a). The interrelationships of the five extant centipede orders had converged on a morphological solution (Edgecombe and Giribet 2004) that differs from the first transcriptomic analysis to sample all five orders (Fernández et al. 2014b) in one key respect: morphology unites Craterostigmomorpha (an order composed of two species of the genus *Craterostigmus*) with the two orders that have strictly epimorphic development, whereas molecules assign the order Lithobiomorpha to that position. Millipedes are the most diverse group of myriapods, with 7753 (Shear 2011) to more than 12,000 (Brewer et al. 2012) named species classified in ca. 16 orders (Brewer and Bond 2013). Millipede phylogeny has until recently been inferred based on small sets of morphological characters (Enghoff 1984; Sierwald et al. 2003; Blanke and Wesener 2014), but some analyses of a few nuclear protein-coding genes have allowed traditional ordinal-level systematics to be tested with molecular data (Regier et al. 2005; Miyazawa et al. 2014).

Until very recently, analyses of arthropod phylogeny based on transcriptomic data have included very few myriapods, typically one species of Diplopoda and one of Chilopoda (Meusemann et al. 2010; Andrew 2011). This undersampling of myriapod diversity is now being rectified, with the first phylogenomic studies aimed at testing major hypotheses about diplopod and chilopod phylogeny appearing. Relationships between eight of the 16 recognized millipede orders have been inferred using 221 genes sampled for nine species (Brewer and Bond 2013), and interrelations of the five extant chilopod orders have been estimated using transcriptomes for seven species and up to 1,934 genes (Fernández et al. 2014b). Rehm et al. (2014) appraised the position of Myriapoda within Arthropoda using up to 181 genes from 8 myriapods, including the first transcriptomic data for Symphyla and some key millipede groups, notably Penicillata (bristly millipedes) and Glomerida (pill millipedes), the latter millipede orders not sampled by Brewer and Bond (2013). However, none of these studies included a dense taxonomic sampling across Myriapoda: to

date, the maximum number of myriapods included in phylogenomic studies is ten, one centipede and nine millipedes (Brewer and Bond 2013).

Myriapods include the oldest fossil remains of terrestrial animals, with millipedes and centipedes both having records as far back as the Silurian (Wilson and Anderson 2004; Shear and Edgecombe 2010), and trace fossils indicate the activity of millipedes in the Ordovician (Wilson 2006). Silurian millipedes exhibit the earliest direct evidence for air breathing in the form of spiracles (Wilson and Anderson 2004). This geological antiquity has brought Myriapoda to the forefront in considerations of the timing of terrestrialisation in animals and plants (Kenrick et al. 2012). Such questions of timing naturally intersect with phylogenetics in the realm of molecular dating. Recent timetrees for myriapods in the context of arthropods as a whole estimate the origin of Myriapoda by the early Cambrian and the divergence of its four classes in the late Cambrian (Rota-Stabelli et al. 2011; Rehm et al. 2014), thus predicting substantial gaps in the early fossil record of the clade.

The open questions about how the main groups of myriapods are related however limit how we interpret the timing of the diversification of a major component of the soil arthropod biota. We contribute to this discussion by presenting a large injection of novel Illumina transcriptome data for myriapods, including previously unsampled millipede orders and a substantially expanded taxonomic coverage for centipedes. Furthermore, we investigate missing data in relation to potential confounding factors in phylogenomic reconstruction (e.g., evolutionary rate, compositional heterogeneity, long branch attraction and heterotachy) to dissect the individual effect of each factor on the recovered phylogenies. Concurrently, we present a morphological character set for the same set of species as sampled transcriptomically and code a set of key fossil species for their preserved morphological characters. These datasets permit diversification time to be explored for Myriapoda.

MATERIAL AND METHODS

Sample collection and molecular techniques

Twenty-five species representing three of the four major groups of myriapods (Diplopoda, Chilopoda and Symphyla) were collected and newly sequenced for this study. Our sampling was designed to maximize representation at the ordinal level in millipedes and at the family level in centipedes. Information on sampling localities can be found in Table 1, and collection details can be found online in the Museum of Comparative Zoology collections database MCZbase (http://mczbase.mcz.harvard.edu). In addition, eight millipedes (Brewer and Bond 2013), seven centipedes (six transcriptomes from Fernández et al. (2014b) and a genome from Chipman et al. (2014)) were retrieved from the Sequence Read Archive (SRA). The following taxa were used as outgroups: an onychophoran (*Peripatopsis overbergiensis*), two crustaceans (*Calanus finmarchicus* and *Daphnia pulex*), an insect (*Drosophila melanogaster*), and six chelicerates (*Limulus polyphemus*, *Liphistius malayanus, Damon variegatus, Mastigoproctus giganteus, Centruroides vittatus* and *Metasiro americanus*). In addition, a seventh chelicerate outgroup, *Anoplodactylus insignis* (Pycnogonida), was newly sequenced for this study. All sequenced cDNA libraries are accessioned in the Sequence Read Archive (Table 1). Tissue preservation and RNA sequencing are as described in Fernández et al. (2014b). All data sets included in this study were sequenced with the Illumina platform. In addition, we retrieved all available sequence data from the pauropod *Euripauropus spinosus* (Regier et al. 2010) in an attempt to place this group in the myriapod tree. However, only 4 genes (out of the 57 Sanger-sequenced genes) were recovered as orthologs with a minimum of 50% of gene occupancy, and the tree showed low support values (see Fig. S1).

Data sanitation, sequence assembly and orthology assignment

De-multiplexed Illumina HiSeq 2500 sequencing results, in FASTQ format, were retrieved from the sequencing facility via FTP. Sequenced results were

6

quality filtered accordingly to a threshold average quality Phred score of 30 and adaptors trimmed using Trimgalore v 0.3.3 (Wu et al. 2011). Ribosomal RNA (rRNA) was filtered out. All known metazoan rRNA sequences were downloaded from GenBank and formatted into bowtie index using 'bowtie-build'. Each sample was sequentially aligned to the index allowing up to two mismatches via Bowtie 1.0.0 (Langmead et al. 2009). Strand-specific *de novo* assemblies were done individually for each specimen in Trinity (Haas et al. 2013) with a path reinforcement distance of 75. Redundancy reduction was done with CD-HIT (Fu et al. 2012) in the raw assemblies (95% similarity). Resulting assemblies were processed in TransDecoder (Haas et al. 2013) in order to identify candidate open reading frames (ORFs) within the transcripts. Predicted peptides were then processed with a further filter to select only one peptide per putative unigene, by choosing the longest ORF per Trinity subcomponent with a python script ('choose_longest_v3.py', http://github//rfernandezgarcia//phylogenomics), thus removing the variation in the coding regions of our assemblies due to alternative splicing, closely-related paralogs, and allelic diversity. Peptide sequences with all final candidate ORFs were retained as multi-fasta files.

We assigned predicted ORFs into orthologous groups across all samples using the Orthologous Matrix algorithm, OMA stand-alone v0.99z (Altenhoff et al. 2011; Altenhoff et al. 2013), which has been shown to outperform alternative approaches (such as reciprocal best hit) in identifying true orthologs and minimizing Type I error in orthology assignment (Altenhoff and Dessimoz 2009).

Phylogenomic analyses and congruence assessment

In order to explore the trade-off between number of genes and matrix completeness, three supermatrices were constructed by varying gene occupancy thresholds: supermatrix I (50% mean gene occupancy, 2,131 genes, 638,722 amino acids), supermatrix II (75% mean gene occupancy, 789 genes, 104,535 amino acids) and supermatrix III (90% gene occupancy, 123 genes, 27,217 amino acids). We also constructed two additional matrices (supermatrix IV, 87%

mean gene occupancy, 40 genes, 12,348 amino acids, and supermatrix V, 87.4% mean gene occupancy, 62 genes, 16,324 amino acids) from subsets of supermatrix III, by excluding genes with the lowest phylogenetic informativeness (Townsend 2007; López-Giráldez et al. 2013). For this purpose, we analyzed the signal-to-noise distribution of the 123 genes of supermatrix III using the method described in Townsend et al. (2012) as implemented in the PhyDesign web server. This method estimates the state space and the evolutionary rates of characters to approximate the probability of phylogenetic signal versus noise due to convergence or parallelism. At every site, based on the rate of character evolution and the character state space, phylogenetic signal is characterized by the probability of observing a parsimony informative synapomorphic site pattern at the leaves of the taxa, while phylogenetic noise is characterized by the probability distribution function over time for homoplastic site patterns that mimic the correct pattern and mislead analyses. In this context, to construct supermatrices IV and V we selected the orthogroups in which the signal was higher than the noise for the shortest internodes and the longest branches in our dataset (t_0 = 10, T = 455), given an ultrametric tree generated by node calibration as described below in this section. Supermatrix IV included the 40 genes with the highest values of phylogenetic informativeness (see Fig. S2). Likewise, to construct supermatrix V we included only the genes in the two upper quartiles in their ranking of phylogenetic informativeness (62 orthologs).

Finally, to account for the effect of evolutionary rate, we ordered the orthogroups in supermatrix I based on their increasing rate and selected the 100 most conserved genes (supermatrix VI; 60.3-99.3% of conserved sites, 70% mean gene occupancy; 29,039 amino acids), the 100 genes with a variation close to the mean (supermatrix VII; 21-23.8% of conserved sites; 61% mean gene occupancy, 30,585 amino acids) and the 100 most variable genes (supermatrix VIII, 0.9-6.3% of conserved sites; 43.6% mean gene occupancy, 25,666 amino acids). Similar criteria have been applied to construct matrices for exploratory purposes in other phylogenomic analyses (e.g., Fernández et al. 2014a; Andrade et al. 2015).

8

The selected orthogroups in supermatrices I, II and III were aligned individually using MUSCLE version 3.6 (Edgar 2004). We then applied a probabilistic character masking with ZORRO (Wu et al. 2012) to account for alignment uncertainty, using default parameters. ZORRO assigns confidence scores ranging from 1 to 10 to each alignment site using a pair hidden Markov model (pHMM) framework. We discarded the positions assigned a confidence score below a threshold of 5 with a custom python script ('cut_zorro.py'). These steps were not necessary in the remaining matrices as they were derived from supermatrices I and III. The aligned, masked orthogroups were then concatenated using Phyutility 2.6 (Smith and Dunn 2008).

Maximum likelihood inference was conducted with PhyML-PCMA (Zoller and Schneider 2013) with 20 nodes, and with ExaML (Aberer and Stamatakis 2013) using the per-site rate (PSR) category model. In order to further test for the effect of heterotachy and heterogeneous substitution rates, we also analyzed some of the matrices in PhyML v.3.0.3 implementing the integrated length (IL) approach (Guindon and Gascuel 2003; Guindon 2013). In this analysis, the starting tree was set to the optimal parsimony tree and the FreeRate model (Soubrier et al. 2012) was selected. Bayesian analyses were conducted with ExaBayes (Stamatakis 2014) and PhyloBayes MPI 1.4e (Lartillot et al. 2013), selecting in this last analysis the site-heterogeneous CAT-GTR model of amino acid substitution (Lartillot and Philippe 2004). Two independent MCMC chains were run for 5,000-10,000 cycles. The initial 25% of trees sampled in each MCMC run prior to convergence (judged when maximum bipartition discrepancies across chains were < 0.1) were discarded as the burn-in period. A 50% majority-rule consensus tree was then computed from the remaining trees. For practical reasons and due to the similar results obtained for the different phylogenetic analysis (see Results and Discussion), not all the analyses were implemented in all the supermatrices (see Fig. 2), but at least one ML and one Bayesian inference analysis per supermatrix were explored (Fig. 2).

To discern whether compositional heterogeneity among taxa and/or within each individual ortholog alignment was affecting the phylogenetic results,

we further analyzed matrices I and II in BaCoCa v.1.1 (Kück and Struck 2014). The relative composition frequency variability (RCFV) measures the absolute deviation from the mean for each amino acid for each taxon. RCFV values were plotted in a heatmap using the R package gplots with an R script modified from Kück and Struck (2014).

To investigate potential incongruence between individual gene trees, we inferred gene trees for each OMA group included supermatrices I, II and III (Fernández et al. 2014a; Fernández et al. 2014b; Laumer et al. 2015). Best-scoring ML trees were inferred for each gene under the selected model of sequence evolution from 100 replicates. Gene trees were decomposed into quartets with SuperQ v.1.1 (Grünewald et al. 2013), and a supernetwork assigning edge lengths based on quartet frequencies was inferred selecting the "balanced" edge-weight optimization function, applying no filter; the supernetworks were visualized in SplitsTree v.4.13.1 (Huson and Bryant 2006). All Python custom scripts can be downloaded from https://github.com/rfernandezgarcia/phylogenomics.

As our data set includes several ancient lineages (in the order of several hundreds of Ma) with a low diversity of extant species, we further evaluated the potential effect of long branch attraction (LBA) in three lineages that show less extant diversity than their extant closest relatives: Craterostigmomorpha, Scutigeromorpha and Polyxenida. Previous morphological and molecular analyses have recovered Scutigeromorpha as the sister group to all other centipede orders (e.g., Shear and Bonamo 1988; Borucki 1996; Giribet et al. 1999; Edgecombe and Giribet 2004; Murienne et al. 2010; Fernández et al. 2014b). The three families of the order include a total of ca. 95 species. More striking is the case of Craterostigmomorpha, with only two extant species (see Results and Discussion). These two orders originated more than 400 Mya, thus exhibiting very long branches prior to their respective diversification. In millipedes, the order Polyxenida includes only 89 species, and it is universally recognized as sister group to the other Diplopoda. As the basalmost groups of both centipedes and millipedes involve long branches, we also explored the

10

effect of LBA in a fourth lineage with long branches and poor taxon representation, Symphyla, which were recovered as sister group to Polyxenida in some of our analyses (see Results and Discussion). To determine whether or not long branch attraction was affecting tree topology, we applied the SAW method of Siddall and Whiting (1999). This method is based on the removal of taxa in a pair suspected to be affected by LBA, one at a time. If after re-running the analysis either of the taxa appears at different branch points in the absence of the other, LBA is postulated. To understand the potential effect of LBA in our data set between (a) Symphyla and Polyxenida, and (b) Scutigeromorpha and Craterostigmomorpha, we pruned one taxon at a time in each case from the most complete matrix, supermatrix III (123 genes), and re-ran ExaBayes and PhyML_PCMA, as described above.

As rogue taxa can frequently have a negative impact on topology or in a bootstrap analysis (see a review in Goloboff and Szumik 2015), we also explored the presence of putative wild card taxa in our data set with RogueNaRok (Aberer et al. 2013), with the aim of identifying a set of taxa that, if pruned from the underlying bootstrap trees, yielded a reduced consensus tree containing additional bipartitions or increased support values.

Morphological data

Morphological characters for the set of species for which transcriptomes were available or newly generated were coded, principally drawing on existing datasets. These include characters bearing on myriapod phylogeny (Rota-Stabelli et al. 2011), higher level chilopod phylogeny (Edgecombe and Giribet 2004; Murienne et al. 2010), diplopod phylogeny (Blanke and Wesener 2014), scutigeromorphs (Edgecombe and Giribet 2006; Koch and Edgecombe 2006), scolopendromorphs (Vahtera et al. 2013), and geophilomorphs (Koch and Edgecombe 2012; Bonato et al. 2014a). Sources for morphological information used for coding fossil terminals is detailed below ("Sources of coding and dates for fossils"). The dataset is available in nexus format as Morphobank Project

11

P2216, "The myriapod tree of life"

(www.morphobank.org/index.php/MyProjects/List/select/project_id/2216).

The morphology dataset consists of 232 characters, of which characters 57, 72, 86, 101 and 109 were scored as ordered/additive. These data were analyzed using parsimony as an optimality criterion in TNT (Goloboff et al. 2008) with equal character weights. A heuristic search used 10,000 random stepwise addition sequences saving 100 trees per replicate, with TBR branch swapping. Since our goal was not to produce a phylogenetic hypothesis derived exclusively from morphological data, no further analyses, such as those using implied weighting, were conducted.

Divergence time inference

In order to infer divergence time in the myriapod phylogeny, we included seven Palaeozoic and Mesozoic fossils in our data set: four centipedes (*Crussolum* sp*., Devonobius delta, Mazoscolopendra richardsoni* and *Kachinophilus pereirai*) and three millipedes (*Cowiedesmus eroticopodus, Archidesmus macnicoli* and *Gaspestria genselorum*). In addition, we included two outgroup fossils, a crustacean (*Rehbachiella kinnekullensis*) and a scorpion (*Proscorpius osborni*). Absolute dates follow the International Stratigraphic Chart v 2014/02.

a) Sources of codings and dates for fossils

Crussolum Shear in Shear et al., 1998 : The scutigeromorph genus *Crussolum* includes one formally named species, *C. crusseratum* Shear in Shear et al. (1998), known from isolated and fragmentary legs from the Middle Devonian Gilboa locality, Schoharie County, New York State. Fossils come from the upper part of the Panther Mountain Formation, dated to the Tioughniogan regional Stage, Givetian in the global time scale. Palynomorphs are consistent with a Givetian age (Richardson et al. 1993). Legs and the forcipular segment of *Crussolum* sp. have been documented from the Windyfield Chert of the Dryden Flags Formation, Aberdeenshire, Scotland, by Anderson and Trewin (2003). An

12

antenna was ascribed to this species by Anderson and Trewin (2003) but has not been used for coding because a hexapod identity cannot be ruled out. Spore assemblages of the Windyfield and stratigraphically underlying Rhynie Chert are dated to the early but not earliest Pragian to early (earliest?) Emsian (*polygonalis-emsiensis* Spore Assemblage Biozone) (Parry et al. 2011). Radiometric dating of the underlying Milton of Noth Andesite at ca 411 Ma (Parry et al. 2011, 2013) has been subject to a dispute over its temporal relationship to hot spring activity associated with the cherts (Mark et al. 2011; Mark et al. 2013) and predates the biostratigraphic dating of the Rhynie Chert relative to the global dating of the base of the Pragian Stage. We apply a date of the 407.6 Ma to the Rhynie and Windyfield cherts, using the Pragian-Emsian boundary as a reference. The coding for *Crussolum* draws mostly on the Windyfield Chert material, and accordingly is dated at 407.6 Ma. The oldest samples of *Crussolum* consist of isolated legs from Ludford Lane in England (Shear et al., 1998), sourced from a horizon 0.15-0.20 m above the base of the Ludlow Bone Bed Member of the Downtown Castle Sandstone Formation. The Ludlow Bone Bed Member is earliest Přidolí in age (Jeram et al., 1990).

Devonobius delta Shear and Bonamo, 1988: Described as the type species of a monotypic centipede order Devonobiomorpha, *Devonobius delta* is coded based on figured material of Shear and Bonamo (1988), all from Gilboa, New York. Stratigraphic details are identical to *Crussolum crusseratum* (see above). A minimum date for the end of the Givetian/base of the Frasnian is applied (382.7 Ma).

Mazoscolopendra richardsoni Mundel, 1979: Codings for this scolopendromorph centipede are based on descriptions and figures by Mundel (1979) and Haug et al. (2014), and personal observation by GDE of material in the Field Museum. Specimens are derived from the Francis Creek Shale Member of the Carbondale Formation, Mazon Creek, Illinois, of Westphalian D age (Shabica and Hay 1997). In the global time scale, this falls within the Moscovian Stage. A date for the top of the Moscovian/base of the Kasimovian is applied (307.0 Ma).

Kachinophilus pereirai Bonato et al., 2014 : Coding of this geophilomorph centipede is based on the original description and direct study of the type material (Bonato et al. 2014b). U-Pd dating of zircons in rock matrix in "burmite" (Burmese amber) establishes a maximum age of 98.79 +/- 0.62Ma (early Cenomanian) for amber inclusions (Shi et al. 2012).

Cowiedesmus eroticopodus Wilson and Anderson, 2004: This species is one of three co-occurring species assigned to Archipolypoda, an extinct superordinal level grouping identified as Chilognatha (Wilson and Anderson, 2004). These represent the oldest body fossils of Diplopoda. Coding is based on the original description and figures of the holotype by Wilson and Anderson (2004) and personal examination by GDE of the specimen in the Australian Museum. This species occurs in the *Dictyocaris* Member of the Cowie Formation in the Stonehaven Group, Stonehaven, Scotland. As summarized by Wilson and Anderson (2004), associated palynomorphs date the unit to the late Wenlock to early Ludlow (Silurian). A minimum date for the early Ludlow uses the base of the Ludfordian Stage (425.6 Ma).

Archidesmus macnicoli Peach, 1882: Selected to supplement *Cowiedesmus* in providing morphological characters for Archipolypoda, coding is based on the revision of this species by Wilson and Anderson (2004). Specimens figured therein come from the Dundee Formation in the Arbuthnott Group, Tillywhandlung Quarry, near Forfar, Scotland. Miospores date the unit to the *micrornatus-newportensis* biozone, of Lochkovian (Lower Devonian) age (Wilson and Anderson 2004). A minimum date is applied to Lochkovian-Pragian boundary (410.8 Ma).

Gaspestria genselorum Wilson, 2006: Coding of this diplopod is based on the original description and published illustrations by Wilson (2006), who interpreted *G. genselorum* as the oldest juliformian, and new photographs of the types provided by R. Miller. Wilson (2006) reviewed the plant megafossil and spore constraints on the age of the two units in which the species occurs, the Cap-aux-Ocs Member of the Battery Point Formation (Gaspé, Québec) and the Campbelltown Formation (New Brunswick). Spores at the Gaspé site are

14

assigned to the *sextantii* Subzone of the *annulatus-lindlarensis* Zone and those of the New Brunswick site to the *Grundiospora* Subzone of the *annulatus-lindlarensis* Zone. Both are of late Emsian age. A minimum date of 393.3 Ma is accordingly applied, constrained by the Emsian-Eifelian boundary.

Proscorpius osborni (Whitfield, 1885): This species was selected as being an anatomically well-understood Silurian scorpion (Whitfield 1885), using the revision by Dunlop et al. (2008) for coding. Formerly known as the 'Bertie Waterlime', the stratigraphic unit from which the material was sourced is the Phelps Member of the Fiddlers Green Formation, in the Bertie Group, New York State. This unit is dated to the Přídolí Series of the Silurian (Dunlop et al. 2008), and assigned a minimal age of 419.2 Ma (dating the Přídolí-Lochkovian boundary).

Rehbachiella kinnekullensis Müller, 1983: The monographic revision by Walossek (1993) was used for coding. Material documented therein was derived from four localities of the late Cambrian *Orsten* in southern Sweden. These are assigned to the *Agnostus pisiformis* Zone and, less commonly, to the overlying *Olenus gibbosus* Zone (late Guzhangian and earliest Paibian, respectively: Nielsen et al. (2014)). Because some of the material is demonstrably Guzhangian, a minimal date is applied to the bases of the Furongian and Paibian, 497 Ma.

b) Divergence time inference through node calibration

Temporal data for the fossils listed above provide constraints on divergence dates. The resolution of *Crussolum* as stem-group Scutigeromorpha in our morphological cladogram (Fig. 3; Fig. S3) constrains the split of Scutigeromorpha and Pleurostigmophora, i.e., crown-group Chilopoda, to at least 419.2 Ma, drawing on the Ludlow Bone Bed material of *Crussolum*. *Devonobius delta* is resolved in our morphological analysis as sister group of Epimorpha but an alternative affinity to *Craterostigmus* has been suggested (Borucki 1996). To allow for uncertainty over the phylogenetic position of *Devonobius* and *Craterostigmus*, the fossil was conservatively used only to constrain crown-group Pleurostigmophora (minimum of 382.7 Ma based on *Devonobius*). Crown-group

15

Epimorpha is constrained to a minimum of 306 Ma by the oldest scolopendromorph *Mazoscolopendra richardsoni*. That species is resolved as total-group Scolopendromorpha based on our morphological analysis (Fig. S3). *Kachinophilus pereirai*, originally assigned to Geophilidae (Bonato et al. 2014b), is resolved based on morphology in a clade with extant geophilids (Fig. S3). As such, it constrains crown-group Adesmata, i.e., the split between Mecistocephalidae and remaining Geophilomorpha, to at least 98.79 Ma. Crown-group Chilognatha is constrained by the occurrence of *Cowiedesmus eroticopodus*, resolved in our morphological analysis as total-group Helminthomorpha (425.6 Ma). *Rehbachiella kinnekullensis* has been interpreted as stem-group Anostraca, i.e., crown-group Branchiopoda (Walozsek, 1993, reference phylogeny Fig. 41 therein), or as stem-group Branchiopoda (Olesen 2009; reference phylogeny in Fig. 14 therein). Our morphological data agree with branchiopod affinities in resolving it as more closely related to *Daphnia* than to *Calanus*. This provides a minimum date for crown-group Altocrustacea (sensu Regier et al. 2010) to 497 Ma. Because of uncertainty with regards to the interrelationships between major pancrustacean groups (including Branchiopoda, Copepoda and Insecta, from which the three exemplars used here were sampled) this date is conservatively applied only to crown-group Pancrustacea. *Proscorpius osborni* is resolved by our morphological data dataset as closest relative of the extant scorpion exemplar *Centruroides*, setting a constraint on the split of Scorpiones from Tetrapulmonata. The relevant crown-group based on our taxon sampling is Arachnopulmonata, dated to at least 419.2 Ma.

The split between Onychophora and Arthropoda was dated between 528 million years (the minimum age for Arthropoda used by Lee et al. 2013 on the basis of the earliest *Rusophycus* traces in the early Fortunian, dating the base of Cambrian Stage 2) and 558 million years. The latter was used as the root of Panarthropoda (Lee et al. 2013) based on dating of White Sea Vendian strata that contain the oldest plausible bilaterian fossils.

Divergence dates were estimated using the Bayesian relaxed molecular clock approach as implemented in PhyloBayes v.3.3f (Lartillot et al. 2013). An auto-correlated model of clock relaxation was applied as it has been shown to provide a significantly better fit than uncorrelated models on phylogenomic datasets (Lepage et al. 2007; Rehm et al. 2011). The calibration constraints specified above were used with soft bounds (Yang and Rannala 2006) under a birth-death prior in PhyloBayes, this strategy having been found to provide the best compromise for dating estimates (Inoue et al. 2010). Two independent MCMC chains were run for 5,000-7,000 cycles, sampling posterior rates and dates every 10 cycles. The initial 25% were discarded as burn-in. Posterior estimates of divergence dates were then computed from the remaining samples of each chain.

RESULTS AND DISCUSSION

Phylogenetic relationships of Myriapoda

All analyses conducted, regardless of methodology or data set, recover monophyly of Myriapoda, Diplopoda, Chilopoda and Symphyla, but the interrelationships among the myriapod classes are less stable (Fig. 2). With regards to the position of Symphyla, two alternatives are found: Symphyla either unite with Diplopoda as predicted by the Progoneata hypothesis or Chilopoda and Diplopoda unite as a clade to the exclusion of Symphyla. The latter result was also found in many of the analyses by Rehm et al. (2014). A chilopod-diplopod clade has not been anticipated morphologically, though certain morphological characters fit such a grouping. For example, chilopods and diplopods share a series of imbricated comb lamellae on the mandibles that are lacking in symphylans and pauropods (Edgecombe and Giribet 2002). This character was proposed as a potential myriapod autapomorphy but the standard phylogenetic hypotheses forced reversals/losses in pauropods and symphylans. Under one of our two scenarios, although in the absence of pauropods, this

17

character would be interpreted as a synapomorphy of Chilopoda + Diplopoda. In the analyses designed to account for LBA, such as the SAW analysis (Fig. S4) with Polyxenida removed and the PhyloBayes analysis with the CAT model of heterogeneity, Symphyla was recovered as sister group to the clade Diplopoda + Chilopoda, suggesting that Symphyla could have been attracted to Diplopoda due to the long branch of Polyxenida. It need be noted, however, that a position of Symphyla closer to Diplopoda than to Chilopoda has been endorsed morphologically (Progoneata hypothesis) (Dohle 1980; Edgecombe 2004, 2011).

The 4-gene analysis including the sequences available for a pauropod recovered Pauropoda as sister group to Symphyla with posterior probability = 0.99. However, the relationships with the other myriapod groups remained unsupported, and the pauropod-symphylan group (=Edafopoda) is observed to be associated with anomalous relationships between orders of both Chilopoda and Diplopoda (Fig. S1). A denser taxon sampling of the two lesser myriapod groups, Symphyla (e.g., including representatives of the other family, Scolopendrellidae) and Pauropoda, would be definitely key to resolve the backbone of the myriapod tree with greater confidence.

Diplopoda: Within Diplopoda, all analyses (apart from the one including a pauropod; Fig. S4) support a sister group relationship between Penicillata (sampled by *Eudigraphis*) and all other millipedes, the latter forming the traditional clade Chilognatha. Chilognatha is united by a calcified cuticle and the male actively transferring sperm to the female genital opening, among other apomorphic characters (Enghoff 1984; Blanke and Wesener 2014). The fundamental division within Chilognatha likewise corresponds to traditional morphology-based systematics, including a sister group relationship between Pentazonia (generally pill millipedes) and Helminthomorpha (long-bodied millipedes).

Relationships within Pentazonia depart from the standard morphological hypothesis. The latter usually unites Glomerida and Sphaerotheriida as sister groups (collectively forming a clade of pill millipedes named Oniscomorpha) to the exclusion of Glomeridesmida. In contrast to this, however, our analyses

18

resolve Sphaerotheriida (*Cyliosoma*) as sister group to Glomerida (*Glomeris*) + Glomeridesmida (*Glomeridesmus*). The only previous molecular studies to include all three orders of Pentazonia are the three-gene analysis of Regier et al. (2005) and the same dataset combined with morphological data by Sierwald and Bond (2007). As in our analyses, these yielded an "unexpected" (Regier et al. 2005: p. 155) grouping of Glomeridesmida and Glomerida to the exclusion of Sphaerotheriida. Monophyly of Oniscomorpha has been defended morphologically based on the presence of a collum much smaller than the following tergites, the second tergite being much larger than following tergites, and a the lack of coxal pouches on legs (Blanke and Wesener 2014). The status of these characters as synapomorphies is disputed by our trees, and instead are interpreted as plesiomorphies of Pentazonia.

In Helminthomorpha, the traditional clades Colobognatha and Eugnatha are each monophyletic, as found in prior analyses of largely the same set of millipede species (Brewer and Bond, 2013), as well as by sampling different exemplars for three nuclear coding genes (Miyazawa et al. 2014). The colobognathan terminals are *Brachycybe* (order Platydesmida) and *Petaserpes* (order Polyzoniida); transcriptomic data remain unavailable for two other colobognathan orders, Siphonocryptida and Siphonophorida. Colobognatha and Eugnatha were originally proposed in the classical era of myriapod systematics (Attems 1926; Verhoeff 1928), but were discarded in the influential millipede classification by Hoffman (1980), who considered that the characteristic reduced mouthparts of Colobognatha were prone to convergence. They have, however, consistently been endorsed in cladistic analyses (Enghoff 1984; Sierwald et al. 2003; Blanke and Wesener 2014). As likewise found by Brewer and Bond (2013), Eugnatha includes two groups that have a long history of usage in morphology-based systematics. One of these unites Chordeumatida (sampled by *Cleidogona*) and Callipodida (sampled by *Abacion*), which are classified together as Coelochaeta. Another clade consists of the ring-forming millipedes, Juliformia, which unites the sampled members of Julida (*Cylindroiulus*), Spirostreptida (*Cambala*) and Spirobolida (*Narceus*). Our analyses variably resolve the three-

taxon problem of the juliform orders with either Julida or Spirobolida as sister group of Spirostreptida. Both of these conflict with the weak morphological support for Spirobolida and Julida as closest relatives (Blanke and Wesener 2014), but have been found in other molecular studies. A Julida + Spirostreptida clade was supported by three nuclear protein-coding genes (Miyazawa et al. 2014), whereas nuclear ribosomal genes (Cong et al. 2009) and mitochondrial gene rearrangements and amino acid sequences (Woo et al. 2007) supported a Spirobolida + Spirostreptida clade. The relationships of the juliform orders need be regarded as an open question for the time being.

Two of the most contentious issues in millipede systematics involve the relationships of Polydesmida and of Stemmiulida. From the perspective of morphology, Polydesmida has been considered the sister group of Juliformia as a "ring-forming" clade (Enghoff et al. 1993) or as sister group to a putative clade named Nematophora (Sierwald et al. 2003; Blanke and Wesener 2014). The latter group, named for the shared presence of preanal spinnerets, traditionally includes Stemmiulida, Callipodida and Chordeumatida (Enghoff 1984; Sierwald et al. 2003). Spinnerets are, however, more widely shared by Polydesmida (Shear 2008) and this character has been cited as synapomorphic for a clade containing Polydesmida + Nematophora (Blanke and Wesener 2014). Our results, like those of Brewer and Bond (2013), do not retrieve monophyly of Nematophora in its traditional guise: Chordeumatida + Callipodida do not unite with Stemmiulida, the latter being resolved either with Polydesmida or with Juliformia. Optimization on the topology of Brewer and Bond (2013) as well as those obtained here suggests that spinnerets are an autapomorphy of Eugnatha as a whole and were lost in Juliformia.

Chilopoda: Half of the analyses retrieved the widely accepted division of Chilopoda into Notostigmophora (Scutigeromorpha) and Pleurostigmophora (the other four orders). However, in the other half Scutigeromorpha and Craterostigmomorpha were recovered as a clade. This result has never been advocated morphologically. Both clades are invariantly at the base of the tree, as recovered in all the SAW analyses (Fig. S6). However, their interrelationships

20

remain unresolved. A possible cause of this phylogenetic conundrum could be the low diversity of Craterostigmomorpha, which comprises only two extant species. Within Pleurostigmophora, the Amalpighiata hypothesis (Craterostigmus sister group of all other pleurostigmophorans; Fernández et al. 2014b) withstood all procedures designed to target systematic error, such as increased gene sampling and the exploration of complex evolutionary models, and was favored over the Phylactometria hypothesis (Lithobiomorpha sister group of all other pleurostigmophorans; Edgecombe and Giribet 2004). The relationships between Lithobiomorpha and Epimorpha (= Scolopendromorpha and Geophilomorpha) remained elusive, as several analyses recovered Scolopendromorpha as sister group to the rest of this clade, defying nearly all previous morphological and molecular evidence. Notably, all the matrices recovering this topology include the highest gene occupancy (supermatrices III-V) at the expense of numbers of genes (see below for further details).

All chilopod orders are well supported and relationships within each of the large orders correspond to previous phylogenetic hypotheses, and for most to traditional taxonomy. The three families of Scutigeromorpha resolve with Pselliodidae (*Sphendononema*) as sister group to Scutigerinidae (*Scutigerina*) and Scutigeridae (*Scutigera*). The same topology has consistently been inferred using targeted sequencing of a few loci (Edgecombe and Giribet 2006, 2009; Butler et al. 2010; Giribet and Edgecombe 2013), but contrasts with the morphological placement of Scutigerinidae as sister group to Pselliodidae and Scutigeridae (Koch and Edgecombe 2006). Lithobiomorpha comprises two clades that correspond to its two families, Henicopidae (*Anopsobius* and *Paralamyctes*) and Lithobiidae (*Lithobius* and *Eupolybothrus*). Scolopendromorpha consists of a blind clade and an ocellate clade. The blind clade, supported in all analyses, unites members of Cryptopidae, Scolopocryptopidae and Plutoniumidae whereas the ocellate clade is the family Scolopendridae. The number of events of eye loss in Scolopendromorpha had been a subject of some debate (Vahtera et al. 2012) but the transcriptomic data strongly corroborate weaker signal from targeted gene sequencing (Vahtera et

21

al. 2013) and morphology (Koch et al. 2009; Koch et al. 2010) for blindness in the common ancestor of Cryptopidae and Scolopocryptopidae + Plutoniumidae. All analyses herein resolve Scolopocryptopidae as paraphyletic, *Scolopocryptops* being more closely related to *Theatops* (Plutoniumidae) than to *Newportia*. This result implies that the shared presence of 23 leg-bearing trunk segments in Scolopocryptopinae (*Scolopocryptops*) and Newportiinae (*Newportia*) is either acquired convergently from a 21-segmented ancestor or is homologous but was reversed (to 21 segments) in Plutoniumidae. Monophyletic Scolopendridae consists of Otostigminae (*Alipes* and *Rhysida*) and Scolopendrinae (*Akymnopellis* and *Scolopendropsis*), precisely mirroring classical and current taxonomy. Geophilomorpha also resolves along traditional lines into Placodesmata (Mecistocephalidae: *Mecistocephalus* and *Tygarrup*) and Adesmata, a clade composed of all other geophilomorphs. Resolution within Adesmata conflicts with the most recent analysis of geophilomorph relationships (Bonato et al. 2014); they found a clade containing Oryidae, Himantariidae and Schendylidae (Himantarioidea), while our analyses likewise recover a Himantariidae + Schendylidae clade but place *Notiphilides* (Oryidae) as sister group to Geophilidae, a result obtained for every analysis of each data set. However, we were not able to fully test the monophyly of Geophiloidea, as Zelanophilidae and Gonibregmatidae are not represented in our analyses. Our three Geophilidae (representatives of Geophilidae and two recently synonymized families Dignathodontidae and Linotaeniidae) constitute a well-supported and stable clade.

Tempo of diversification in myriapods

The dated phylogeny for the 123 gene matrix reflects a diversification of Myriapoda in the early and middle Cambrian, supporting a Cambrian diversification event, as suggested recently for this lineage (Rota-Stabelli et al. 2013). This dating continues to substantially predate body fossil or even trace fossil evidence for Myriapoda, the oldest trackways that can be ascribed to

22

myriapods with reasonable confidence being Late Ordovician in age (Johnson et al. 1994; Wilson 2006). Diversification of Diplopoda is inferred to occur in the late Cambrian to mid Silurian, and the diversification of Chilopoda in the Early Ordovician to Middle Devonian (Fig. 3). The diversification times of the millipede and centipede orders included in our analyses are thus generally congruent with the timing inferred in previous studies (Meusemann et al. 2010; Rehm et al. 2011; Brewer and Bond 2013; Fernández et al. 2014b), and do not greatly predate first occurrences of major chilopod and diplopod lineages in the fossil record. On the other hand, an estimated Cambrian diversification of myriapods implies a large gap in the fossil record. Fossil candidates for stem-group Myriapoda in marine, freshwater or terrestrial sediments of Cambrian age remain unknown or unidentified (Edgecombe 2004; Shear and Edgecombe 2010). Nonetheless, the occurrence of crown-group Pancrustacea as early as Cambrian Stage 3 (Edgecombe and Legg 2014) predicts a ghost lineage for Myriapoda to at least that time, ca 517 Ma.

In centipedes, the additional transcriptomes with respect to previous studies (Fernández et al. 2014b) resulted in slightly older dates, but it did not affect their origin. For instance, the latter study inferred the diversification of Pleurostigmophora in the Middle Ordovician to Early Carboniferous, the diversification of Amalpighiata in the Silurian-Carboniferous, and the diversification of Epimorpha from the Early Devonian to the early Permian. In this study, the recovered dates ranged from the Early Ordovician to the Middle Devonian in Pleurostigmophora, from the Middle Ordovician to the Early Carboniferous in Amalpighiata, and Epimorpha from the Late Devonian to the mid Carboniferous.

The controversy over the systematic position of *Craterostigmus* (i.e., whether Phylactometria or Amalpighiata is monophyletic) relates to lineages that have ancient, rapidly diverging stem groups but relatively shallow crown groups, conditions expected to contribute to a difficult phylogenetic problem. Despite a divergence from other living chilopod orders by the Late Silurian, the two extant species of *Craterostigmus* have a mean date for their divergence from

each other in the Cretaceous, and these species are almost indistinguishable morphologically (Edgecombe and Giribet 2008). Likewise, Scutigeromorpha, with which *Craterostigmus* groups in various analyses, has a stem dating to the Silurian but the deepest split between its three extant families has a mean date in the Triassic. Although the group is also conservative morphologically, we have maximized phylogenetic diversity here for a clade that has been well resolved phylogenetically and which increased its diversification rate around 100 million years ago (Giribet and Edgecombe 2013). This is effectively like Craterostigmomorpha, where all the extant diversity is comparatively recent and old branches are unrepresented in our analyses because of extinction.

Using morphology to test phylogenomic hypothesis

Results herein demonstrate that most systematic hypotheses for Myriapoda are strongly supported across the explored sets of genes, optimality criteria and programs. For example, the deep interrelationships within the four large centipede orders are stable across all analyses (Fig. 2). These relationships also show a high degree of congruence with morphological trees and morphology-based classifications. In the case of, for example, Scolopendromorpha, we can conclude that a high degree of confidence can be placed in the results.

Where hypotheses vary according to different methods or data partitions, in some cases morphology provides a strong arbiter for evaluating the rival hypotheses. For example, the 123-gene analyses typically recover Lithobiomorpha as sister group of Geophilomorpha, whereas the larger gene samples generally group Geophilomorpha with Scolopendromorpha. The former hypothesis has, to our knowledge, only been recovered in a single, non-numerical phylogenetic analysis, based on a single character of sperm structure (Jamieson 1986). In contrast, a geophilomorph-scolopendromorph clade – the classical Epimorpha – receives morphological support from the perspectives of development, behavior, external morphology and internal anatomy (eight

24

autapomorphies listed by Edgecombe 2011). This ability of morphology to select between rival hypotheses that are each based on vast pools of data affirms the continued relevance of morphological characters in phylogenetics (e.g., Giribet 2015; Lee and Palci 2015; Wanninger 2015) and permit delving into the possible reasons for the discordance between the different analyses.

Furthermore, access to a morphological dataset allowed us to establish the position of fossil terminals for node calibration using the exact same set of extant terminals used in phylogenomic analyses (as in Sharma and Giribet 2014). Most recent justification for the necessity of morphological data in molecular dating has focused on its indispensability for so-called tip dating or total evidence dating (e.g., Giribet 2015; Lee and Palci 2015; Pyron 2015). However, even in the standard node calibration approach as employed here, the position of fossils can be determined in the context of the precise taxonomic sample used in other parts of the study, rather than cobbling together justifications for node calibrations from external (or possibly mixed) sources.

The effect of confounding factors in phylogenomic reconstruction: matrices with high taxon occupancy recovered the 'wrong' topologies

Assessing the effect of missing data on phylogenetic inference has received substantial attention from phylogeneticists over many years, but unambiguous conclusions have been evasive. Properties of the data sets themselves, such as different rates of evolution and compositional heterogeneity, can have a strong influence on the accuracy of phylogenetic inferences, irrespective of the amount or the pattern of missing data. The combined effect of these factors complicates the problem, and no study has dissected out in depth the effect of each parameter independently. The different matrices constructed and the multitude of analyses conducted to address the potential effects of a diversity of confounding factors in phylogenomic inference (missing data, compositional heterogeneity, heterotachy, differential evolutionary rate, levels of phylogenetic informativeness, model misspecification, concatenation and gene-

25

tree incongruence) yield congruent results (see Fig. 2, Fig. S2, Fig. S3, Fig. S5, Fig. S6). The main three conflicting nodes (marked as 1, 2 and 3 in Fig. 2) are recovered in approximately half of the analyses, indicating that these artifactual/methodological explanations for the incongruence alone are unlikely to have driven the recovered topologies.

The potential impact of missing data on phylogenomic inference accuracy has been poorly investigated, particularly for higher-level (=deep and old) relationships, but gene completeness has often been advocated as the only justification for selecting genes. This has sometimes been used as a justification for using target-enrichment approaches to phylogeny versus 'phylotranscriptomics' (e.g., Lemmon and Lemmon 2013). In the past decades, several studies have investigated the effects of missing data in phylogenetic reconstruction (e.g., Maddison 1993; Norell et al. 1995; Norell and Wheeler 2003; Wiens 2003; Simmons 2012), with some recent studies focusing on phylogenomic data sets with hundreds or thousands of orthologous genes (e.g., Rokas et al. 2003; Philippe et al. 2004; Roure et al. 2013). These studies have suggested a negative (e.g., missing data have a deleterious effect on accuracy) or neutral (e.g., missing data *per se* do not directly affect inference) effect of the missing data. The negative effect is mainly due to a reduced detection of multiple substitutions, exacerbating systematic errors such as long-branch attraction (LBA), as incomplete species are less efficient in breaking long branches (Roure et al. 2013)--but that study demonstrated this using a matrix too small to resolve deep metazoan relationships. Perhaps for this reason other studies have shown that adding incomplete taxa is not deleterious *per se* as long as enough informative character states are present for each species, but analyzing too few complete characters could reduce accuracy because of a lack of phylogenetic signal (Wiens 2003; Philippe et al. 2004; Wiens 2006; Lemmon et al. 2009). In this study, we show for the first time that high matrix occupancy can lead to 'incorrect' inferences: all the different analyses conducted with the most complete matrix (supermatrix III) resulted in a topology of centipedes at odds with well-established phylogenetic hypotheses and with nearly all other analyses

26

using more genes or selecting genes based on other criteria (informativeness, compositional homogeneity, etc.), and curiously it was the matrix that showed the highest level of conflict between individual gene trees (Fig. S6). These three contradicting ideas on the effect of missing data in phylogenetic inference hinder the experimental design in phylogenomics, and highlight the necessity of further exploring its effect in new data sets.

The conflicting nodes between centipede orders are unlikely to result from incomplete taxon sampling, as our data set includes all extant families of centipedes with the exception of two geophilomorph families (Zelanophilidae and Gonibregmatidae) and one monotypic family of Scolopendromorpha (Mimopidae). In this context, biological explanations for the incongruence, such as incomplete lineage sorting consistent with a scenario of rapid radiation, or lineage-specific high extinction, may need to be considered. In fact, Craterostigmomorpha, the taxon responsible for some of the major instability, comprises only two species in Tasmania and New Zealand, constituting an old lineage with depauperate extant diversity and conservative morphology, some of the conditions often used to refer to living fossils (Werth and Shear 2014).

CONCLUDING REMARKS

In this study we explored the phylogeny of Myriapoda and the internal relationships of its largest clades, Diplopoda and Chilopoda, using published genomes and transcriptomes as well as novel transcriptomic data for 25 species. For this we constructed an array of data sets to independently optimize gene number, gene occupancy, phylogenetic informativeness, or gene conservation and analyzed them using different phylogenetic methods and evolutionary models. In addition we generated a morphological data matrix of 232 characters that was used to precisely place a set of fossils subsequently used for node calibration analyses. Our results largely corroborate those of previous analyses, especially with respect to the millipede ordinal relationships and added resolution to the centipede tree, sampled at the family-level. However, we

identified a few conflicting nodes across analyses to discover that for the most part the matrices optimizing gene occupancy produced topologies at odds with morphology, development, prior molecular data sets, or most notably, our own analyses using matrices with larger numbers of genes (with lower average gene occupancy) or matrices optimizing phylogenetic informativeness and gene conservation. This calls for caution when selecting data sets based strictly on matrix completeness and adds further support to previous notions that a large diversity of genes, even to the detriment of matrix occupancy (contra Roure et al. 2013), may be a feasible solution to analyzing phylogenetic relationships among deep animal clades (e.g., Hejnol et al. 2009). Finally, our dating analysis continues to support results from prior studies in placing the diversification of Myriapoda in the Cambrian, prior to any recognizable myriapod fossil record, whereas the estimated diversification of both millipedes and centipedes is closer to the existing fossil record. Further refinement of the myriapod tree of life will require addition of pauropods, more symphylans, and a few missing lineages of centipedes and millipedes.

REFERENCES

Aberer, A.J., Krompass, D., Stamatakis, A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Systematic Biology 62:162-166.

Aberer, A.J., Stamatakis, A. 2013. ExaML: Exascale Maximum Likelihood. Program and documentation available at: http://sco.h-its.org/exelixis/web/software/examl/index.html.

Altenhoff, A.M., Dessimoz, C. 2009. Phylogenetic and functional assessment of orthologs inference projects and methods. PLoS Computational Biology 5.

Altenhoff, A.M., Gil, M., Gonnet, G.H., Dessimoz, C. 2013. Inferring hierarchical orthologous groups from orthologous gene pairs. PLoS ONE 8:e53786.

Altenhoff, A.M., Schneider, A., Gonnet, G.H., Dessimoz, C. 2011. OMA 2011: orthology inference among 1000 complete genomes. Nucleic Acids Research 39:D289-D294.

Anderson, L.I., Trewin, N.H. 2003. An Early Devonian arthropod fauna from the Windyfield Cherts, Aberdeenshire, Scotland. Palaeontology 46:457-509.

Andrade, S.C.S., Novo, M., Kawauchi, G.Y., Worsaae, K., Pleijel, F., Giribet, G., Rouse, G.W. 2015. Articulating "archiannelids": Phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. Molecular Biology and Evolution 32:2860-2875.

Andrew, D.R. 2011. A new view of insect–crustacean relationships II. Inferences from expressed sequence tags and comparisons with neural cladistics. Arthropod Structure & Development 40:289-302.

Attems, C.G. 1926. Myriopoda. Pages 402 *in* Handbuch der Zoologie, 4(1) (W. Kükenthal, and T. Krumbach, eds.). de Gruyter, Berlin.

Blanke, A., Wesener, T. 2014. Revival of forgotten characters and modern imaging techniques help to produce a robust phylogeny of the Diplopoda (Arthropoda, Myriapoda). Arthropod Structure & Development 43:63-75.

Bonato, L., Drago, L., Murienne, J. 2014a. Phylogeny of Geophilomorpha (Chilopoda) inferred from new morphological and molecular evidence. Cladistics 30:485-507.

Bonato, L., Edgecombe, G.D., Minelli, A. 2014b. Geophilomorph centipedes from the Cretaceous amber of Burma. Palaeontology 57:97-110.

Borucki, H. 1996. Evolution und Phylogenetisches System der Chilopoda (Mandibulata, Tracheata). Verhandlungen des naturwissenschaftlichen Vereins in Hamburg 35:95-226.

Brewer, M.S., Bond, J.E. 2013. Ordinal-level phylogenomics of the arthropod class Diplopoda (millipedes) based on an analysis of 221 nuclear protein-coding loci generated using next-generation sequence analyses. PLoS ONE 8:e79935.

Brewer, M.S., Sierwald, P., Bond, J.E. 2012. Millipede taxonomy after 250 years: classification and taxonomic practices in a mega-diverse yet understudied arthropod group. PLoS ONE 7:e37240.

Butler, A.D., Edgecombe, G.D., Ball, A.D., Giribet, G. 2010. Resolving the phylogenetic position of enigmatic New Guinea and Seychelles Scutigeromorpha (Chilopoda): a molecular and morphological assessment of Ballonemini. Invertebrate Systematics 24:539-559.

Chipman, A.D., Ferrier, D.E., Brena, C., Qu, J., Hughes, D.S., Schroder, R., Torres-Oliva, M., Znassi, N., Jiang, H., Almeida, F.C., Alonso, C.R., Apostolou, Z., Aqrawi, P., Arthur, W., Barna, J.C., Blankenburg, K.P., Brites, D., Capella-Gutierrez, S., Coyle, M., Dearden, P.K., Du Pasquier, L., Duncan, E.J., Ebert, D., Eibner, C., Erikson, G., Evans, P.D., Extavour, C.G., Francisco, L., Gabaldon, T., Gillis, W.J., Goodwin-Horn, E.A., Green, J.E., Griffiths-Jones, S., Grimmelikhuijzen, C.J., Gubbala, S., Guigo, R., Han, Y., Hauser, F., Havlak, P., Hayden, L., Helbing, S., Holder, M., Hui, J.H., Hunn, J.P., Hunnekuhl, V.S., Jackson, L., Javaid, M., Jhangiani, S.N., Jiggins, F.M., Jones, T.E., Kaiser, T.S., Kalra, D., Kenny, N.J., Korchina, V., Kovar, C.L., Kraus, F.B., Lapraz, F., Lee, S.L., Lv, J., Mandapat, C., Manning, G., Mariotti, M., Mata, R., Mathew, T., Neumann, T., Newsham, I., Ngo, D.N., Ninova, M., Okwuonu, G., Ongeri, F., Palmer, W.J., Patil, S., Patraquim, P., Pham, C., Pu, L.L., Putman, N.H., Rabouille, C., Ramos, O.M., Rhodes, A.C., Robertson, H.E., Robertson, H.M., Ronshaugen, M., Rozas, J., Saada, N., Sanchez-Gracia, A., Scherer, S.E., Schurko, A.M., Siggens, K.W., Simmons, D., Stief, A., Stolle, E., Telford, M.J., Tessmar-Raible, K., Thornton, R., van der Zee, M., von Haeseler, A., Williams, J.M., Willis, J.H., Wu, Y., Zou, X., et al. 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. PLoS Biology 12:e1002005.

Cong, P.Y., Xia, X.H., Yang, Q. 2009. Monophyly of the ring-forming group in Diplopoda (Myriapoda, Arthropoda) based on SSU and LSU ribosomal RNA sequences. Progress in Natural Science 19:1297-1303.

Dohle, W. 1980. Sind die Myriapoden eine monophyletische Gruppe? Eine Diskussion der Verwandtschaftsbeziehungen der Antennaten. Abhandlungen und Verhandlungen des Naturwissenschaftlichen Vereins in Hamburg (Neue Folge) 23:45-104.

Dohle, W. 1997. Are the insects more closely related to the crustaceans than to the myriapods? Entomologica Scandinavica Suppl. 51:7-16.

Dong, Y., Sun, H., Guo, H., Pan, D., Qian, C., Hao, S., Zhou, K. 2012. The complete mitochondrial genome of *Pauropus longiramus* (Myriapoda: Pauropoda): Implications on early diversification of the myriapods revealed from comparative analysis. Gene 505:57-65.

Dunlop, J.A., Penney, D., Tetlie, O.E., Anderson, L.I. 2008. How many species of fossil arachnids are there? The Journal of Arachnology 36:267-272.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.

Edgecombe, G.D. 2004. Morphological data, extant Myriapoda, and the myriapod stem-group. Contributions to Zoology 73:207-252.

Edgecombe, G.D. 2011. Phylogenetic relationships of Myriapoda. Pages 1-20 *in* Treatise on Zoology - Anatomy, Taxonomy, Biology. The Myriapoda, Volume 1 (A. Minelli, ed.) Brill.

Edgecombe, G.D., Giribet, G. 2002. Myriapod phylogeny and the relationships of Chilopoda. Pages 143-168 *in* Biodiversidad, taxonomía y biogeografía de artrópodos de México: Hacia una síntesis de su conocimiento (J. E. Llorente Bousquets, and J. J. Morrone, eds.). Prensas de Ciencias, Universidad Nacional Autónoma de México, Mexico D.F.

Edgecombe, G.D., Giribet, G. 2004. Adding mitochondrial sequence data (16S rRNA and cytochrome *c* oxidase subunit I) to the phylogeny of centipedes (Myriapoda, Chilopoda): an analysis of morphology and four molecular loci. Journal of Zoological Systematics and Evolutionary Research 42:89-134.

Edgecombe, G.D., Giribet, G. 2006. A century later — a total evidence re-evaluation of the phylogeny of scutigeromorph centipedes (Myriapoda : Chilopoda). Invertebrate Systematics 20:503-525.

Edgecombe, G.D., Giribet, G. 2008. A New Zealand species of the trans-Tasman centipede order Craterostigmomorpha (Arthropoda : Chilopoda) corroborated by molecular evidence. Invertebrate Systematics 22:1-15.

Edgecombe, G.D., Giribet, G. 2009. Phylogenetics of scutigeromorph centipedes (Myriapoda: Chilopoda) with implications for species delimitation and historical biogeography of the Australian and New Caledonian faunas. Cladistics 25:406-427.

Edgecombe, G.D., Legg, D.A. 2014. Origins and early evolution of arthropods. Palaeontology 57:457-468.

Enghoff, H. 1984. Phylogeny of millipedes - a cladistic analysis. Zeitschrift Fur Zoologische Systematik Und Evolutionsforschung 22:8-26.

Enghoff, H., Dohle, W., Blower, J.G. 1993. Anamorphosis in millipedes (Diplopoda)—the present state of knowledge with some developmental and phylogenetic considerations. Zoological Journal of the Linnean Society 109:103-234.

Fernández, R., Hormiga, G., Giribet, G. 2014a. Phylogenomic analysis of spiders reveals nonmonophyly of orb weavers. Current Biology 24:1772-1777.

Fernández, R., Laumer, C.E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P.P., Pérez-Porro, A.R., Edgecombe, G.D., Giribet, G. 2014b. Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. Molecular Biology and Evolution 31:1500-1513.

Fu, L.M., Niu, B.F., Zhu, Z.W., Wu, S.T., Li, W.Z. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-3152.

Gai, Y., Song, D., Sun, H., Yang, Q., Zhou, K. 2008. The complete mitochondrial genome of *Symphylella* sp. (Myriapoda: Symphyla): Extensive gene order rearrangement and evidence in favor of Progoneata. Molecular Phylogenetics and Evolution 49:574-585.

Giribet, G. 2015. Morphology should not be forgotten in the era of genomics—a phylogenetic perspective. Zoologischer Anzeiger 256:96-103.

Giribet, G., Carranza, S., Riutort, M., Baguñà, J., Ribera, C. 1999. Internal phylogeny of the Chilopoda (Myriapoda, Arthropoda) using complete 18S rDNA and partial 28S rDNA sequences. Philosophical Transactions of the Royal Society of London B Biological Sciences 354:215-222.

Giribet, G., Edgecombe, G.D. 2013. Stable phylogenetic patterns in scutigeromorph centipedes (Myriapoda : Chilopoda : Scutigeromorpha): dating the diversification of an ancient lineage of terrestrial arthropods. Invertebrate Systematics 27:485-501.

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

Goloboff, P.A., Farris, J.S., Nixon, K.C. 2008. TNT, a free program for phylogenetic analysis. Cladistics 24:774-786.

Goloboff, P.A., Szumik, C.A. 2015. Identifying unstable taxa: Efficient implementation of triplet-based measures of stability, and comparison with Phyutility and RogueNaRok. Molecular Phylogenetics and Evolution 88:93-104.

Grünewald, S., Spillner, A., Bastkowski, S., Bogershausen, A., Moulton, V. 2013. SuperQ: computing supernetworks from quartets. IEEE/ACM Transactions on Computational Biology and Bioinformatics / IEEE, ACM 10:151-60.

Guindon, S. 2013. From trajectories to averages: an improved description of the heterogeneity of substitution rates along lineages. Systematic Biology 62:22-34.

Guindon, S., Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696-704.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8:1494-1512.

Haug, J.T., Haug, C., Schweigert, G., Sombke, A. 2014. The evolution of centipede venom claws - Open questions and possible answers. Arthropod Structure & Development 43:5-16.

Hejnol, A., Obst, M., Stamatakis, A., M., O., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E.G., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proceedings of the Royal Society B: Biological Sciences 276:4261-4270.

Hoffman, R.L. 1980. Classification of the Diplopoda. Muséum d'Histoire Naturelle, Genève.

Huson, D.H., Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23:254-267.

Inoue, J., Donoghue, P.C.J., Yang, Z.H. 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Systematic Biology 59:74-89.

Jamieson, B.G.M. 1986. The spermatozoa of the Chilopoda (Uniramia): an ultrastructural review with data on dimorphism in *Ethmostigmus rubripes* and phylogenetic discussion. Journal of Submicroscopic Cytology 18:543-558.

Johnson, E.W., Briggs, D.E.G., Suthren, R.J., Wright, J.L., Tunnicliff, S.P. 1994. Non-marine arthropod traces from the subaerial Ordovician Borrowdale Volcanic Group, English Lake District. Geological Magazine 131:395-406.

Kenrick, P., Wellman, C.H., Schneider, H., Edgecombe, G.D. 2012. A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. Philosophical Transactions of the Royal Society B: Biological Sciences 367:519-36.

Koch, M., Edgecombe, G.D. 2006. Peristomatic structures in Scutigeromorpha (Chilopoda): a comparative study, with new characters for higher-level systematics. Zoomorphology 125:187-207.

Koch, M., Edgecombe, G.D. 2012. The preoral chamber in geophilomorph centipedes: comparative morphology, phylogeny, and the evolution of centipede feeding structures. Zoological Journal of the Linnean Society:no-no.

Koch, M., Edgecombe, G.D., Shelley, R.M. 2010. Anatomy of *Ectonocryptoides* (Scolopocryptopidae: Ectonocryptopinae) and the phylogeny of blind Scolopendromorpha (Chilopoda). International Journal of Myriapodology 3:51-81.

Koch, M., Pärschke, S., Edgecombe, G.D. 2009. Phylogenetic implications of gizzard morphology in scolopendromorph centipedes (Chilopoda). Zoologica Scripta 38:269-288.

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

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

Lartillot, N., Philippe, H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular Biology and Evolution 21:1095-1109.

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

Laumer, C.E., Hejnol, A., Giribet, G. 2015. Nuclear genomic signals of the "microturbellarian" roots of platyhelminth evolutionary innovation. eLife 4:e05503.

Lee, Michael S.Y., Palci, A. 2015. Morphological phylogenetics in the genomic age. Current Biology 25:R922-R929.

Lee, M.S.Y., Soubrier, J., Edgecombe, G.D. 2013. Rates of phenotypic and genomic evolution during the Cambrian Explosion. Current Biology 23:1-7.

Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Systematic Biology 58:130-145.

Lemmon, E.M., Lemmon, A.R. 2013. High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44:99-121.

Lepage, T., Bryant, D., Philippe, H., Lartillot, N. 2007. A general comparison of relaxed molecular clock models. Molecular Biology and Evolution 24:2669-2680.

López-Giráldez, F., Moeller, A.H., Townsend, J.P. 2013. Evaluating phylogenetic informativeness as a predictor of phylogenetic signal for metazoan, fungal, and mammalian phylogenomic data sets. Biomed Research International 2013:621604.

Maddison, W.P. 1993. Missing data versus missing characters in phylogenetic analysis. Systematic Biology 42:576-581.

Mark, D.F., Rice, C.M., Fallick, A.E., Trewin, N.H., Lee, M.R., Boyce, A., Lee, J.K.W. 2011. 40Ar/39Ar dating of hydrothermal activity, biota and gold mineralization in the Rhynie hot-spring system, Aberdeenshire, Scotland. Geochimica et Cosmochimica Acta 75:555-569.

Mark, D.F., Rice, C.M., Trewin, N.H. 2013. Discussion on 'A high-precision U-Pb age constraint on the Rhynie Chert Konservat-Lagerstatte: time scale and other implications' Journal, Vol. 168, 863-872. Journal of the Geological Society 170:701-703.

Meusemann, K., von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P., Ebersberger, I., Walzl, M., Pass, G., Breuers, S., Achter, V., von Haeseler, A., Burmester, T., Hadrys, H., Wägele, J.W., Misof, B. 2010. A phylogenomic approach to resolve the arthropod tree of life. Molecular Biology and Evolution 27:2451-2464.

Miyazawa, H., Ueda, C., Yahata, K., Su, Z.H. 2014. Molecular phylogeny of Myriapoda provides insights into evolutionary patterns of the mode in post-embryonic development. Scientific Reports 4:4127.

Muller, K.J. 1983. Crustacea with preserved soft parts from the Upper Cambrian of Sweden. Lethaia 16:93-109.

Mundel, P. 1979. The centipedes (Chilopoda) of the Mazon Creek. Pages 361-378 in Mazon Creek Fossils (M. Nitecki, ed.) Academic Press, New York.

Murienne, J., Edgecombe, G.D., Giribet, G. 2010. Including secondary structure, fossils and molecular dating in the centipede tree of life. Molecular Phylogenetics and Evolution 57:301-313.

Nielsen, A.T., Weidner, T., Terfelt, F., Hoyberget, M. 2014. Upper Cambrian (Furongian) biostratigraphy in Scandinavia revisited: definition of superzones. GFF 136:193-197.

Norell, M.A., Gladstein, D., Wheeler, W.C. 1995. On missing data in phylogenetic analysis. Journal of Vertebrate Paleontology 15:46A.

Norell, M.A., Wheeler, W. 2003. Missing entry replacement data analysis: A replacement approach to dealing with missing data in paleontological and total evidence data sets. Journal of Vertebrate Paleontology 23:275-283.

Olesen, J. 2009. Phylogeny of Branchiopoda (Crustacea) – character evolution and contribution of uniquely preserved fossils. Arthropod Systematics & Phylogeny 67:3-39.

Parry, S.F., Noble, S.R., Crowley, Q.G., Wellman, C.H. 2011. A high-precision U–Pb age constraint on the Rhynie Chert Konservat-Lagerstätte: time scale and other implications. Journal of the Geological Society, London 168:863-872.

Parry, S.F., Noble, S.R., Crowley, Q.G., Wellman, C.H. 2013. Reply to Discussion on 'A high-precision U-Pb age constraint on the Rhynie Chert Konservat-Lagerstatte: time scale and other implications' Journal, 168, 863-872. Journal of the Geological Society 170:703-706.

Peach, B.N. 1882. On some fossil myriapods from the lower old red sandstone. Proceedings Physical Society of Edinburgh vii:pp. 177-188.

Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W.H., Casane, D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Molecular Biology and Evolution 21:1740-1752.

Pyron, R.A. 2015. Post-molecular systematics and the future of phylogenetics. Trends in Ecology & Evolution 30:384-389.

Regier, J.C., Shultz, J.W., Ganley, A.R., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J.E., Cummings, M.P., Martin, J.W., Cunningham, C.W. 2008. Resolving arthropod phylogeny: Exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Systematic Biology 57:920-938.

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

Regier, J.C., Wilson, H.M., Shultz, J.W. 2005. Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. Molecular Phylogenetics and Evolution 34:147-158.

Rehm, P., Borner, J., Meusemann, K., von Reumont, B.M., Simon, S., Hadrys, H., Misof, B., Burmester, T. 2011. Dating the arthropod tree based on large-scale transcriptome data. Molecular Phylogenetics and Evolution 61:880-887.

Rehm, P., Meusemann, K., Borner, J., Misof, B., Burmester, T. 2014. Phylogenetic position of Myriapoda revealed by 454 transcriptome sequencing. Molecular Phylogenetics and Evolution 77C:25-33.

Richardson, J.B., Bonamo, P.M., McGregor, D.C. 1993. The spores of *Leclercqia* and the dispersed spore morphon *Acinosporites lindlarensis* Riegel: a case of gradualistic evolution. Bulletin of the Natural History Museum, London (Geology):121-155.

Rokas, A., Williams, B.L., King, N., Carroll, S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798-804.

Rota-Stabelli, O., Campbell, L., Brinkmann, H., Edgecombe, G.D., Longhorn, S.J., Peterson, K.J., Pisani, D., Philippe, H., Telford, M.J. 2011. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. Proceedings of the Royal Society B: Biological Sciences 278:298-306.

Rota-Stabelli, O., Lartillot, N., Philippe, H., Pisani, D. 2013. Serine codon-usage bias in deep phylogenomics: pancrustacean relationships as a case study. Systematic Biology 62:121-133.

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

Shabica, C.W., Hay, A.A. 1997. Richardson's guide to the fossil fauna of Mazon Creek. Northeastern Illinois University, Chicago.

Sharma, P.P., Giribet, G. 2014. A revised dated phylogeny of the arachnid order Opiliones. Frontiers in Genetics 5:255.

Shear, W. 2011. Class Diplopoda de Blainville in Gervais, 1844. Pages 159-164 *in* Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (Z.-Q. Zhang, ed.) Magnolia Press, Auckland.

Shear, W.A. 2008. Spinnerets in the milliped order Polydesmida, and the phylogenetic signifi cance of spinnerets in millipeds (Diplopoda). International Journal of Myriapodology 2:123-146.

Shear, W.A., Bonamo, P.M. 1988. Devonobiomorpha, a new order of centipeds (Chilopoda) from the middle Devonian of Gilboa, New York State, USA, and the phylogeny of centiped orders. American Museum Novitates 2927:1-30.

Shear, W.A., Edgecombe, G.D. 2010. The geological record and phylogeny of Myriapoda. Arthropod Structure & Development 39:174-190.

Shear, W.A., Jeram, A.J., Selden, P.A. 1998. Centipede legs (Arthropoda, Chilopoda, Scutigeromorpha) from the Silurian and Devonian of Britain and the Devonian of North America. American Museum Novitates 3231:1-16.

Shi, G., Grimaldi, D.A., Harlow, G.E., Wang, J., Wang, J., Yang, M., Lei, W., Li, Q., Li, X. 2012. Age constraint on Burmese amber based on U–Pb dating of zircons. Cretaceous Research 37:155-163.

Siddall, M.E., Whiting, M.F. 1999. Long-branch abstractions. Cladistics 15:9-24.

Sierwald, P., Bond, J.E. 2007. Current status of the myriapod class Diplopoda (millipedes): taxonomic diversity and phylogeny. Annual Review of Entomology 52:401-420.

Sierwald, P., Shear, W.A., Shelley, R.M., Bond, J.E. 2003. Millipede phylogeny revisited in the light of the enigmatic order Siphoniulida. Journal of Zoological Systematics and Evolutionary Research 41:87-99.

Simmons, M.P. 2012. Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics 28:208-222.

Smith, S.A., Dunn, C.W. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24:715-716.

Soubrier, J., Steel, M., Lee, M.S.Y., Sarkissian, C.D., Guindon, S., Ho, S.Y.W., Cooper, A. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. Molecular Biology and Evolution 29:3345-3358.

Stamatakis, A. 2014. ExaBayes User's Manual.

Townsend, J.P. 2007. Profiling phylogenetic informativeness. Systematic Biology 56:222-231.

Townsend, J.P., Su, Z., Tekle, Y.I. 2012. Phylogenetic signal and noise: Predicting the power of a data set to resolve phylogeny. Systematic Biology 61:835-849.

Vahtera, V., Edgecombe, G.D., Giribet, G. 2012. Evolution of blindness in scolopendromorph centipedes (Chilopoda: Scolopendromorpha): insight from an expanded sampling of molecular data. Cladistics 28:4-20.

Vahtera, V., Edgecombe, G.D., Giribet, G. 2013. Phylogenetics of scolopendromorph centipedes: Can denser taxon sampling improve an artificial classification? Invertebrate Systematics 27:578-602.

Verhoeff, K.W. 1928. Klasse Diplopoda I. Pages 1071 *in* Bronn's Klassen und Ordnungen des Tier-Reichs 5. Gliederfüssler: Arthropoda, 2. Myriapoda Akademische Verlagsgesellschaft, Leipzig.

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

Wanninger, A. 2015. Morphology is dead – long live morphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics. Frontiers in Ecology and Evolution 3:54.

Werth, A.J., Shear, W.A. 2014. The evolutionary truth about living fossils. American Scientist 102:434-443.

Whitfield, R.P. 1885. An American Silurian scorpion (*Palaeophonus osborni*). Science 6:87.

Wiens, J.J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Systematic Biology 52:528-538.

Wiens, J.J. 2006. Missing data and the design of phylogenetic analyses. Journal of Biomedical Informatics 39:34-42.

Wilson, H.M. 2006. Juliformian millipedes from the Lower Devonian of Euramerica: implications for the timing of millipede cladogenesis in the Paleozoic. Journal of Paleontology 80:638-649.

Wilson, H.M., Anderson, L.I. 2004. Morphology and taxonomy of Paleozoic millipedes (Diplopoda : Chilognatha : Archipolypoda) from Scotland. Journal of Paleontology 78:169-184.

Woo, H.J., Lee, Y.S., Park, S.J., Lim, J.T., Jang, K.H., Choi, E.H., Choi, Y.G., Hwang, U.W. 2007. Complete mitochondrial genome of a troglobite millipede *Antrokoreana gracilipes* (Diplopoda, Juliformia, Julida), and juliformian phylogeny. Molecules and Cells 23:182-191.

Wu, M., Chatterji, S., Eisen, J.A. 2012. Accounting for alignment uncertainty in phylogenomics. PLoS ONE 7:e30288.

Wu, Z.P., Wang, X., Zhang, X.G. 2011. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. Bioinformatics 27:502-508.

Yang, Z., Rannala, B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution 23:212-226.

Zoller, S., Schneider, A. 2013. Improving phylogenetic inference with a semiempirical amino acid substitution model. Molecular Biology and Evolution 30:469-479.

Zwick, A., Regier, J.C., Zwickl, D.J. 2012. Resolving discrepancy between nucleotides and amino acids in deep-level arthropod phylogenomics: Differentiating serine codons in 21-amino-acid models. PLoS ONE 7:e47450.

Figure 1. Live habitus of myriapod exemplars from this study. A, *Scutigerella* sp. from Great Smoky Mountains National Park, Tennessee, USA; B, *Eudigraphis taiwaniensis*, from Kenting National Park, Taiwan; C, Sphaerotheriid from Helderberg Nature Reserve, South Africa; D, *Brachycybe lecontii* from Great Smoky Mountains National Park, E, *Abacion* sp. from Great Smoky Mountains National Park; F, *Scutigerina weberi* from KwaZulu-Natal, South Africa; G, *Craterostigmus crabilli* from Kahurangi National Park, New Zealand; H, *Theatops spinicaudus* from Great Smoky Mountains National Park; I, *Notiphilides grandis* from Reserva Ducke, Manaus, Amazonas, Brazil.
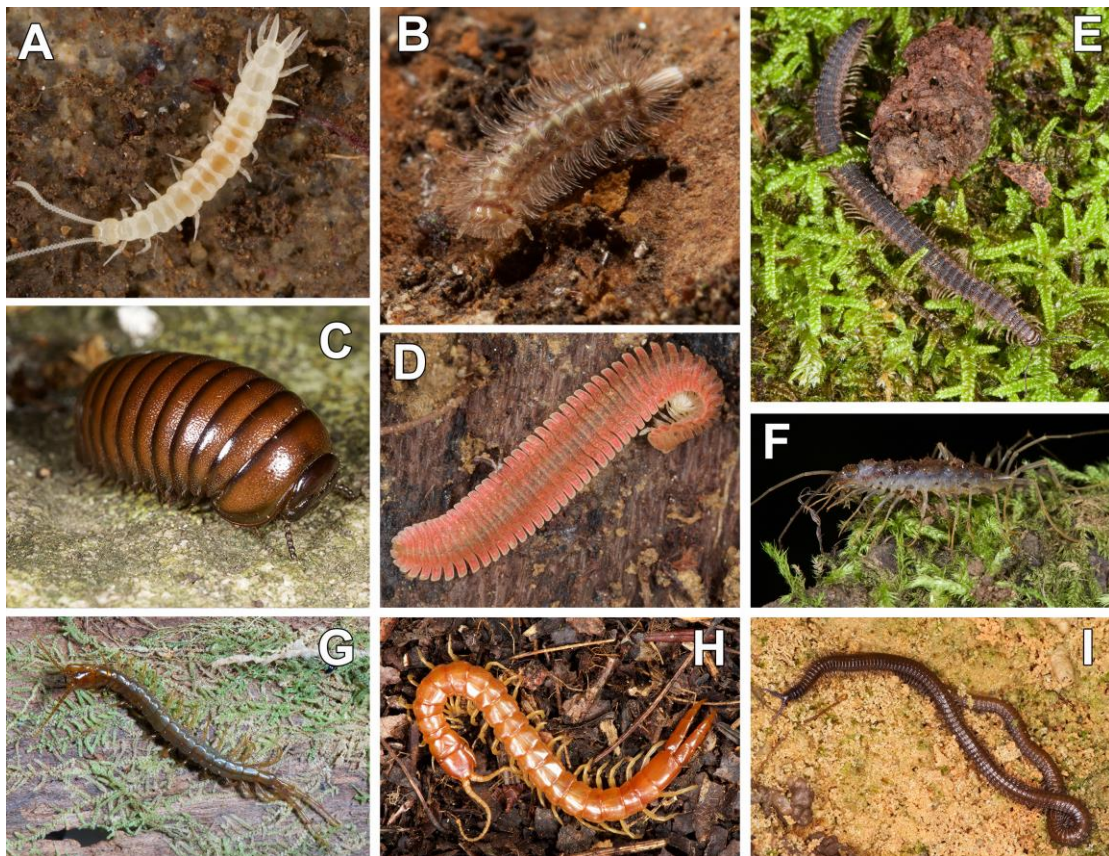
Figure 2. Summary of the evolutionary relationships among myriapods. Up. Maximum likelihood phylogenomic hypothesis of myriapod interrelationships (supermatrix I, Ln L ExaML -15764355.248567). Checked matrices in each node represent nodal support for the different analyses in supermatrices I to VIII (see Material and Methods for further information). PP, PhyML-PCMA; EM, ExaML; PB, PhyloBayes; EB, ExaBayes; and PIL, ML analysis with integrated branch length as implemented in PhyML. Filled squares indicate nodal support values higher than 0.95/ 0.90/95 (posterior probability, PB and EB/Shimodaira-Hasegawa-like support, PP/bootstrap, EM and PIL). White squares indicate lower nodal support. Nodal support for the clades represented with letters A to U, and the alternative topologies to the three conflicting nodes (named 1, 2 and 3) are shown.

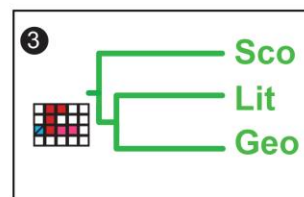Figure 3. Chronogram of myriapod evolution for the 123-gene data set with 95% highest posterior density (HPD) bar for the dating under the uncorrelated gamma model. Nodes that were calibrated with fossils are indicated with a star placed at the age of the fossil.
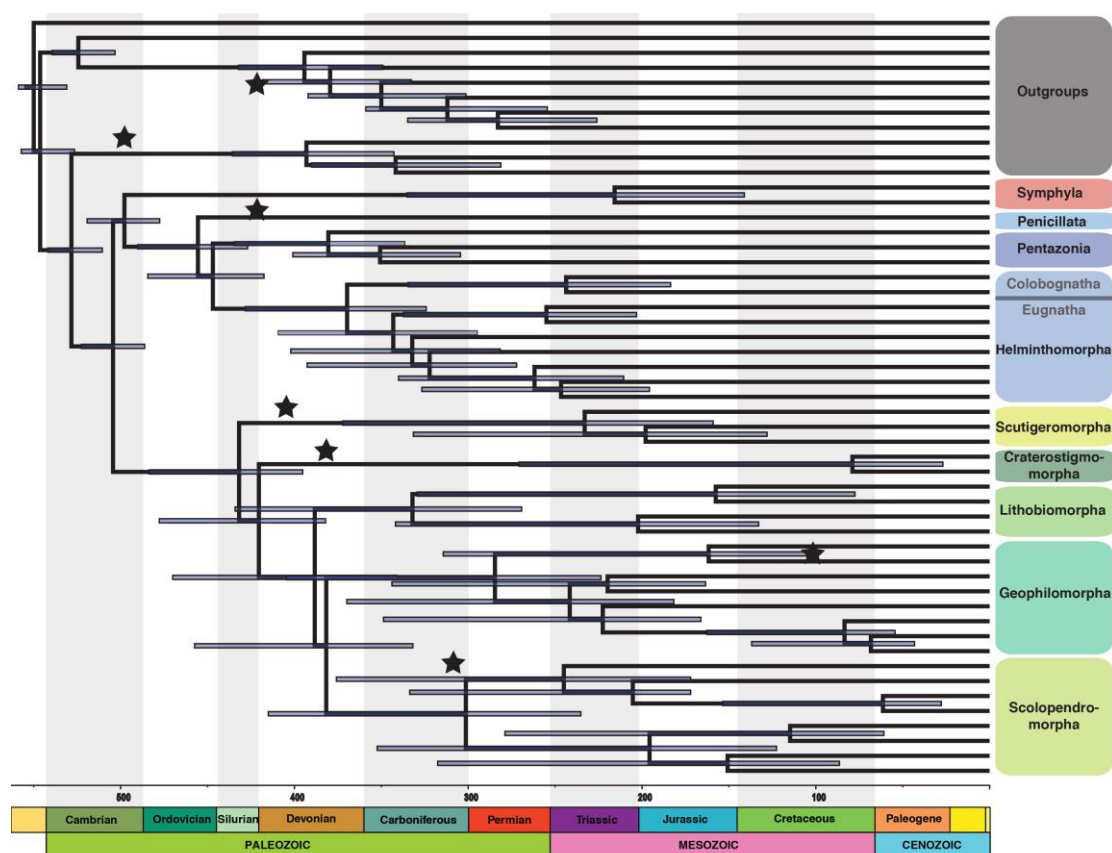
Figure S1. Phylogenetic hypothesis of the interrelationships of Pauropoda, Symphyla, Diplopoda and Chilopoda inferred from a 4-gene data set using ExaBayes.
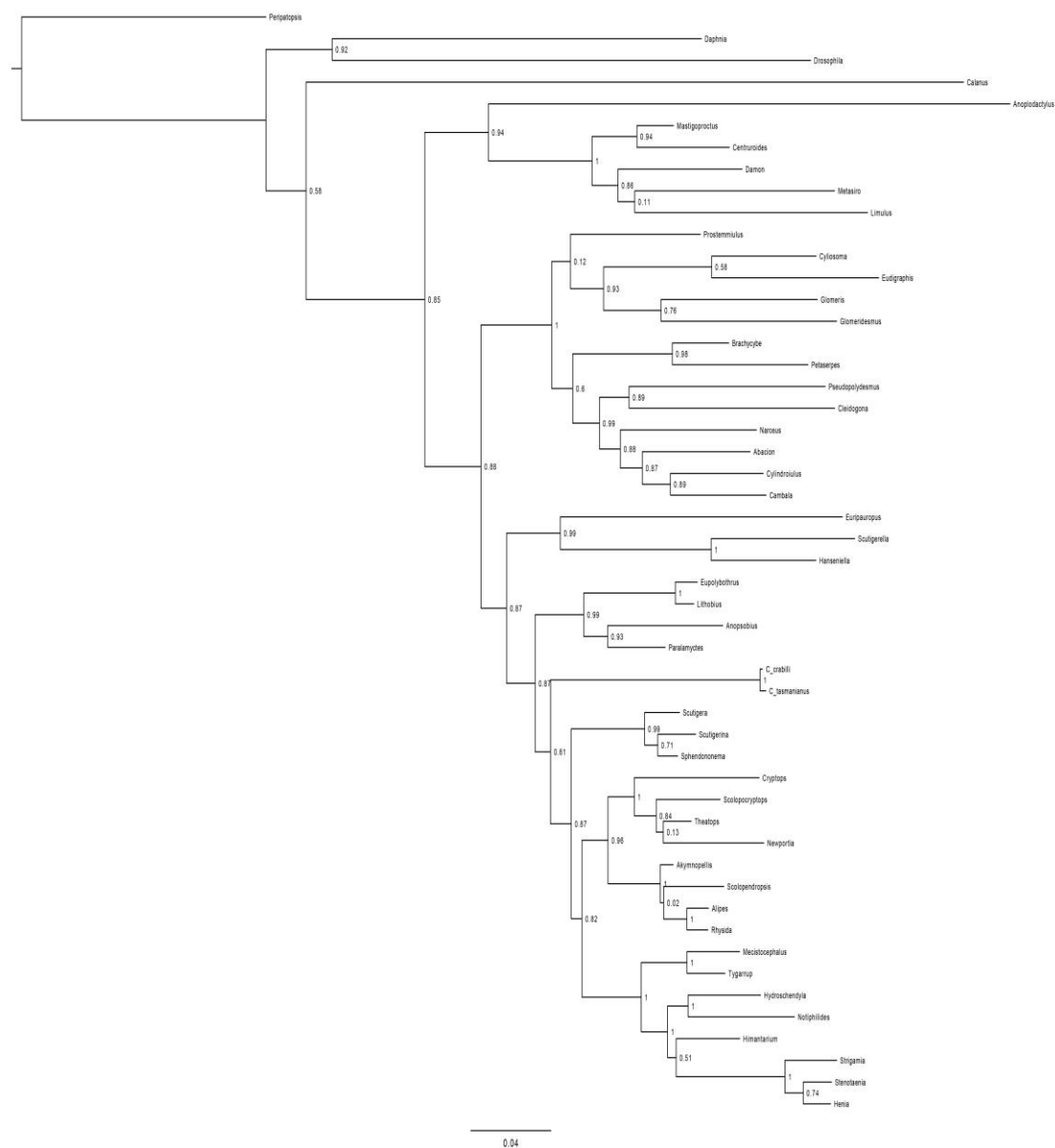
Figure S2. Phylogenetic informativeness profile of the genes included in supermatrices III (a), VI (b) and V (c) during the interval of time corresponding to the diversification of all myriapod clades (as delimited by blue lines in the ultrametric tree in (a)).
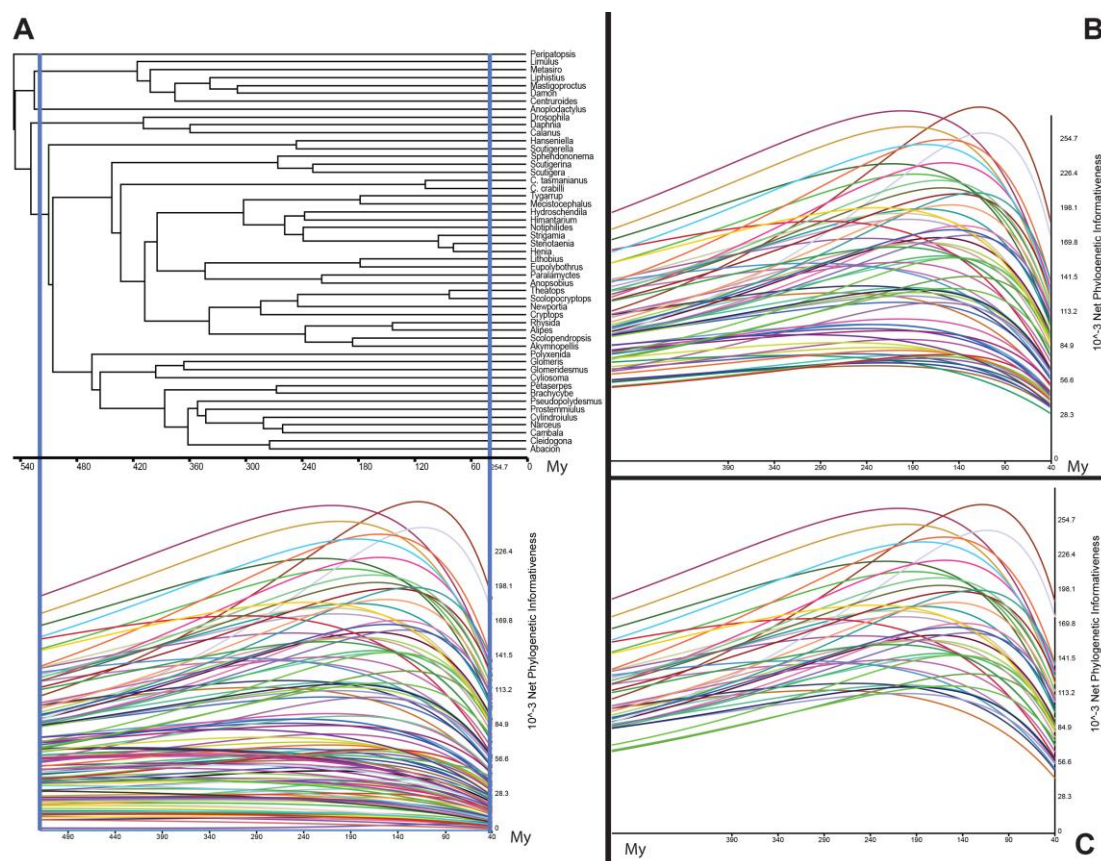
Figure S3. Phylogenetic hypothesis of Myriapoda based on 232 morphological characters coded for both extant and extinct species (see Material and Methods for further details). Strict consensus of 2515 shortest cladograms (420 steps). Fossil taxa are marked with a dagger symbol.

Peripatopsis capensis
Anoplodactylus insignis
Limulus polyphemus
Metasiro americanus
Centruroides vittatus
Proscorpius osborni †
Liphistius malayanus
Damon variegatus
Mastigoproctus giganteus
Calanus finmarchicus
Daphnia pulex
Rehbachiella kinnekullensis †
Apis mellifera
Drosophila melanogaster
Eudigraphis taiwanensis
Glomeridesmus sp.
Glomeris marginata
Cyliosoma sp.
Cowiedesmus eroticopodus †
Archidesmus macnicoli †
Pseudopolydesmus sp.
Gaspestria genselorum †
Narceus americanus
Cambala annulata
Cylindroiulus sp.
Petaserpes sp.
Brachycybe lecontii
Cleidogona sp.
Prostemmiulus sp.
Abacion magnum
Hanseniella sp.
Scutigerella sp.
Crussolum spp. †
Scutigerina weberi
Sphendononema guildingii
Scutigera coleoptrata
Anopsobius giribeti
Paralamyctes validus
Lithobius forficatus
Eupolybothrus cavernicolus
Craterostigmus tasmanianus
Craterostigmus crabilli
Devonobius delta †
Mecistocephalus guildingii
Tygarrup javanicus
Hydroschendyla submarina
Notiphilides grandis
Himantarium gabrielis
Stenotaenia linearis
Henia brevis
Strigamia maritima
Kachinophilus pereirai †
Mazoscolopendra richardsoni †
Cryptops hortensis
Newportia adisi
Theatops spinicaudus
Scolopocryptops sexspinosus
Rhysida longipes
Alipes grandidieri
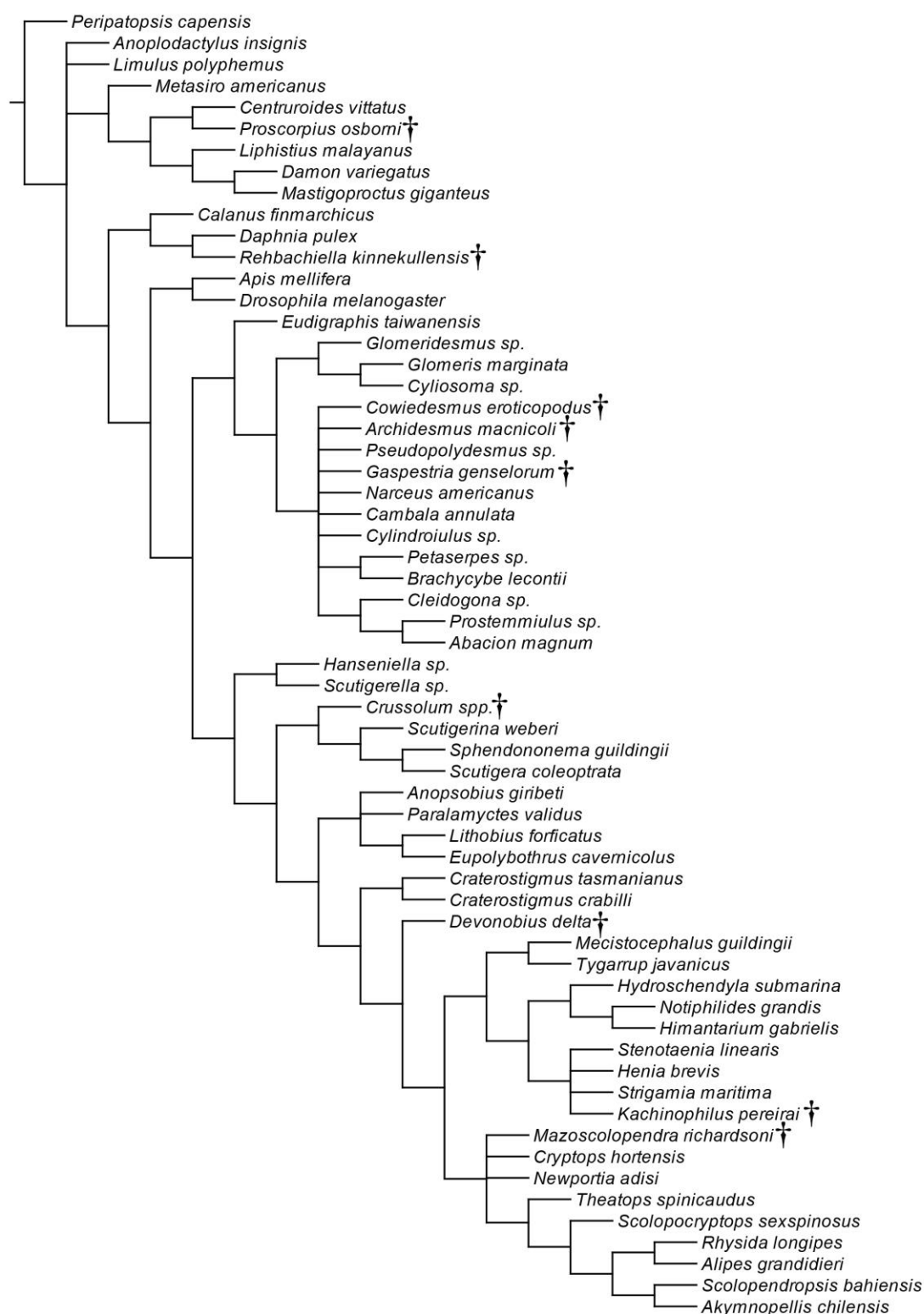Scolopendropsis bahiensis
Akymnopellis chilensis

46

Figure S4. Topologies recovered in the SAW analyses to test for LBA between a) Symphyla and Polyxenida, and b) Scutigeromorpha and Craterostigmomorpha.
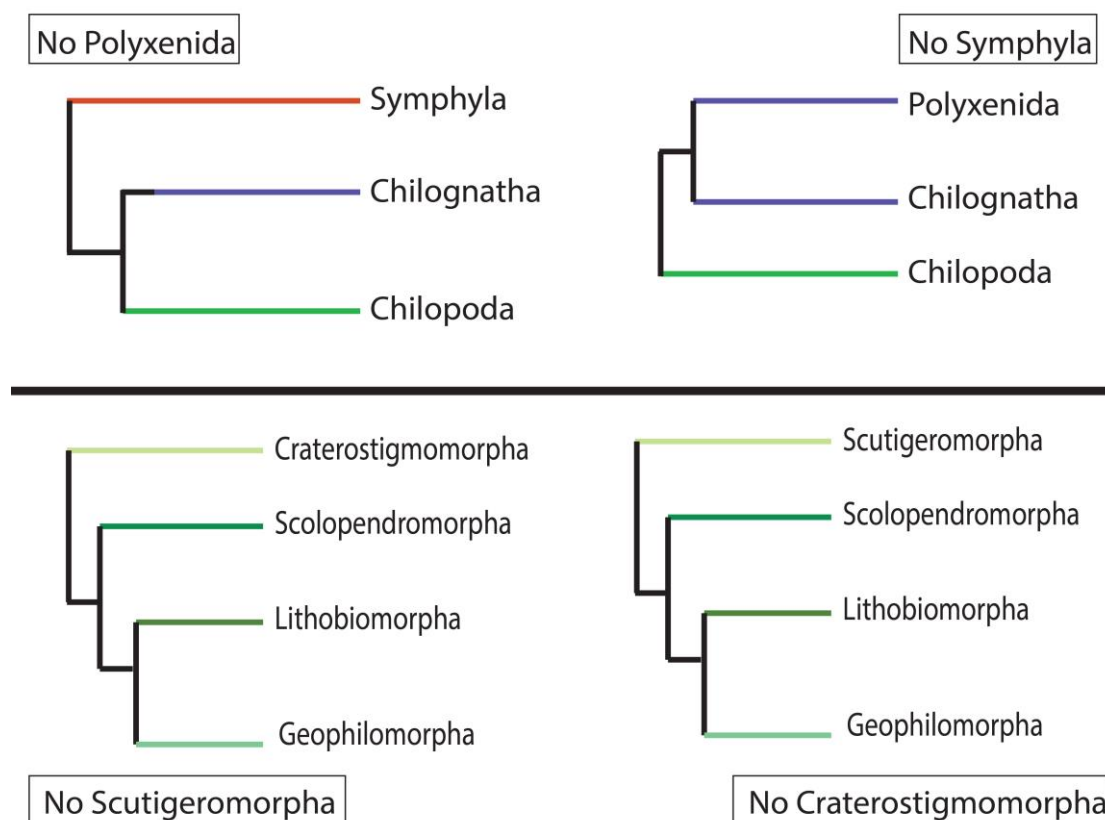
Figure S5. BaCoCa results showing compositional homogeneity values per taxon and per gene in supermatrices II and III (results in supermatrix I not shown due to a lack of resolution related to the high number of genes included in the matrix).
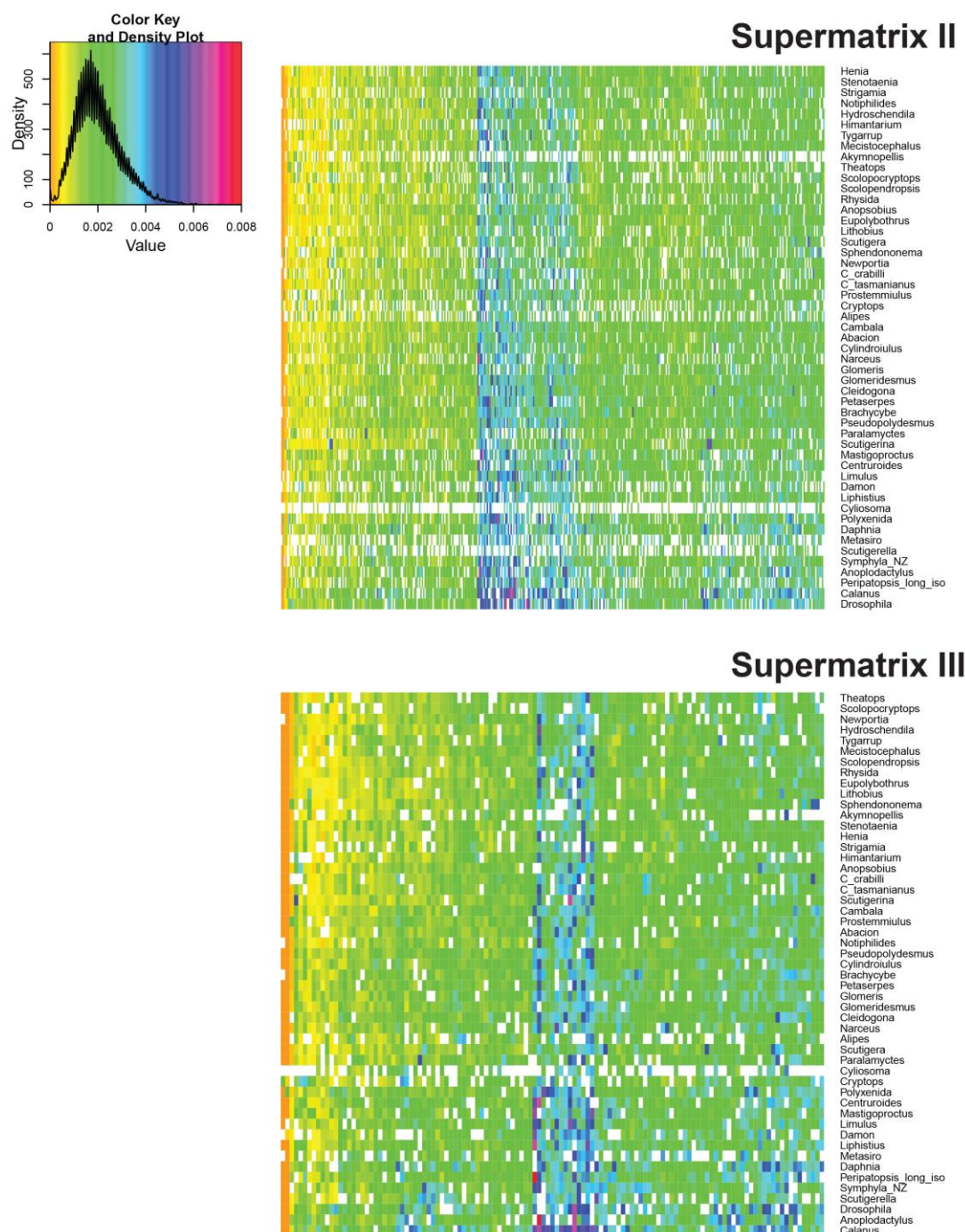
Figure S6. Supernetwork representation of quartets derived from individual ML gene trees, for the genes concatenated in supermatrices I (a), II (b) and III (c). Phylogenetic conflict is represented by reticulations and short branches.
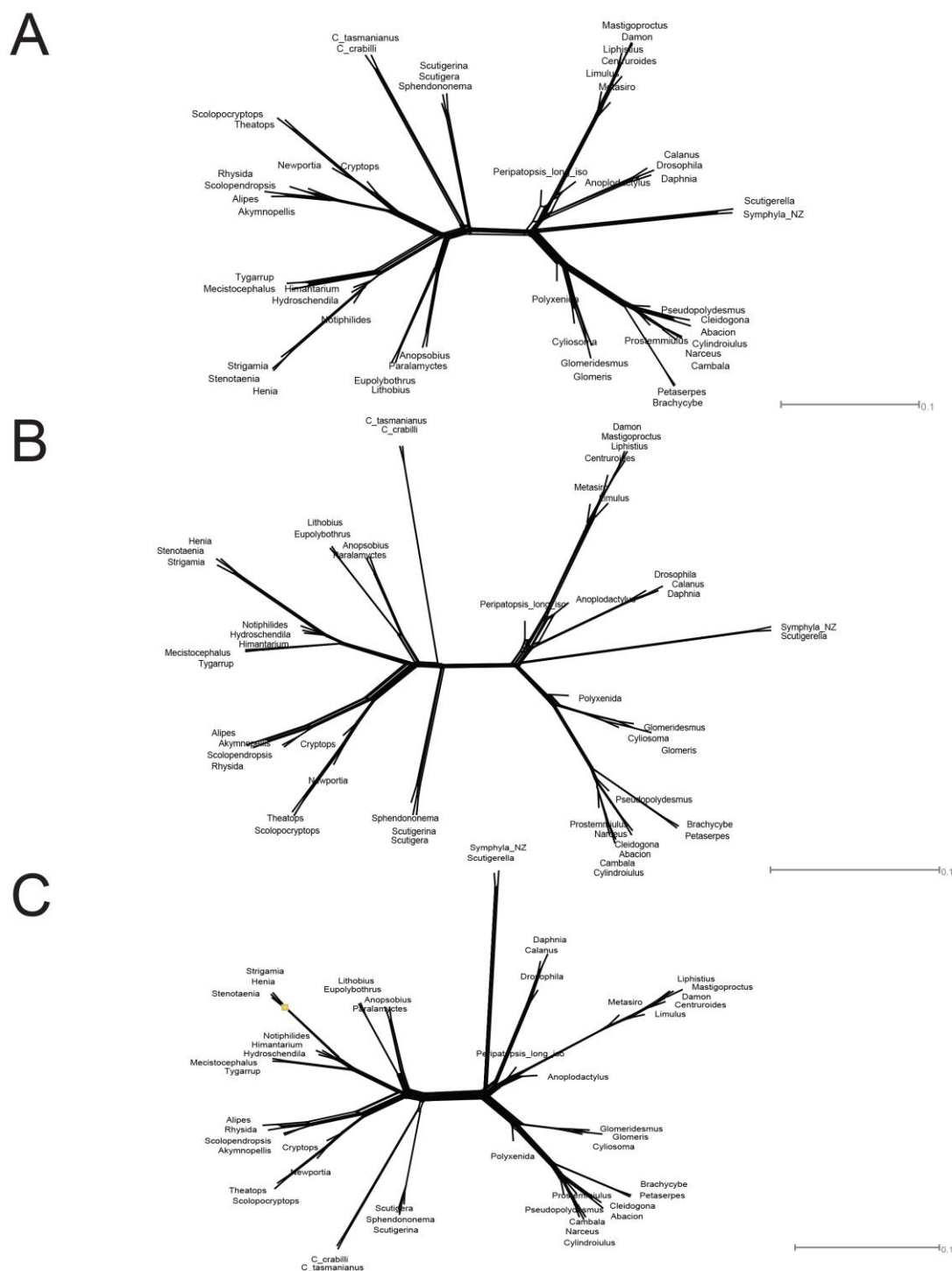
Table 1. List of specimens sequenced and analyzed in the present study. The catalogue numbers and links to the Museum of Comparative Zoology (MCZ) and Sequence Read Archive (SRA) accession numbers are shown.

Species	Source	MCZ voucher	SRA #	MCZ link
Sphendononema guildingii	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Scutigerina weberi	Illumina HiSeq (this study)	IZ-32088	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:32088
Scutigera coleoptrata	Illumina HiSeq (this study)	IZ-204015	SRR1158078	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:20415
Paralamyctes validus	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Anopsobius giribeti	Illumina HiSeq (this study)	IZ-35360	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:35360
Lithobius forficatus	Illumina HiSeq (Fernández et al., 2014)	IZ-131534	SRR1159752	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:131534
Eupolybothrus cavernicolus	Illumina (Stoev et al., 2013)	-	ERX311347	-
Craterostigmus tasmanianus	Illumina HiSeq (Fernández et al., 2014)	IZ-128299	SRR1157986	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:130583
Craterostigmus crabilli	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Cryptops hortensis	Illumina HiSeq (Fernández et al., 2014)	IZ-130583	SRR1153457	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:130583
Theatops spinicaudus	Illumina HiSeq (this study)	IZ-46873	*to assign*	*to assign*
Newportia adisi	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Scolopocryptops sexspinosus	Illumina HiSeq (this study)	IZ-44070	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:44070
Akymnopellis chilensis	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Scolopendropsis bahiensis	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Alipes grandidieri	Illumina HiSeq (Fernández et al., 2014)	IZ-130616	SRR619311	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:130616
Rhysida longipes	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Tygarrup javanicus	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Mecistocephalus guildingii	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Notiphilides grandis	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Himantarium gabrielis	Illumina HiSeq (Fernández et al., 2014)	IZ-131564	SRR1159787	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:131564

50

Hydroschendyla submarina	Illumina HiSeq (this study)	IZ-32096	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:32096
Stenotaenia linearis	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Henia brevis	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Strigamia maritima	Genome available	-	-	
Eudigraphis taiwaniensis	Illumina HiSeq (this study)	IZ-128912	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:128912
Glomeris marginata	Illumina HiSeq (this study)	IZ-43690	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:43690
Cyliosoma sp.	Illumina HiSeq (this study)	IZ-44064	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:44064
Glomeridesmus sp.	Illumina (Brewer & Bond, 2013)	-	SRX3267 75	
Prostemmiulus sp.	Illumina (Brewer & Bond, 2013)	-	SRX3267 82	
Pseudopolydesmus sp.	Illumina (Brewer & Bond, 2013)	-	SRX3267 79	
Petaserpes sp.	Illumina (Brewer & Bond, 2013)	-	SRX3267 77	
Brachycybe lecontii	Illumina (Brewer & Bond, 2013)	-	SRX3267 76	
Abacion magnum	Illumina (Brewer & Bond, 2013)	-	SRX3267 81	
Cleidogona sp.	Illumina (Brewer & Bond, 2013)	-	SRX3267 80	
Cambala annulata	Illumina (Brewer & Bond, 2013)	-	SRX3267 83	
Cylindroiulus sp.	Illumina HiSeq (this study)	IZ-43724	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:43724
Narceus americanus	Illumina HiSeq (this study)	IZ-44069	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:44069
Hanseniella sp.	Illumina HiSeq (this study)	to be assigned	*to assign*	*to assign*
Scutigerella sp.	Illumina HiSeq (this study)	IZ-46890	*to assign*	http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:46890
Outgroups				
Peripatopsis overbergiensis	Sharma et al. (2014)			
Anoplodactylus insignis	Illumina HiSeq (this study)			
Damon variegatus	Sharma et al. (2014)			
Liphistius malayanus	Sharma et al. (2014)			
Mastigoproctus giganteus	Sharma et al. (2014)			
Metasiro americanus	Sharma et al. (2014)			
Centruroides vittatus	Sharma et al. (2014)			

Limulus polyphemus	Sharma et al. (2014)			
Daphnia pulex	Genome available			
Calanus finmarchicus	Lentz et al. (2014)			
Drosophila melanogaster	Genome available			