

Genome of the Rusty Millipede, *Trigoniulus corallinus*, Illuminates Diplopod, Myriapod, and Arthropod Evolution

Nathan J. Kenny¹, Xin Shen², Thomas T.H. Chan¹, Nicola W.Y. Wong¹, Ting Fung Chan², Ka Hou Chu³, Hon-Ming Lam², and Jerome H.L. Hui^{1,*}

¹Simon F.S. Li Marine Science Laboratory of School of Life Sciences and Center of Soybean Research of State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong

²Center of Soybean Research of State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong

³Simon F.S. Li Marine Science Laboratory of School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong

*Corresponding author: E-mail: jeromehui@cuhk.edu.hk.

Accepted: April 11, 2015

Data deposition: Data has been deposited in NCBI databases (Bioproject PRJNA260872, Biosample SAMN03048671, experiment SRX700727) and made available for download as noted in text.

Abstract

The increasing availability of genomic information from the Arthropoda continues to revolutionize our understanding of the biology of this most diverse animal phylum. However, our sampling of arthropod diversity remains uneven, and key clade such as the Myriapoda are severely underrepresented. Here we present the genome of the cosmopolitanly distributed Rusty Millipede *Trigoniulus corallinus*, which represents the first diplopod genome to be published, and the second example from the Myriapoda as a whole. This genomic resource contains the majority of core eukaryotic genes (94.3%), and key transcription factor classes that were thought to be lost in the Ecdysozoa. Mitochondrial genome and gene family (transcription factor, Dscam, circadian clock-driving protein, odorant receptor cassette, bioactive compound, and cuticular protein) analyses were also carried out to shed light on their states in the Diplopoda and Myriapoda. The ready availability of *T. corallinus* recommends it as a new model for evolutionary developmental biology, and the data set described here will be of widespread utility in investigating myriapod and arthropod genomics and evolution.

Key words: millipede, Myriapoda, Diplopoda, genome, mitochondrial genome, Arthropoda.

Introduction

The Myriapoda contains greater than 13,000 described species and is one of the most speciose metazoan subphyla. Recent molecular analysis suggests that the Myriapoda is the sister group to the Pancrustacea (Regier et al. 2010; fig. 1A). Precise myriapod interrelationships remain uncertain (Rehm et al. 2014), but the Myriapoda is generally known to contain 1) approximately 3,000 centipede species in the predatory order Chilopoda; 2) 200 species of symphalans, which have some similarities to centipedes, but with rather fewer segments and a constrained pattern of limb growth (Barnes 1982); 3) the order Pauropoda, which consists of around 700 species of soil-dwelling arthropods; and 4) the most species-rich order, Diplopoda, with around 8,000 millipede species (Chapman 2005). Despite all their diversity (Brewer and Bond 2013), only a single myriapod genome (that of centipede *Strigamia maritima*) has been publically released (Chipman

et al. 2014), limiting our ability to draw inferences into myriapod biology, as well as providing only a single outgroup from this clade for comparison to the Pancrustacea.

Millipedes can be found worldwide performing vital ecological roles as detritovores (Snyder et al. 2009; Shelley and Golovatch 2011), and can be easily distinguished from other myriapods—segments posterior to the fourth from the head have two pairs of limb per segment (diplosegments) (Barnes 1982). The oldest terrestrial metazoan fossil is that of a millipede (*Pneumodesmus newmani*) dated at approximately 428 Myr of age (Wilson and Anderson 2004). Millipedes have the ability to synthesize a range of defensive chemical components to ward off predators (Shear et al. 2007, 2010), and a genomic resource will be useful to understand the metabolism of these organisms for a range of research into these novel pathways, which could represent a source for fungicides and other novel bioproducts (Roncadori et al. 1985).

© The Author(s) 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

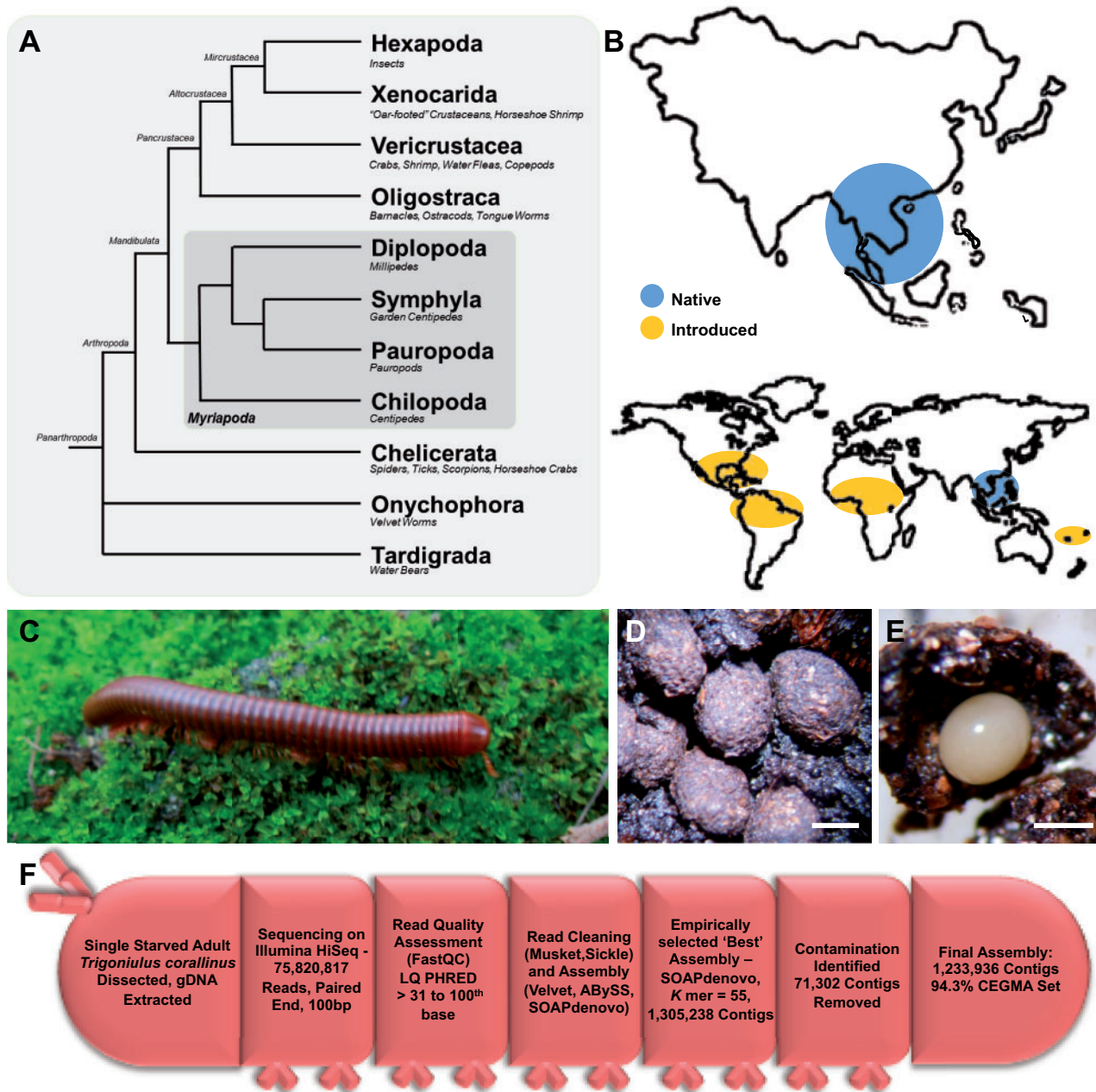


Fig. 1.—*Trigoniuulus corallinus* biology. (A) Phylogenetic relationships of diplopods and related clades. Phylogeny based on Regier et al. (2010) (not all clades shown). (B) *Trigoniuulus corallinus* distribution. Native range indicated broadly (blue), with known introduced populations indicated in yellow. Note that introduced *T. corallinus* may well be found elsewhere. (C) Adult *T. corallinus*, approximately 5 cm in length. (D) Egg capsules gathered in captivity as described in Materials and Methods. (E) Egg case dissected to show single egg. In both (D) and (E), white scale bar represents 1 mm in length. (F) Overview of genomic DNA sequencing, genome assembly procedures, and final data figures from this study.

Here we present the genomic sequence of one such species, the Rusty Millipede *Trigoniuulus corallinus* (Gervais 1847, also called the amber or coral millipede). This species is found broadly around Southeast Asia, from Myanmar to Taiwan (Shelley and Lehtinen 1999), and has been introduced to the Caribbean, Central and South America, islands around the Pacific, and most recently to the southern United States

(Shelley et al. 2006). Early reports also recorded this species in Kew Gardens in the early 1900s (Pocock 1902, 1906), but more recent evidence of their presence is lacking, suggesting that this colony may be extinct (Stoev et al. 2010). Adults usually grow to approximately 5 cm in length, and the embryonic development of this species has been subject to a small amount of previous study (e.g., Shinohara et al. 2007).

Information from a diplopod outgroup is of wide utility for the discernment of ancestral arthropod characters when compared with better known insect model organisms and new crustacean models, such as the water flea *Daphnia* and shrimp *Neocaridina* (Colbourne et al. 2011; Kenny et al. 2014). Furthermore, the cosmopolitanly distributed diplopod *T. corallinus* is easily cultivable and will breed in captivity, providing a ready resource for embryological study. *Trigoniulus corallinus* therefore represents a potentially intriguing model for future work in developmental, ecological and evolutionary spheres of scientific investigation, and the initial genomic resources presented here will be of great interest to a variety of fields.

Materials and Methods

Animal Husbandry

An adult *T. corallinus* was sourced locally from The Chinese University of Hong Kong campus, and our species identification was confirmed by the Agriculture, Fisheries and Conservation Department of the Hong Kong Government. *Trigoniulus corallinus* can be seen in large numbers mating in pairs in the late afternoon from late May to early June in wooded areas around Hong Kong. Millipedes were collected as mating pairs in the late afternoon, and kept in 5-L plastic aquaria at room temperature, with petri dishes filled with moistened autoclaved soil providing both a basic source of nutrition and a breeding substrate. Millipedes were also fed on apple slices with a small quantity of Zoo Med Repti Calcium. Eggs were laid in clusters of 5–20 at the base of petri dishes approximately 1 week after collection, which could be distinguished from fecal pellets by their larger size, clustering together in a small area, and relatively dry appearance. Eggs can be removed from their capsules by gentle dissection in 1×phosphate buffered saline. Photographs were taken using a Nikon Coolpix S2700.

Genomic DNA Extraction and Sequencing

Genomic DNA was extracted from the organism after starvation using a DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's protocol, and sequenced on the Illumina HiSeq2000 platform by Beijing Genomics Institute (BGI) Hong Kong.

Quality Control and Assembly

FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed April 25, 2015) was used for initial assessment of read quality. After initial assembly trials using ABySS (Simpson et al. 2009) and Velvet (Zerbino and Birney 2008) assembly software, SOAPdenovo2 (Luo et al. 2012) was empirically selected for further optimization. After initial assembly trials, Bowtie (Langmead and Salzberg 2012) was used to discern the true library fragment size (nominally 170 bp) which was found to average 183.05 bp, with

a standard deviation of 12.47 bp. The final “best” assembly presented here was then assembled using a *k*-mer size of 55, a pair_num_cutoff of 3, and a map_len value of 32. Deconseq 0.4.2 was used to remove potential contamination using custom human, arthropod, bacterial, and protist databases (using all available genomic sequence for the latter two clades, as downloaded from National Center for Biotechnology Information [NCBI] on September 16, 2014). Statistics for final assemblies were then determined using a perl script (available from the authors on request).

Gene Identification

For all analyses, the more complete (100 bp+) data set was utilized. To determine coverage of the core eukaryotic gene cassette, CEGMA (Parra et al. 2007) was run with all default settings. For identification of other genes, ncbi-blast-2.2.23+ (Altschul et al. 1990) tBLASTn searches were run using genes of known homology from the NCBI nr database as query sequences. The sequence of the contigs putatively identified was then reciprocally blasted (BLASTx) to the NCBI nr database for confirmation of identity.

Phylogenetic Analysis

Amino acid sequences identified in the *T. corallinus* genome were aligned with those of known homology downloaded from the NCBI nr database, using MAFFT version 7 (Kato and Standley 2013) under the L-INS-i strategy, and alignments saved in fasta format and imported into MEGA 6 (Tamura et al. 2013) for alignment curation, removal of gaps and phylogenetic analysis under the WAG (Whelan and Goldman)+Freqs model, four gamma categories, and all other default prior settings. Bayesian phylogenetic analysis was performed with MrBayes v3.2.1-x64 software (Huelsenbeck and Ronquist 2001) under the WAG model and all other default priors. Markov chain Monte Carlo searches were performed for the number of generations stated in figure legends, sampled every 100 generations, until the average standard deviation of split frequencies was less than 0.01, thus indicating convergence. The first 25% of samples were discarded as “burn-in” in all cases. Bayesian trees were displayed in Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>, last accessed April 25, 2015).

Mitochondrial DNA Analysis

Millipede mitochondrial genome sequences were downloaded from the nr database and blasted against our resource using BLASTn (Altschul et al. 1990) and all default settings. Sequences putatively identified as *T. corallinus* mitochondrial DNA (mtDNA) sequences were used to design primers using Primer3Plus (<http://primer3plus.com/>, last accessed April 25, 2015) and polymerase chain reactions (PCRs) on genomic DNA (extracted as above) were performed to confirm the sequence of the mitochondrial genome, with Sanger direct

sequencing performed by Techdragon Hong Kong and BGI. After the first round of PCR and assembly, additional primers were used to sequence and confirm the structure of the mtDNA, the sequence of which can be found in [supplementary file S2, Supplementary Material](#) online. DOGMA was used to demarcate protein coding and rRNA sequences (Wyman et al. 2004) with alignment confirmation performed with known arthropod mitochondrial sequences when necessary. tRNAscan-SE 1.21 (Lowe and Eddy 1997) was used to find tRNA gene sequences. The mitochondrial gene map was drawn using OGDRAW (Lohse et al. 2007). The nucleotide sequences of the 13 protein-coding genes found in all myriapod mtDNA sequences publically available as of October 28, 2014 ([supplementary file S2, table 2.2, Supplementary Material](#) online) were individually aligned in ClustalX (Larkin et al. 2007) and concatenated. jModelTest 2 (Darriba et al. 2012) was used to perform a nested likelihood ratio test to determine the best fitting model of sequence evolution, (GTR [general time reversible] + 4G + i), which was used in subsequent experimentation. Maximum-likelihood inference was performed using PhyML 3.0 (Guindon et al. 2010) with 1,000 bootstrap replicates and Bayesian inference using MrBayes 3.1 (MPI version, Ronquist and Huelsenbeck 2003) under the chosen substitution model with all other default priors. Chains converged after 1,000,000 generations, sampled every 100 iterations, with the average standard deviation of split frequencies less than 0.01. Trees were summarized with the first 25% of samples discarded as “burn-in” before display.

Results

Sequence Data, Assembly, and Contamination Removal

A. T. corallinus gDNA sample was sequenced on the Illumina HiSeq2000 platform by BGI Hong Kong. Reads were provided from an external server, and a summary of these data can be seen in table 1. The data have been uploaded to NCBI's Short Read Archive (Bioproject PRJNA260872, Biosample SAMN03048671, experiment SRX700727). Read quality as assayed using FastQC was found to be good (lower quartile Phred score greater than 31 through to the 100th base for both read data sets), as can be observed in [supplementary file S3, Supplementary Material](#) online. After initial trials of a range of de novo assembly software, SOAPdenovo2 (Luo et al. 2012) was selected for further optimization. Initial assemblies were used to find expected coverage and average sequenced fragment size, and a final optimized assembly with settings as described in the Materials and Methods section was performed, resulting in 1,305,238 contigs with a size of 100 bp or above, before the removal of contaminating sequence.

Initial analysis of our read data detected ribosomal gene sequence from non-Millipede species. Myriapods in general and *T. corallinus* populations in particular are often prone to infection by gregarine parasites, protists which inhabit the

Table 1

Summary Statistics Relating to Reads Used in Genome Assembly

Platform	Illumina HiSeq2000
Read length	100 bp
Average insert size (expected 170 bp)	183.05 bp
Insert size standard deviation	12.47 bp
Number of paired-end reads	75,820,817
Average GC%	38

Table 2

Summary Statistics, Final Genome Assembly

SOAPdenovo (<i>k</i> 55)	200 bp Minimum Contig Size	100 bp Minimum Contig Size
Number of contigs	348,141	1,233,936
Max contig length (bp)	39,922	39,922
Mean contig length (bp)	885.88	337.93
Median contig length (bp)	380	120
N50 contig length (bp)	1,829	955
No. of contigs in N50	41,048	82,105
No. of contigs >1 kb	78,806	78,806
No. of bases, total	308,411,367	416,979,918
No. of bases in contigs >1 kb	205,268,103	205,268,103
GC content %	41.74	41.13

intestinal tract of a range of arthropods and other invertebrates (Chang et al. 2004). We therefore assayed our genomic resource for evidence of these species, using known ribosomal RNA sequences. Although we did not recover any evidence of *Gregarinasina* infection (as seen in Chang et al. 2004), fragmentary ribosomal RNA sequences with BLAST identity to a number of ciliophoran protist species were recovered in our data set before contamination removal. To ensure our data were not more broadly contaminated with protist sequence, all protist and bacterial genomic data present in NCBI databases as of September 16, 2014 were used as the basis for comparison and removal using DeconSeq Standalone 0.4.3 (Schmieder and Edwards 2011). A total of 71,302 contigs were removed from our initial assembly (5.46%). Contigs removed will include shared repetitive elements, and as a result our data set should not be used for making conclusions about such regions of the genome. Statistics for the final assembly after the removal of contamination, given for contigs 200 and 100 bp in size and greater, are presented in table 2.

Using Jellyfish at a *k*-mer size of 31 and the “Estimate Genome Size” perl script (Ryan 2013; github.com/josephryan/, last accessed April 25, 2015), our estimate of genome sequence coverage for this species is 14.23 \times , with our estimated haploid genome size therefore being around 538 Mb. The genome size is two times larger than the *S. maritima* genome (290 Mb; Chipman et al. 2014) but is of average for myriapods (www.genomesize.com, last accessed April 25, 2015). No evidence could be gained from *k*-mer plots for large amounts of heterozygosity (as per Liu et al. 2013),

which could be due to the presence of a considerable proportion of unique k -mers in the data. Our genome assembly contains almost 900,000 contigs of size 100–200 bp. As such, we have made two versions of our assembly available, one containing all contigs, and one containing only those of size 200 bp and greater. These final assemblies have been uploaded to our lab website, and are directly available from tinyurl.com/millipede-genome100bp and tinyurl.com/millipede-genome200bp for ease of download. In total, 78,806 contigs greater than 1 kb in length are recovered, and this size, while at its lower limit too small for complete gene models (centipede *S. maritima* median total gene size: 3,170 bp), is sufficient to span most protein-coding domains.

CEGMA (Parra et al. 2007, 2009) results showing coverage of the expected metazoan gene complement are excellent—171/248 ultraconserved Core Eukaryotic Genes (CEGs) are recovered as present in our data (68.95%). From BLAST results (tBLASTn, $E < 10^{-9}$), 432/458 KOG groups (94.3%) were shown to be present as noted in [supplementary file S1, Supplementary Material](#) online, with 349/458 (76.2%) supported by all six species used by CEGMA. The missing 26 Eukaryotic Orthologous Groups (KOG)s groups are listed in [supplementary file S1, Supplementary Material](#) online. As the genome of *S. maritima* contained 95.1% (cf. *T. corallinus* 94.3%) of this data set, we suggest that we have recovered the majority of coding sequence in our assembly, albeit at low contiguity. Furthermore, taken together with the results of our targeted investigations into specific gene families, detailed below, we find no evidence for large-scale gene loss in the *T. corallinus* genome.

Comparison with the *S. maritima* Genome

The recently published *S. maritima* genome data set was distinguished by its high levels of conservation across the majority of characterized gene families (Chipman et al. 2014). To further test whether this “prototypical arthropod cassette” was also found in the millipede genome, we compared the complete annotated *S. maritima* proteome to the *T. corallinus* genome. Of the 26,950 peptides in the *Drosophila melanogaster* genome (BDGP5.23 release), 15,995 (59.4%) possess a putative ortholog in the millipede genome (tBLASTn, $E < 10^{-9}$). This is comparable to the number of *S. maritima* genes possessing a hit in the *D. melanogaster* proteome. Of 15,008 *S. maritima* peptides (Release 23, downloaded from Ensembl, September 18, 2014), 9,168 (61.1%) possess a putative ortholog in *D. melanogaster* (BLASTp, $E < 10^{-9}$).

Of 15,008 proteins in the *S. maritima* genome, 8,459 (56.4%) possessed a putative ortholog in our data set (tBLASTn, $E < 10^{-9}$). Surprisingly, this was fractionally lower than *D. melanogaster* complement recovery. Previous estimates have suggested that 32% of genes present in the *Strigamia* genome are the result of gene duplication events unique to the Myriapoda lineage (Chipman et al. 2014,

phylomeDB analysis), but the low percentage of centipede proteins recovered in our genome may mean that these genes are largely *Strigamia*- or centipede-specific, rather than shared across the Myriapoda.

Widespread Dscam Paralogy—A Myriapod Trait?

It was noted in investigations of the *S. maritima* genome that duplication/paralogs were used to generate coding sequence diversity, where insects use alternate splicing (Chipman et al. 2014). To investigate whether this trait is observed more broadly in the Myriapoda, we have assayed cases noted from that species in the *T. corallinus* genome. The most pronounced example of duplication observed in the centipede was the Down Syndrome Cell Adhesion Molecule (Dscam) gene family, where over 100 unique loci were noted (Chipman et al. 2014). Using BLAST to compare *S. maritima* Dscam homologs to our data set (tBLASTn, cutoff 10^{-9} ; [supplementary file S1, Supplementary Material](#) online), we recover 43 contigs containing the *Dscam* Ig7 domain.

We therefore suggest that myriapods share duplication at some loci, rather than utilizing splice variation to generate gene-level diversity, as suggested in Chipman et al. (2014). As duplication rather than splice variation has also been recently noted in the chelicerate *Ixodes scapularis* (Brites et al. 2013), it is likely that the alternative splicing of *Dscam*, and perhaps other genes, has evolved independently in the Pancrustacea (Chipman et al. 2014).

However, care should be taken before assuming duplicates to be present ancestrally on the basis of duplicate loci in *S. maritima*, as duplications observed in that species are not necessarily shared across all myriapods. For instance, a single copy of gene *cap'n'collar* (*cnc*) has also been noted as exhibiting paralogy rather than splice variation in *S. maritima* (nucleotide and amino acid sequence; [supplementary file S1, Supplementary Material](#) online), but only a single example is found in *T. corallinus*. Although proving absence is difficult, the clear homology of this gene to its orthologs and our generally high recovery of coding regions of the genome gives us confidence that a second paralog would be spotted if present. The duplication of *cnc* in *S. maritima* therefore seems to be specific to that species, rather than a myriapod symplesiomorphy.

Trigoniulus corallinus Circadian Clock-Driving Proteins: A Less Derived Myriapod?

In some ways the terrestrial lifestyle of *T. corallinus* suggests that it may possess more ancestrally shared characters than *S. maritima*, given the subterranean, marine lifestyle of the latter species. This is borne out by investigation of known gene families, whose absence in *S. maritima* could not previously be discerned to have resulted from selective pressure in the geophilomorph centipedes or ancestral loss. Perhaps unsurprisingly, given the presence of classical ocelli in this species, canonical opsin genes are found in this genome. Classical

circadian clock-driving proteins, such as period (*per*), Clock (*Clk*) and cycle (*cyc*), are all found in the *T. corallinus* genome (supplementary file S1, Supplementary Material online), further reinforcing the lineage-specific nature of loss in the *S. maritima* resource (Chipman et al. 2014). The presence of such genes robustly underlines the utility of a diploid genomic resource.

Mitochondrial Genome Analysis

Myriapods have been previously noted as displaying diversity in mtDNA sequence and gene structure (e.g., Gai et al. 2008; Lavrov et al. 2002), and the sequencing of the *T. corallinus* genome provided an ideal opportunity to add to our presently

poor sampling of myriapod mtDNA sequences—as of October 28, 2014, 14 myriapod mtDNA sequences were available in public sources, a fraction of the 13,000 extant species of this clade.

The *T. corallinus* mitochondrial genome is a 14,907 bp, circular molecule, possessing the typical metazoan 13 protein-coding genes, 22 transfer RNA genes, and 2 ribosomal RNA genes (fig. 2). These are distributed with 22 genes on the majority-strand (α), and 15 on the minority-strand (β). Full details, along with start/stop codon and positional information can be found in supplementary file S2, Supplementary Material online, along with full mtDNA sequence and primers used. Of the 13 protein-coding genes, the typical metazoan start codon, “ATN,” is used by 12, whereas *cox1* employs

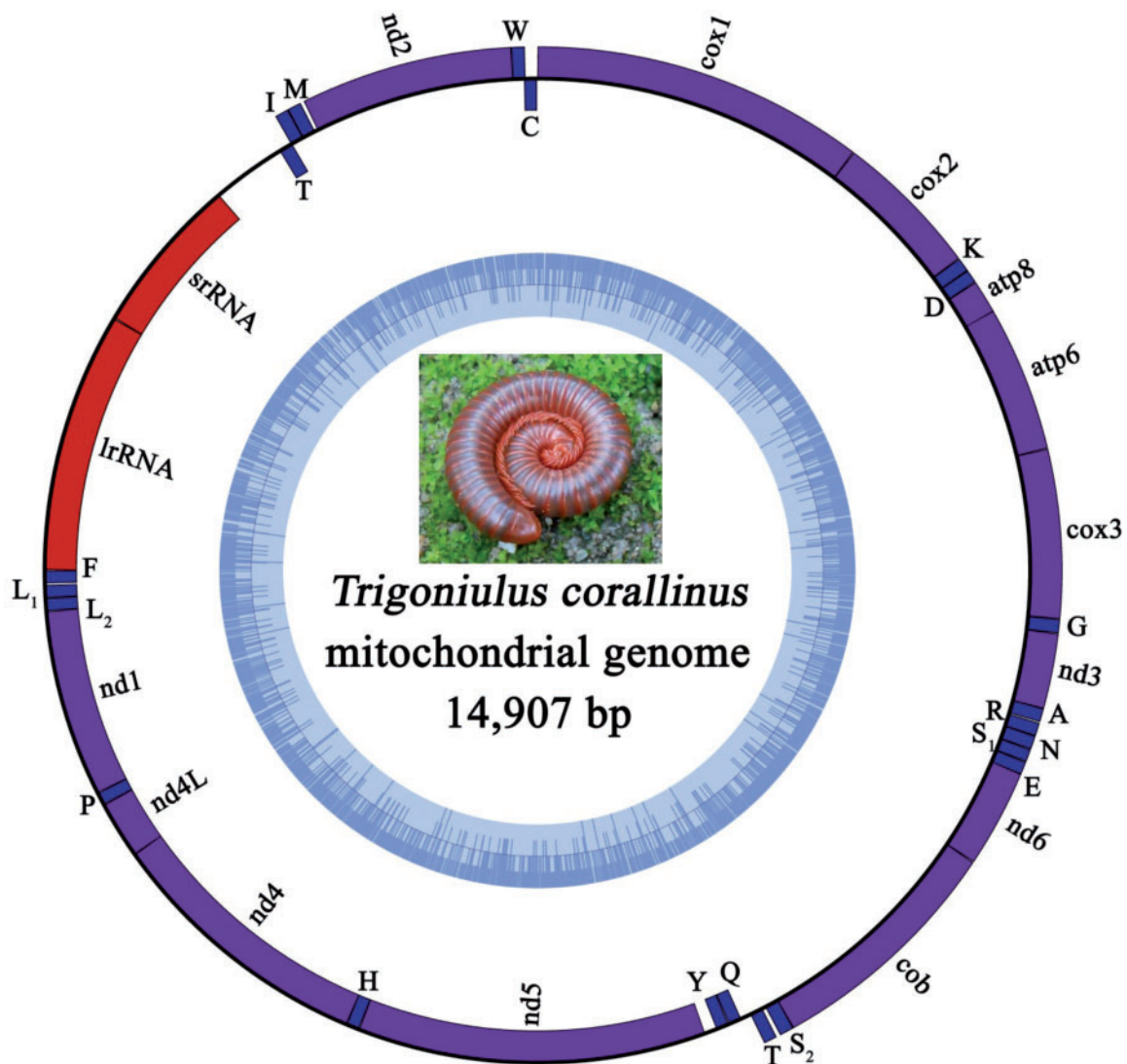


FIG. 2.—The organization of the *T. corallinus* (Myriapoda: Trigoniulidae) mitochondrial genome. Orientation of genes (transcription clockwise or anticlockwise represented outside or inside the form, respectively) is represented by the outside circle. Local GC content, (GC dark blue, AT light blue) represented on the inner ring. Image displayed in OrganellarGenomeDRAW (Lohse et al. 2007). Photograph by the authors.

“ACG.” Six genes, however, use an atypical incomplete stop codon. Twelve genes overlap in coding sequence with another, and a total of 510 noncoding base pairs were noted, with 360 of these between the *srRNA* and *trnT* genes. As shown in [supplementary file S2, table 2.2, Supplementary Material](#) online, the AT% of the mitochondrial genome is high, at 72%, which is not unusual for myriapods as a whole but is quite markedly different to the mtDNA sequence of the closest sister taxon to *T. corallinus* yet available, that of *Narceus annularis*. The gene arrangements of these two species are only slightly different (fig. 3), with the location of the *trnF* gene altered. The *trnV* gene seen in other myriapods is also absent, and in its place is a second *trnT* gene, which must have marked effects on the mitochondrial biology of *T. corallinus*. It is possible that this second *trnT* gene came about as a result of an apparent tandem duplication in the Spirobolid lineage (Lavrov et al. 2002), followed by differential patterns of loss or mutation. This species' mtDNA gene arrangement generally closely resembles the unusual arrangement seen in its sister taxa and that of the Spirostreptid *Thyropygus* sp. DVL-2001, rather than mirroring the arthropod ground plan, as the Lithobiomorph centipedes *Lithobius forficatus* and *Bothroplys* sp. SP-2004 do. The duplication event proposed in Lavrov et al. (2002) therefore seems generally conserved in Spirobolid millipedes, and likely predates the divergence of the Spirobolid and Spirostreptid lineages.

We utilized the concatenated alignments of nucleotide sequences from protein-coding genes to attempt to reconstruct myriapod phylogeny (fig. 4), and can firmly assign *T. corallinus* and *N. annularis* as sister groups with maximal support. Most local interrelationships are recovered as expected given previous knowledge (e.g., Brewer et al. 2013) and with firm bootstrap support. However, the location of the Symphyla in particular in our analysis, as sister taxa to all other Myriapoda, runs counter to many lines of evidence, which generally place Symphyla as seen in figure 1A. We note, however, that some recent evidence (Rehm et al. 2014) using a range of models supports this topology. The placement of the Pauropoda within the Diplopoda is also interesting, given the extensive character loss which must be inferred if this is correct. Difficulties in resolving deep phylogenetic relationships with mtDNA data, both generally in the Arthropoda (Cameron et al. 2004) and within the Myriapoda in particular (Brewer et al. 2013; Rehm et al. 2014) have been noted previously, and we suggest that deeper sampling in the Myriapoda is a necessity if mtDNA is to be used as a character for phylogenetic inference with confidence, given the high rates of change and reorganization seen in this clade.

Developmental Gene Cassette

Insects have been the workhorses of genetics and developmental biology since that field began. We therefore know far more about how genes control development in insects than in

any other clade in the Arthropoda. With the publication of data from noninsect arthropods, our knowledge of the underpinnings of many developmental processes is maturing greatly (e.g., Sin et al. 2014). Information from the *T. corallinus* genome can also help this process. The sequences of many other key developmental genes and transcription factors are also found in the *T. corallinus* genomic data set. High levels of recovery of several well-categorized gene families allow us to both confirm deep sequence recovery in our data set and clarify previously opaque areas of diplopod and arthropod molecular evolution. Table 3 shows our recovery of several classes of transcription factors.

Perhaps, the most well-cataloged developmental transcription factor genes are the ANTP-class homeobox genes including HoxL genes which perform a variety of key roles in the organization of development (Hui et al. 2012). Our recovery of these genes was excellent, with 12 genes found of the approximately 15 (Hui et al. 2012) that would be expected to be present (fig. 5A, table 3). The three missing genes, *Pb/Hox 2*, *Eve/Evx*, and *Unpg/Gbx*, are all ancestrally present in the Urbilateria (Hui et al. 2012). It is difficult to confirm true absence from an incomplete genome, but despite attempts to find gene sequences, both inside and outside the homeodomain, of these missing genes using gene sequences from *S. maritima* no evidence of presence could be garnered. A potential ParaHox gene *Xlox* homolog was found in *S. maritima* (Chipman et al. 2014) with some similarity to *Zen/Hox3*. This would represent the first ecdysozoan *Xlox* sequence if orthology could be confirmed. However, in *T. corallinus* only a clear *Zen/Hox3* gene was present and no *Xlox* homolog was observed. Given that ParaHox genes are known to exist in placozoans, cnidarians, and the last common ancestor of bilaterians (Hui et al. 2008, 2009; Mendivil Ramos et al. 2012), the loss of *Xlox* is indeed widespread in the Arthropoda.

Other than the HoxL class, our data set also offers interesting insights. As in the centipede genome, a clear *DMBX* ortholog is identified in the genome of *T. corallinus*, whose sequence can be found in [supplementary file S1, Supplementary Material](#) online. Its presence in myriapods and thus representatives of all bilaterian superphyla confirms further the presence of this gene at the base of the bilaterian radiation, and, along with its presence in annelids (Takahashi and Holland 2004; Kenny and Shimeld 2012) suggests a means by which the origin of its role in demarcating the mid-brain/hindbrain boundary may be tested.

The Forkhead Box (Fox) gene class of transcription factors is responsible for mediating a broad range of cellular activity. Known for its “winged helix” forkhead motif, derived from the helix-turn-helix class of genes to which they belong, these genes are easily recognizable and play key roles in metazoan growth and development (Shimeld, Degnan, et al. 2010; Shimeld, Boyle, et al. 2010). However, how they have evolved is often difficult to discern. The *FoxJ1*, *FoxJ2/3*, and *FoxL1*

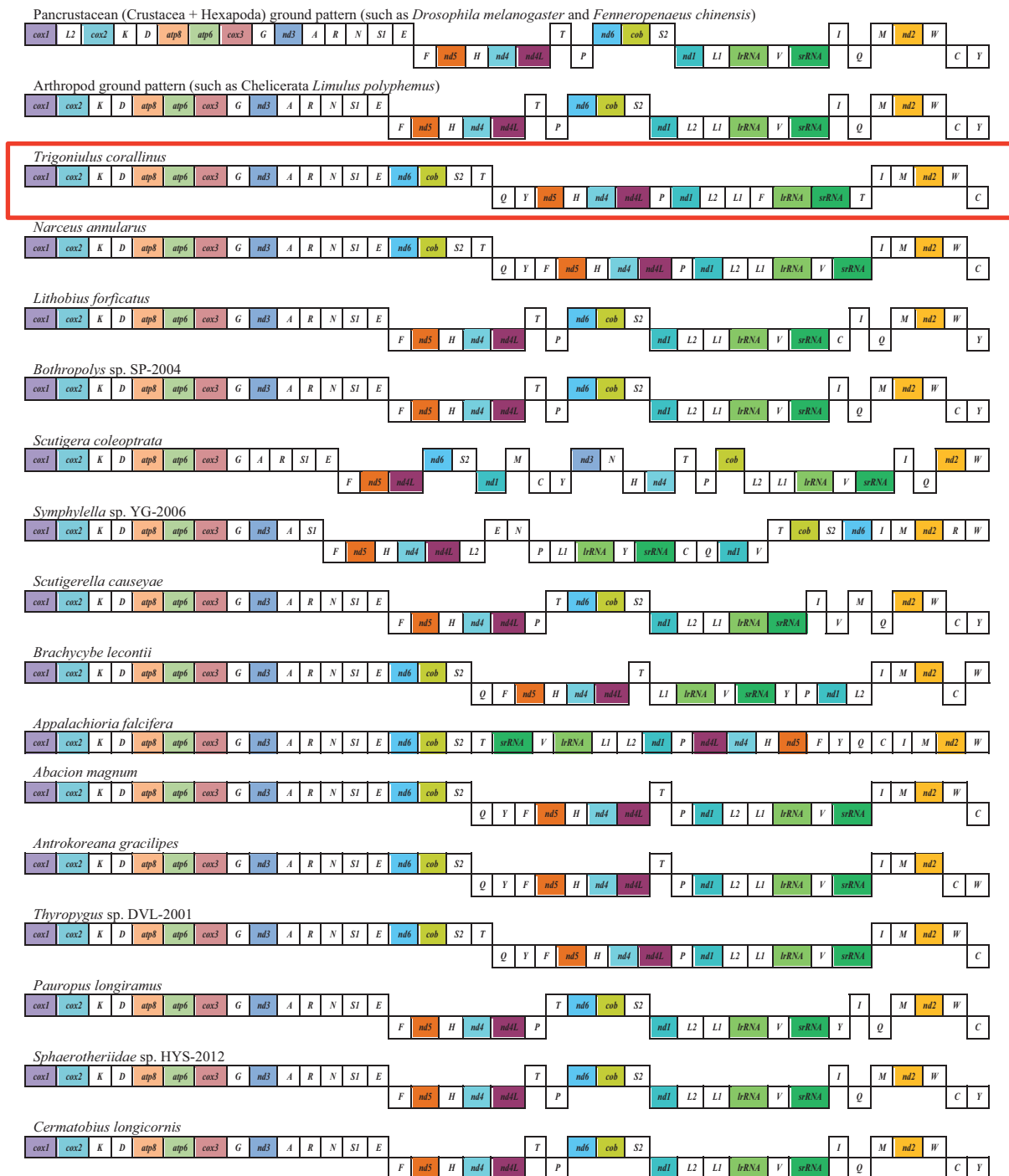


FIG. 3.—Mitochondrial genome gene order across the Myriapoda, as compared with Panarthropod and Hexapod ground patterns, with that of *T. corallinus* boxed in red. Note similarities to *N. annularis*, also a member of the Spirobollidae. Genes colored for ease of recognition.

subfamilies of Forkhead box genes, for instance, are absent from insects (Shimeld, Degnan, et al. 2010; Shimeld, Boyle, et al. 2010). In the *T. corallinus* genome we recover almost every subfamily of Fox genes we would expect to find in this species (table 3), confirming further the sequencing depth of

our genome. The identity of these was proven by phylogenetic inference, as can be seen in figure 5B, with robust bootstrap support providing evidence of orthology. This includes the *FoxJ2/3* and *FoxL1* genes, which have also been identified in *S. maritima* (Chipman et al. 2014), confirming that losses in

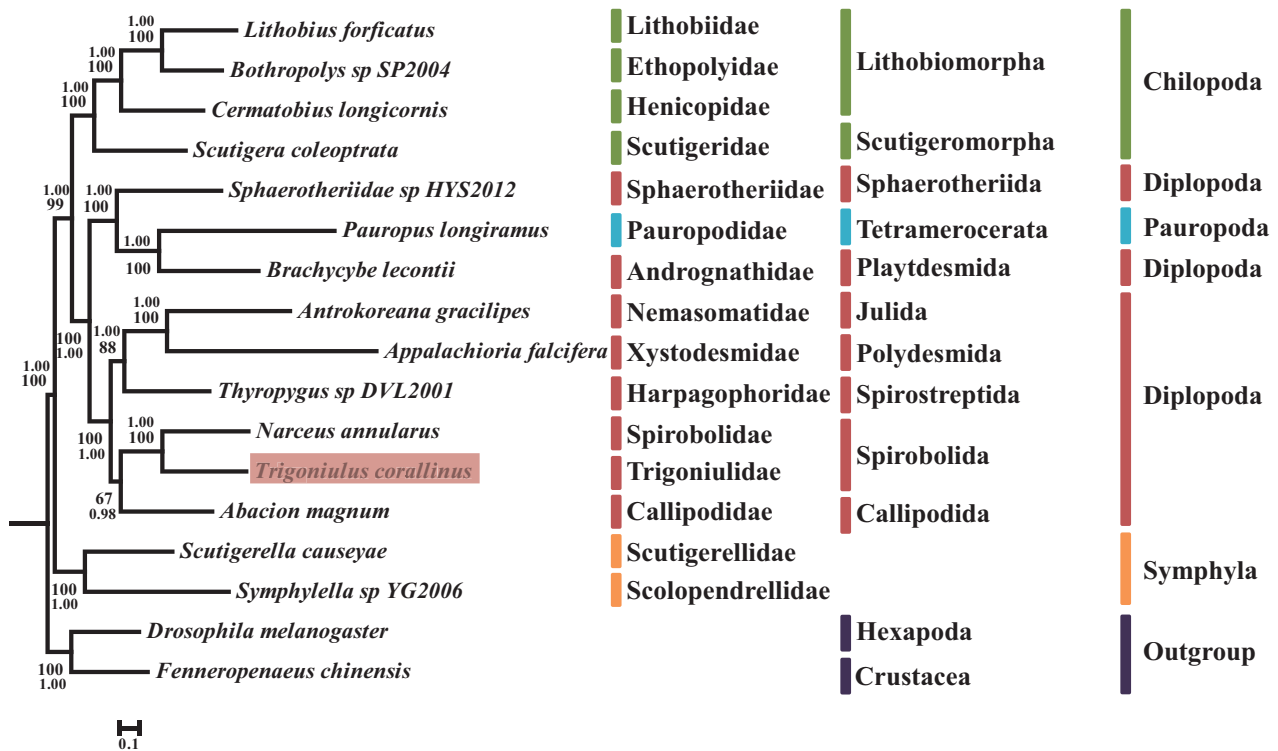


FIG. 4.—Myriapod mtDNA interrelationships: Tree recovered by Bayesian and maximum-likelihood inference of Myriapod and Panarthropod interrelationships, on the basis of concatenated coding nucleotide sequences of 13 protein-coding genes, performed as described in Materials and Methods. *Trigoniulus corallinus* boxed in red for ease of identification. Familial, Order, and Class level classifications shown at right. Note: Symphalans are positioned as sister taxa to all other Myriapods, in marked contrast to established phylogenies. Bootstrap proportions (as percentage, 1,000 replicates) and Bayesian posterior probabilities (maximum 1.00) shown at base of nodes. Scale bar represents sequence changes per site at unit distance.

Table 3

Recovery of Developmentally Important Transcription Factor Families, and Notable Absences in Our Data Set

Gene Classes	Homologs Recovered	Unexpected Absences
Homeobox genes (HoxL)	12	<i>Evx</i> , <i>Gbx</i> (<i>Unpg</i>), <i>Hox2</i> (<i>Pb</i>)
Fox genes	16	<i>Fox J1</i>
Sox genes	5	<i>Sox B1</i>
T-box genes	5	<i>Tbx 15/18/22</i>

insects are specific to that clade. The only absentee from our data set is *FoxJ1*, which is found in *S. maritima*. We note one orphan Fox gene, noted on our trees as “Fox Unknown Homology,” with weak resemblance to Fox K and Fox C protein sequences under BLAST-based comparison and a clear forkhead domain structure which may represent a highly derived form of *FoxJ1*. If this gene is indeed absent from our data set, its frequent loss may require further investigation across the Arthropoda, to determine why its role as a key regulator of ciliogenesis in other clades is not retained (Yu et al. 2008).

The SOX family of transcription factors was also targeted for detailed investigation. These HMG class genes play diverse roles in sex determination, growth, and development

(Koopman et al. 2004). The phylogenetic relationships of these genes with those of known orthology can be seen in figure 6A, and an interesting contrast can be seen in the diversity of *T. corallinus* and *S. maritima* SOX family members. Although in *S. maritima* these genes seem to have undergone duplication in the Sox B2 lineage (Chipman et al. 2014), *T. corallinus* possesses a far more typical protostome cassette in this regard and has retained a *SoxF* gene, confirming that this loss is specific to the centipede’s lineage rather than shared myriapod-wide. The only exception to the strong orthology with well-categorized SOX clades comes in the form of a divergent HMG-box like sequence, found in both *S. maritima* and *T. corallinus*. This sequence diverges greatly from known SOX clades, and is therefore pulled toward to base of the tree by long branch attraction. The sequence for the *T. corallinus* ortholog of this sequence is given in [supplementary file S1, Supplementary Material](#) online.

The T-box family was also examined in detail, and similarly to other families showed almost complete recovery (table 3, fig. 6B). Of the families reasonably expected to be found in the Myriapoda, only *Tbx 15/18/22* was absent from our data set. The presence of two *Midline/H15* genes in both *S. maritima* and *T. corallinus* suggests that these genes could have

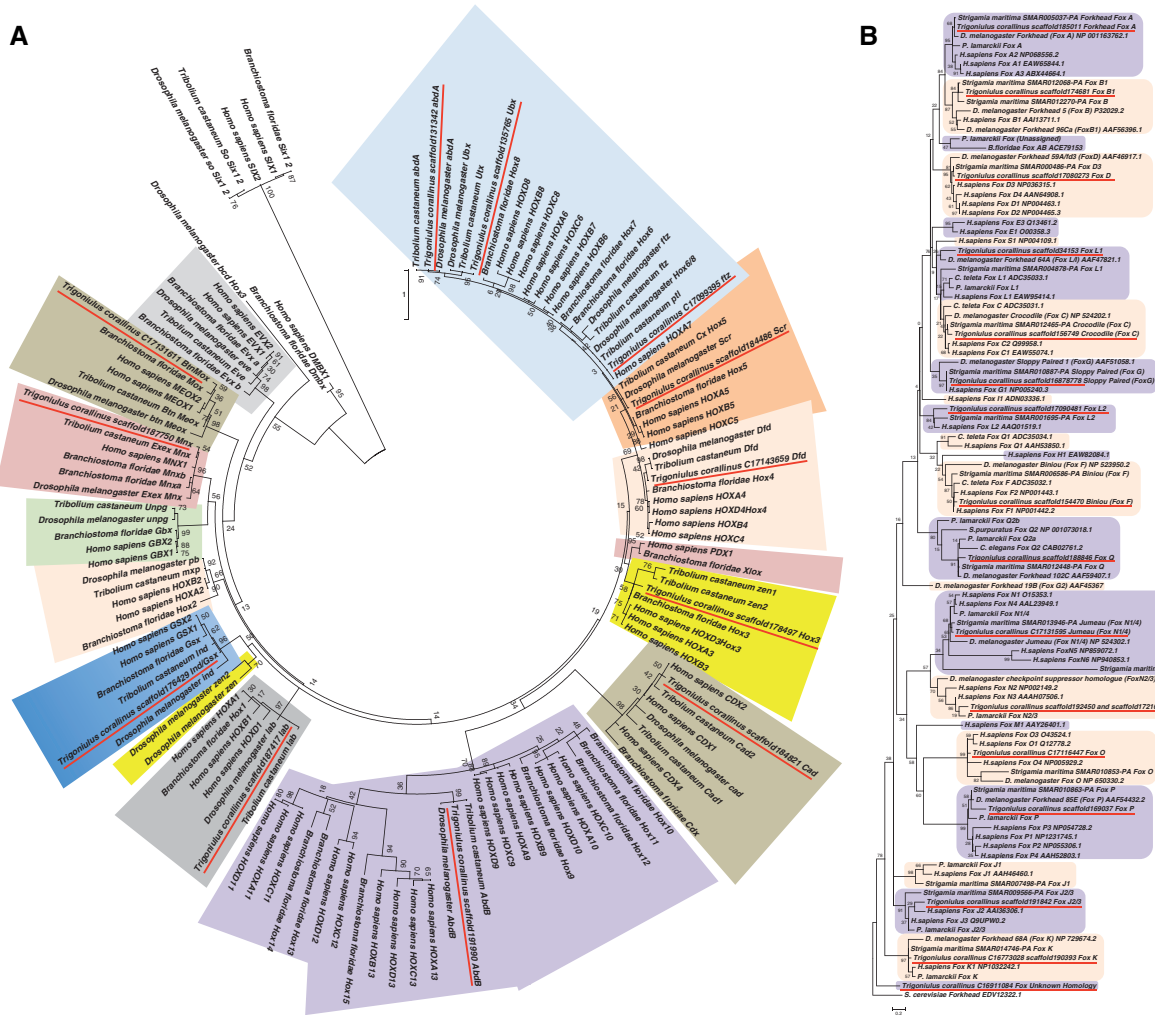


Fig. 5.—(A) ANTP-class *HoxL* and *Fox* family gene interrelationships as inferred by maximum-likelihood reconstruction. Sequences include *T. corallinus* (underlined in red) amino acid sequences, and those of known orthology downloaded from HomeoDB (Zhong et al. 2008; Zhong and Holland 2011). After removal of all columns containing one or more gaps, a final, 57 amino acid alignment spanning the homeodomain was used as the basis of phylogenetic inference. Bayesian phylogeny using the same sequences can be found in [supplementary file S4, Supplementary Material](#) online. Numbers at base of nodes represent bootstrap percentages (from 1,000 replicates). Scale bar at top center represents substitutions per site at given unit distance. Colored boxes denote individual *HoxL* gene families. Note: *Drosophila melanogaster Zen* and *Bicoid* genes differ in sequence markedly from similar genes, and fall outside their known orthologs. Tree rooted using *Six* and *DMBX* gene amino acid sequences. All sequences and alignment used can be found in [supplementary file S1, Supplementary Material](#) online. (B) Reconstruction of interrelationships of the *Fox* genes of *T. corallinus* (underlined, red) with those of known homology from other species, inferred by maximum-likelihood reconstruction. Sequences downloaded from NCBI's nr database have accession numbers as noted in next to taxa names, whereas others were sourced from the Ensembl Metazoa web resource (*S. maritima*), or taken from previously published sources (*P. lamarcki*; Kenny and Shimeld 2012). After removal of all columns containing one or more gaps, a final, 68 amino acid alignment spanning the Forkhead domain was used to infer phylogeny. Bayesian phylogeny using the same sequences can be found in [supplementary file S4, Supplementary Material](#) online. Numbers at base of nodes represent bootstrap percentages (from 1,000 replicates). Scale bar represents substitutions per site at given unit distance. Tree rooted using *Saccharomyces cerevisiae* Forkhead (EDV12322.1) amino acid sequence. All sequences and the alignment used can be found in [supplementary file S1, Supplementary Material](#) online.

duplicated early in the arthropod radiation and undergone extensive gene conversion in the interim, but the lack of evidence of paralogy could also mean that these loci are prone to independent duplication, for reasons unknown. *Trigoniulus*

corallinus retains a *Tbx 4/5* gene which is absent from the *S. maritima* data set, and only possesses a single *Brachyury* homolog, compared with the two found in the centipede (Chipman et al. 2014).

The recovery of developmentally important transcription factor genes in our data set is near-total. We show evidence of presence of in excess of 80% of the expected transcription factor complements of a variety of key families, a considerable advance on presently extant data sets, and a figure which compares positively with that found in other genomes. These data will be vital for inferring the ancestral complements and functions of a variety of key genes across the Arthropoda.

Chemosensory Genes and Receptors

Given the lifestyle of diplopods, the ability to sense and respond to a variety of environmental cues, and particularly the ability to assay the chemical composition of their habitat, is key to allowing them to find food and one another while avoiding predators and the worst environmental effects. The colonization of land by insects and myriapods occurred independently (Regier et al. 2010), and thus it is interesting to consider whether chemosensory genes evolved before or after their diversification.

In *S. maritima*, no representatives of the Odorant Binding Protein (Pelosi 1994) or CheA/B families (Starostina et al. 2009) could be identified, leading to the conclusion that these genes are possible insect novelties (Chipman et al. 2014). This is corroborated by our data set, where we were unable to note even putative homology to any sequence in our genome when we searched our data set using a number of genes from these families (tBLASTn, *E* cutoff 1). The Chemosensory Protein (CSP) family (Pelosi et al. 2006) possesses two orthologs in our data set, as seen in *S. maritima*. These differ markedly in amino acid sequence from the centipede orthologs, however, which may reflect the differences in environments inhabited by these species.

Again similarly to observations in *S. maritima*, no Odorant Receptor (OR) genes could be identified in our data set using a variety of sequences of known orthology at lenient blast settings (tBLASTn, *E* cutoff 1). This is consistent with the findings of Robertson et al. (2003), who suggested the OR family represents an insect novelty. Surprisingly we were unable to identify many Gustatory Receptor (GR) genes in our data set, finding only two sequences with clear homology to this family (sequences, [supplementary file S1, Supplementary Material](#) online), a marked contrast with *S. maritima*, which possesses 77 of these genes. GR genes are known to have existed in the arthropod common ancestor, and have been observed in arachnids and crustaceans as well as in the centipede, but the marked diversification seen in *S. maritima* seems limited to that species. A relatively large number of Ionotropic Receptor (IR) genes can be found in the *T. corallinus* genome, although not as many as the 69 observed in *S. maritima*. A total of 23 IR sequences (listed in [supplementary file S1, Supplementary Material](#) online) were found in our data set, of all IR classes. This restricted complement relative to the

centipede may reflect the use of duplication to build diversity in *S. maritima*, as noted earlier in this article.

Our data set therefore corroborates the hypothesis that the Odorant Binding Protein, CheA/B, and OR gene families are insect novelties. The CSP, GR, and IR families are present in the Rusty Millipede, but differ greatly in sequence identity and complement number to that found in the centipede. This likely reflects the differences in environment and behavior exhibited by these species, but may also be a consequence of an increased tendency to duplicate genes for the creation of genetic diversity in *S. maritima*. Increased sequencing efforts in the Myriapoda will allow testing of these hypotheses further.

Myriapods as Novel Chemical Sources

Myriapods have been noted previously as potential sources of novel bioactive chemical compounds, with centipedes attracting attention as their venoms may be of interest to medicine (Undheim et al. 2014). Millipedes are known to exude a cocktail of defensive chemicals to deter predators, with hydrogen cyanide, benzaldehyde, and quinone derivatives playing a key role (Blum 1981). These could be of interest to pharmaceutical companies in a range of contexts, as even capuchin monkeys have been observed using millipedes as dispensers of insect repellent (Weldon et al. 2003). The investigation of genomic pathways resulting in the production of such compounds in the Diplopoda is, however, still very much in its infancy, and this genome therefore represents a potent source of information on these pathways. We used three families of quinone biosynthesis pathway genes recently investigated and functionally confirmed in the beetle *Tribolium castaneum* (Li et al. 2013) to provide a basis for understanding the diversity of these pathways in *T. corallinus*. Using the *T. castaneum* *quinone-less vitellogenin-like (VTGL)*, *quinone-less arylsulfatase b (ARSB)*, and *quinone-less multidrug resistance protein (MRP)* gene sequences to search our data set, we identified possible homologs in our genome data set (tBLASTn, *E* cutoff 10^{-9} , *T. castaneum* numbers given in brackets): 3 (14) *VTGL*-like sequences, 20 (4) *ARSB* sequences, and 121 (17) *MRP* sequences were putatively identified in our millipede genome, all of which can be found in [supplementary file S1, Supplementary Material](#) online. This large diversity of *ARSB*- and *MRP*-like genes, while at present of completely unknown functional role, means that millipedes have a large cassette of possible genes to deploy in producing quinone derivatives and other chemicals. Such diversity reflects the known novelty of their biochemical cassette and provides fertile ground for future investigations into novel bioactive agents. We observe that *VTGL* diversity is however markedly low compared with *T. castaneum*, which should be noted in further investigations.

Myriapods also represent possible sources of antifungal and antibacterial agents. These can be recognized by similarity to known genes in key domains. For instance, [fig. 7](#) shows the alignment of two novel mycin (antifungal) genes found in the

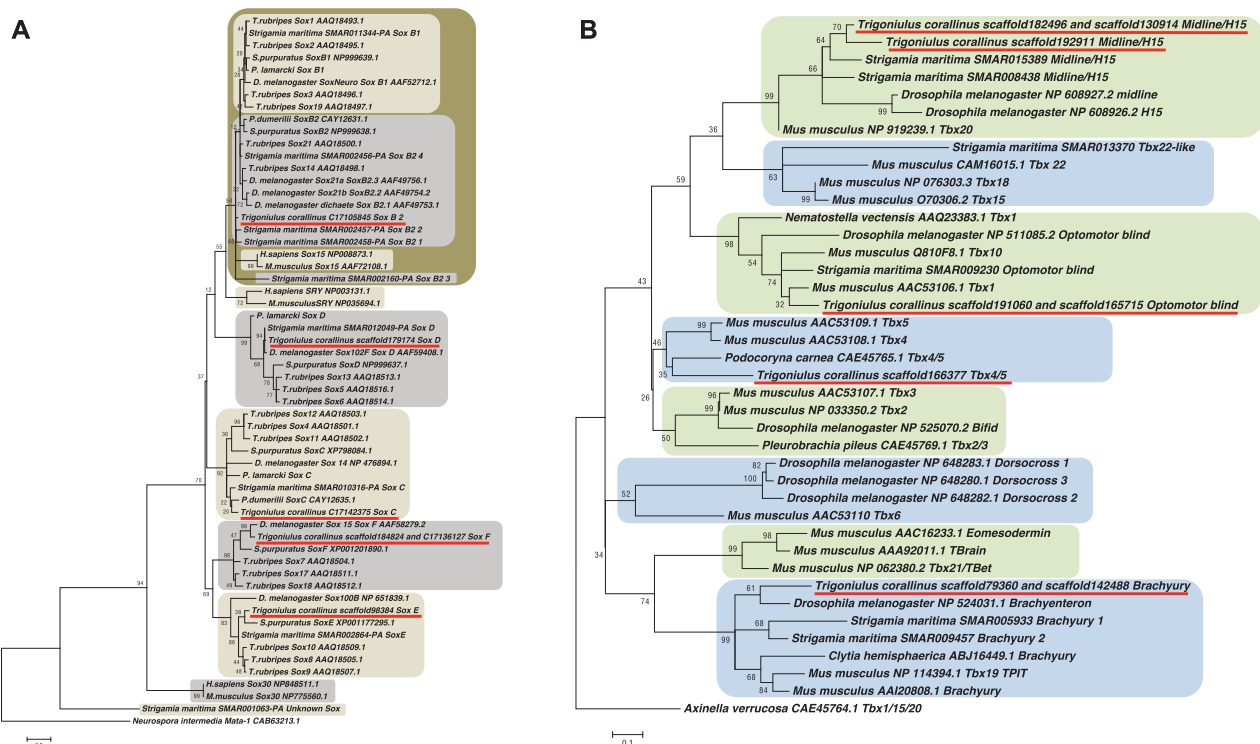


Fig. 6.—Phylogeny of Sox and T-box class genes: (A) Phylogeny of the Sox genes of *T. corallinus* (underlined, red) as inferred by maximum-likelihood reconstruction in MEGA. Sequences of known homology were downloaded from NCBI’s nr database with accession numbers given on taxa labels or downloaded from the Ensembl Metazoa web resource (*S. maritima*). All columns containing one or more gaps were removed and the resulting 80 amino acid alignment spanning the HMG domain was used to reconstruct phylogeny. Bayesian phylogeny reconstructed with the same alignment can be found in the [supplementary file S4, Supplementary Material](#) online. Bootstrap percentages (from 1,000 replicates) are given at base of nodes. Substitutions per unit distance given by scale bar at base of figure. Colored boxes indicate generally inferred Sox gene clades, with the enigmatic interrelationships of SoxB genes noted. *Neurospora intermedia* Mata-1 (CAB63213.1) amino acid sequence has been used to root the tree. All sequences and the alignment used can be found in [supplementary file S1, Supplementary Material](#) online. (B) Phylogeny of the T-box genes of *T. corallinus* (underlined, red) as inferred by maximum-likelihood reconstruction in MEGA. Sequences were downloaded from NCBI’s nr database with accession numbers as noted in figure or sourced from the Ensembl Metazoa web resource (*S. maritima*). After removal of all columns containing one or more gaps, a final, 88 amino acid alignment spanning the T-box domain was used to infer phylogeny. A Bayesian phylogeny using this alignment is presented in [supplementary file S4, Supplementary Material](#) online. Numbers at base of nodes represent bootstrap percentages (from 1,000 replicates). Scale bar represents substitutions per site at given unit distance. Colored boxes demarcate commonly inferred T-box clades. Tree rooted using *Axinella verrucosa* CAE45764.1 Tbx1/15/20 amino acid sequence. All sequences and alignments used can be found in [supplementary file S1, Supplementary Material](#) online.

millipede genome to those known from *D. melanogaster* and *Caenorhabditis remanei*. As with the known mycin sequences, these novel proteins possess clear similarity to the gamma-thionin (PF00304) antifungal domains found in plants, but the differences in sequence outside the core domain mean that they may have novel characteristics. These and other novel bioagents may be of use and interest to a range of biochemical industries, given the often compromising niche inhabited by the millipede.

Cuticular Protein Diversity

The exoskeleton of arthropods is made up of the polysaccharide chitin, along with a range of cuticular proteins. Many

kinds of cuticular protein are known, but the CPR family is the most common, with up to 150 genes found in some species of arthropod (Willis 2010). These proteins can be recognized by a conserved, approximately 64 amino acid sequence (Rebers and Willis 2001). Searches using sequences of known orthology of our data set identified 26 instances of this sequence in the *T. corallinus* genome. This is slightly fewer than the 39 members identified in the *S. maritima* genome (Chipman et al. 2014), and as in that data set, both RR-1 (flexible cuticle) and RR-2 (rigid cuticle) associated forms of these genes can be identified, entirely consistent with the mandibulate origin of the RR-1 family as proposed in Chipman et al. (2014). The nucleotide and amino acid

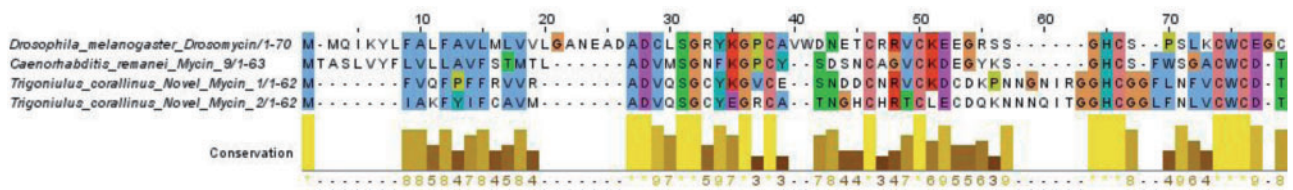


Fig. 7.—Novel mycin sequences: Alignment of known mycin (antifungal) genes from *D. melanogaster*, *Caenorhabditis remanei*, and novel mycins from *T. corallinus*. Alignment visualized in Jalview, colored with ClustalX identity. Note regions of high conservation around shared cysteine residues, with more divergence elsewhere, particularly at the N terminus of these sequences.

sequences of these putatively identified domains can be found in [supplementary file S1, Supplementary Material](#) online.

Discussion

The resource detailed here represents a vital data set for beginning investigation into the evolution of a variety of traits in the Diplopoda, Myriapoda, and the Arthropoda more generally. It is a draft genomic data set with deep coverage of the coding complement and high recovery of the total expected genome size. With a broader sampling of genomic diversity, our ability to infer the true origin of genes and phenotypes will allow us greater biological insight, and allow firm conclusions to be drawn as to the reasons for the success and diversity of the arthropod lineage. This data set also contains a variety of intriguing findings. Millipedes were the first arthropods to emerge onto land, and possess a variety of unique adaptations which have contributed to their success over their 400-Myr old history. This genomic resource will allow us to investigate the diversity of millipede novelties, such as their defensive chemicals and fungicides, with much more vigor than previous studies have been able to accomplish.

We can also be more confident of assertions regarding myriapod genomics with the advent of this resource. For example, the extensive gene duplication seen in *S. maritima* is far less prevalent in this resource. Although some duplication as a means to build genomic diversity seems to occur in this clade, it is of much more limited scope, and the scale of gene duplication seen in the centipede should therefore not necessarily be expected in other members of the Myriapoda. Our recovery of transcriptomic cassettes also allows a different angle to examine origin and loss of traits in arthropods from a molecular perspective—we can confirm, for instance, the presence of *DMBX*, *FoxJ2/3*, and *FoxL1* in the arthropod common ancestor, and show firm evidence for the existence of *Tbx 4/5* and *SoxF* in this lineage.

The advent of myriapod data sets will also allow a comprehensive revisiting of how extant gene complements were co-opted into the formation of new organs, particularly oxygen exchange systems, after terrestrialization in insects and myriapods (e.g., Grillo et al. 2014; Sánchez-Higueras et al. 2014). Although genes involved in the development of respiratory

organs, such as *apterous* (e.g., Damen et al. 2002, sequence; [supplementary file S1, Supplementary Material](#) online), are present in our data set, evidence remains to be gathered on any potential role of these genes in respiration in the Myriapoda. This is because terrestrialization occurred independently in insects and myriapods, and we are cautious about inferring functional homology between genes in these species without expression data. Further research using this genome as a resource will however offer a unique insight into how terrestrialization was accomplished independently in these speciose and successful metazoan clades.

The utility of myriapod mitochondrial genomes for the reconstruction of arthropod phylogenetics has been the subject of some recent debate, due to the extensive rearrangement seen in the myriapod taxa (Brewer et al. 2013). The *T. corallinus* mitochondrial genome appears similar in general arrangement to that of other related millipedes, and particularly that of the only other Spirobolid mitochondrial genome yet sequenced, that of *N. annularis*. By other metrics however, such as AT%, the *T. corallinus* genome is quite different to even that of its closest sequenced relative. These differences have obfuscated phylogenetic inference using mtDNA sequences in the Myriapoda in the past (e.g., Brewer et al. 2013) and this was also found to an extent in our data, with Symphalans posited as the sister group of all other myriapods, a likely artifact, perhaps caused by high rates of change in this clade. Further sampling of myriapod mtDNA is necessary to unravel the complex and intriguing patterns of evolution seen in these organelles in this clade (Lavrov et al. 2000, 2002; Brewer et al. 2013), and our data will be a vital addition to this effort.

Trigonius corallinus has much potential as a model species. Although the centipede *S. maritima* has an already well-developed community and history as a scientific model, we suggest that *T. corallinus* may be of broader utility as a myriapod model organism. It is a cosmopolitan species available commercially when not collectible in the wild. Unlike *S. maritima* it can be kept in the laboratory and will breed in controllable conditions. Although protocols for its use in a variety of developmental contexts remain to be established, there is no reason why this species could not become an important system for developmental biology research.

As a developmental and genetic model, there is much that could be learnt from a millipede model species. Segmentation, a subject of much interest in the insect developmental community, is presently still to be completely understood in the Myriapoda, with Janssen et al. (2004) suggesting that it may even be separately programmed dorsally and ventrally, at least in *Glomeris marginalia*. Millipedes may also be of developmental utility for understanding the original opening position of the genitalia in arthropods, given the peculiar cephalic location of this opening in this clade, and for understanding neurogenesis across the Arthropoda (e.g., Dove and Stollewerk 2003). Millipedes also undergo periodomorphosis—adult to adult molts, with sexually mature and intercalary instars (Verhoeff 1923; Sahli 1990). Understanding how this is regulated, in contrast with the final adult molts seen in other arthropods, is likely to benefit greatly from a genomic resource.

This data set is therefore a basic resource of interest to a very wide field of scientific endeavor. It contains the majority of coding sequence of *T. corallinus*, a fact that will allow it to be used as a comparison point to other myriapods—notably *S. maritima*—and to the Arthropoda as a whole. This will allow us to learn much about trait evolution in this most diverse of metazoan phyla, and provides a firm basis for further work in the still underinvestigated Diplopoda. The results detailed here will allow work to begin in earnest on a range of developmental, physiological, and genetic questions, providing for the first time a diplopod data set with which to address them. Furthermore, this genome resource will be able to shed some light on the unique pathways possessed by the Myriapoda for chemical defense, and investigation into the antifungal and antipredation adaptations of this underresearched lineage will be greatly aided. As well as being the first diplopod genomic resource publically available, *T. corallinus* also has much to recommend it as a model organism. The combination of this genomic sequence data and the traits and distribution of *T. corallinus* means that this species represents an ideal comparison point for gaining an understanding of the evolution and diversification of the most diverse and ecologically important phylum on the planet today—the Arthropoda.

Supplementary Material

Supplementary files S1–S4 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors thank the AFCD of the HK Government for confirming identity of the millipede species. They also thank BGI for their help in sequencing. This work was supported by the Lo Kwee-Seong Biomedical Research Fund and Lee Hysan Foundation to H-M.L. and J.H.L.H.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Barnes RD. 1982. Invertebrate zoology. Philadelphia (PA): Holt-Saunders International. p. 817–818.
- Brewer MS, Bond JE. 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(11):e79935.
- Brewer MS, Swafford L, Spruill CL, Bond JE. 2013. Arthropod phylogenetics in light of three novel millipede (Myriapoda: Diplopoda) mitochondrial genomes with comments on the appropriateness of mitochondrial genome sequence data for inferring deep level relationships. *PLoS One* 8(7):e68005.
- Brites D, Brena C, Ebert D, Du Pasquier L. 2013. More than one way to produce protein diversity: duplication and limited alternative splicing of an adhesion molecule gene in basal arthropods. *Evolution* 67(10):2999–3011.
- Blum MS. 1981. Chemical defenses of arthropods. London: Academic Press.
- Cameron SL, Miller KB, D'Haese CA, Whiting MF, Barker SC. 2004. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Arthropoda). *Cladistics* 20:534–557.
- Chang WL, Yang CY, Huang YC, Chao D, Chen TW. 2004. Prevalence and observation of intestine-dwelling gregarines in the millipede *Trigoniulus corallinus* (Spirobolida: Pachybolidae) collected from Shoushan, Kaohsiung, Taiwan. *Formos Entomol.* 24:137–145.
- Chapman AD. 2005. Numbers of living species in Australia and the world. Canberra (ACT): Department of the Environment and Heritage, p. 23.
- Chipman AD, et al. 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biol.* 12(11):e1002005.
- Colbourne JK, et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* 331(6017):555–561.
- Damen WG, Saridaki T, Averof M. 2002. Diverse adaptations of an ancestral gill: a common evolutionary origin for wings, breathing organs, and spinnerets. *Curr Biol.* 12(19):1711–1716.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 9:772.
- Dove H, Stollewerk A. 2003. Comparative analysis of neurogenesis in the myriapod *Glomeris marginata* (Diplopoda) suggests more similarities to chelicerates than to insects. *Development* 130(10):2161–2171.
- 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. *Mol Phylogenet Evol.* 49(2):574–585.
- Grillo M, Casanova J, Averof M. 2014. Development: a deep breath for endocrine organ evolution. *Curr Biol.* 24(1):R38–R40.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321.
- Hui JH, et al. 2009. Features of the ancestral bilaterian inferred from *Platynereis dumerilii* ParaHox genes. *BMC Biol.* 7(1):43.
- Hui JH, et al. 2012. Extensive chordate and annelid macrosynteny reveals ancestral homeobox gene organization. *Mol Biol Evol.* 29(1):157–165.
- Hui JH, Holland PW, Ferrier DE. 2008. Do cnidarians have a ParaHox cluster? Analysis of synteny around a *Nematostella* homeobox gene cluster. *Evol Dev.* 10(6):725–730.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8):754–755.

- Janssen R, Pripic NM, Damen WG. 2004. Gene expression suggests decoupled dorsal and ventral segmentation in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). *Dev Biol.* 268(1):89–104.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kenny NJ, et al. 2014. Genomic sequence and experimental tractability of a new decapod shrimp model, *Neocaridina denticulata*. *Mar Drugs.* 12(3):1419–1437.
- Kenny NJ, Shimeld SM. 2012. Additive multiple k-mer transcriptome of the keelworm *Pomatoceros lamarckii* (Annelida; Serpulidae) reveals annelid trochophore transcription factor cassette. *Dev Genes Evol.* 222:325–339.
- Koopman P, Schepers G, Brenner S, Venkatesh B. 2004. Origin and diversity of the Sox transcription factor gene family: genome-wide analysis in *Fugu rubripes*. *Gene* 328:177–186.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947–2948.
- Lavrov DV, Boore JL, Brown WM. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. *Mol Biol Evol.* 19(2):163–169.
- Lavrov DV, Brown WM, Boore JL. 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proc Natl Acad Sci U S A.* 97(25):13738–13742.
- Li J, et al. 2013. Odoriferous defensive stink gland transcriptome to identify novel genes necessary for quinone synthesis in the red flour beetle, *Tribolium castaneum*. *PLoS Genet.* 9(7):e1003596.
- Liu B, et al. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv* 1308:2012.
- Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet.* 52:267–274.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
- Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1):18.
- Mendivil Ramos O, Barker D, Ferrier DE. 2012. Ghost loci imply Hox and ParaHox existence in the last common ancestor of animals. *Curr Biol.* 22(20):1951–1956.
- Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.
- Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft genomes. *Nucleic Acids Res.* 37(1):289–297.
- Pelosi P. 1994. Odorant-binding proteins. *Crit Rev Biochem Mol.* 29:199–228.
- Pelosi P, Zhou JJ, Ban LP, Calvillo M. 2006. Soluble proteins in insect chemical communication. *Cell Mol Life Sci.* 63:1658–1676.
- Pocock RI. 1902. Myriapoda. In: Malden HE, editor. *The Victoria history of the county of Surrey, Vol. 1.* Victoria County History, London. p. 176–178.
- Pocock RI. 1906. Antennata: Myriapoda. *The wild fauna and flora of the Royal Botanic Gardens, Kew. Bulletin of Miscellaneous Information, Additional Series* 5:21–22.
- Rebers JE, Willis JH. 2001. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem Mol Biol.* 31:1083–1093.
- Rehm P, Meusemann K, Börner J, Misof B, Burmester T. 2014. Phylogenetic position of Myriapoda revealed by 454 transcriptome sequencing. *Mol Phylogenet Evol.* 77:25–33.
- Regier JC, et al. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463(7284):1079–1083.
- Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 100:14537–14542.
- Roncadori RW, Duffey SS, Blum MS. 1985. Antifungal activity of defensive secretions of certain millipedes. *Mycologia* 77(2):185–191.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.
- Sahlí F. 1990. On post-adult moults in Julida (Myriapoda, Diplopoda). Why periodomorphosis and intercalaries occur in males. In: *Proceedings of the 7th International Congress of Myriapodology 1987.* (Minelli, A, Ed.). pp. 135–156. E.J. Brill, Leiden.
- Sánchez-Higuera C, Sotillos S, Hombria JCG. 2014. Common origin of insect trachea and endocrine organs from a segmentally repeated precursor. *Curr Biol.* 24(1):76–81.
- Schmieder R, Edwards R. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6:e17288.
- Shear W, Jones T, Miras H. 2007. A possible phylogenetic signal in millipede chemical defenses: the polydesmidan millipede *Leonardesmus injucundus* Shelley & Shear secretes p-cresol and lacks a cyanogenic defense (Diplopoda, Polydesmida, Nearctodesmidae). *Biochem Syst Ecol.* 35:838–842.
- Shear WA, McPherson IS, Jones TH, Loria SF, Zigler KS. 2010. Chemical defense of a troglobiont millipede, *Tetracion jonesi* Hoffman (Diplopoda, Callipodida, Abacionidae). *Int J Myriapol.* 3:153–158.
- Shelley RM, Carmany RM, Burgess J. 2006. Introduction of the millipede, *Trigoniulus corallinus* (Gervais, 1847) (Spiribolida: Trigoniulidae), in Florida, USA. *Entomol News.* 117(2):239–241.
- Shelley RM, Golovatch SI. 2011. Atlas of myriapod biogeography. I. Indigenous ordinal and supra-ordinal distributions in the Diplopoda: perspectives on taxon origins and ages, and a hypothesis on the origin and early evolution of the class. *Insecta Mundi* 0158:1–134.
- Shelley RM, Lehtinen PT. 1999. Diagnoses, synonymies and occurrences of the pantropical millipedes, *Leptogoniulus soromus* (Butler) and *Trigoniulus corallinus* (Gervais) (Spiribolida: Pachybolidae: Trigoniulinae). *J Nat Hist.* 33(9):1379–1401.
- Shimeld SM, Boyle MJ, Brunet T, Luke GN, Seaver EC. 2010. Clustered Fox genes in lophotrochozoans and the evolution of the bilaterian Fox gene cluster. *Dev Biol.* 240:234–248.
- Shimeld SM, Degnan B, Luke GN. 2010. Evolutionary genomics of the Fox genes: origin of gene families and the ancestry of gene clusters. *Genomics* 95:256–260.
- Shinohara K, Hika Y, Nijima K. 2007. Postembryonic development of *Trigoniulus corallinus* (Gervais) (Diplopoda: Spiribolida, Pachybolidae). *Edaphologia* 82:9–16.
- Simpson JT, et al. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19(6):1117–1123.
- Sin YW, et al. 2014. Identification of putative ecdysteroid and juvenile hormone pathway genes in the shrimp *Neocaridina denticulata*. *Gen Comp Endocrinol* 214:167–176.
- Snyder BA, Boots B, Hendrix PF. 2009. Competition between invasive earthworms (*Amyntas corticis*, Megascolecidae) and native North American millipedes (*Pseudopolydesmus erasus*, Polydesmidae): effects on carbon cycling and soil structure. *Soil Biol Biochem.* 41:1442–1449.
- Starostina E, Xu AG, Lin HP, Pikielny CW. 2009. A *Drosophila* protein family implicated in pheromone perception is related to Tay-Sachs GM2-activator protein. *J Biol Chem.* 284:585–594.
- Stoev P, et al. 2010. Myriapods (Myriapoda). *BioRisk* 4:97–130.

- Takahashi T, Holland PW. 2004. Amphioxus and ascidian *Dmbx* homeobox genes give clues to the vertebrate origins of midbrain development. *Development* 131(14):3285–3294.
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis. Version 6.0. *Mol Biol Evol.* 30:2725–2729.
- Undheim EA, et al. 2014. Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes). *Mol Biol Evol.* 31:2124–2148.
- Verhoeff KW. 1923. Periodomorphose. *Zool Anz.* 56(233–238):241–254.
- Weldon PJ, Aldrich JR, Klun JA, Oliver JE, Debboun M. 2003. Benzoquinones from millipedes deter mosquitoes and elicit self-anointing in capuchin monkeys (*Cebus* spp.). *Naturwissenschaften* 90(7):301–304.
- Willis JH. 2010. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem Mol Biol.* 40:189–204.
- Wilson HM, Anderson LI. 2004. Morphology and taxonomy of Paleozoic millipedes (Diplopoda: Chilognatha: Archipolypoda) from Scotland. *J Paleontol.* 78(1):169–184.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255.
- Yu X, Ng CP, Habacher H, Roy S. 2008. FoxJ1 transcription factors are master regulators of the motile ciliogenic program. *Nat Genet.* 40(12):1445–1453.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18(5):821–829.
- Zhong YF, Butts T, Holland PWH. 2008. HomeoDB: a database of homeobox gene diversity. *Evol Dev.* 10(5):516–518.
- Zhong YF, Holland PWH. 2011. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev.* 13(6):567–568.

Associate editor: Jay Storz