# Sizing Up the Onychophoran Genome: Repeats, Introns, and Gene Family Expansion Contribute to Genome Gigantism in *Epiperipatus broadwayi*

Shoyo Sato [1,*], Tauana J. Cunha [1,2], Bruno A.S. de Medeiros [1,2], Danielle E. Khost[3], Timothy B. Sackton [3], and Gonzalo Giribet [1]

[1]Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts
[2]Field Museum of Natural History, Chicago, Illinois
[3]FAS Informatics Group, Harvard University, Cambridge, Massachusetts

*Corresponding author: E-mail: shoyosato@g.harvard.edu.

**Accepted:** 21 January 2023

## Abstract

Genome assemblies are growing at an exponential rate and have proved indispensable for studying evolution but the effort has been biased toward vertebrates and arthropods with a particular focus on insects. Onychophora or velvet worms are an ancient group of cryptic, soil dwelling worms noted for their unique mode of prey capture, biogeographic patterns, and diversity of reproductive strategies. They constitute a poorly understood phylum of exclusively terrestrial animals that is sister group to arthropods. Due to this phylogenetic position, they are crucial in understanding the origin of the largest phylum of animals. Despite their significance, there is a paucity of genomic resources for the phylum with only one highly fragmented and incomplete genome publicly available. Initial attempts at sequencing an onychophoran genome proved difficult due to its large genome size and high repeat content. However, leveraging recent advances in long-read sequencing technology, we present here the first annotated draft genome for the phylum. With a total size of 5.6Gb, the gigantism of the *Epiperipatus broadwayi* genome arises from having high repeat content, intron size inflation, and extensive gene family expansion. Additionally, we report a previously unknown diversity of onychophoran hemocyanins that suggests the diversification of copper-mediated oxygen carriers occurred independently in Onychophora after its split from Arthropoda, parallel to the independent diversification of hemocyanins in each of the main arthropod lineages.

**Key words:** velvet worm, Panarthropoda, hemocyanin, Peripatidae.

## Significance

Onychophora or velvet worms are of interest for understanding the evolution of reproductive biology, biogeography, and the evolution of its sister group Arthropoda, the most diverse lineage of animals. Despite their significance, there are no published genomes for the phylum and only one highly fragmented and incomplete genome is available on GenBank. Here we report the first annotated onychophoran genome, a case of genome gigantism, and a note on the evolution of hemocyanins.

## Introduction

Onychophora, otherwise known as "velvet worms" or "peripatus," are unique among the Metazoa as being the only exclusively terrestrial animal phylum. They are soft-bodied, many-legged, animals that inhabit permanently moist microhabitats. The phylum is divided into two families: the circumtropical Peripatidae, and the temperate Gondwanan Peripatopsidae (Oliveira et al. 2012; Giribet and Edgecombe 2020) that diversified before the breakup of Gondwana and display strong biogeographic affinities

(Monge-Nájera 1995; Murienne et al. 2014; Giribet et al. 2018). Onychophora is particularly notable for their distinctive prey capture mechanism in which they secrete a glue from two modified, anterior appendages to entangle prey (Benkendorff et al. 1999; Haritos et al. 2010; Baer et al. 2017; Baer et al. 2019). Additionally, despite their highly conserved morphology, they are noted for their remarkable diversity of reproductive strategies ranging from oviparity, through ovoviviparity, to placental viviparity which has been implicated in their dispersal capabilities and subsequent radiation (Anderson 1973; Mayer et al. 2015; Baker et al. 2021).

In a broader context, Onychophora hold a pivotal place in metazoan phylogenetics as the sister group to Arthropoda (Dunn et al. 2008; Rota-Stabelli 2010; Laumer et al. 2019) and are traditionally grouped into the clade Panarthropoda along with Tardigrada. The discrepancy in diversity and disparity between Onychophora, with around 200 species, and Arthropoda, the most diverse animal phylum comprising 80% of described living species of animals, raises important questions regarding broad macroevolutionary patterns such as morphological evolution and diversification. Panarthropods have contributed heavily to the field of genomics but taxon representation has been extremely uneven with most of this effort focused on arthropods, and in particular insects, with hundreds of genomes publicly available for this group. Tardigrade genomics has developed into an emerging field and the group is now represented by four genomes on National Center for Biotechnology Information (NCBI) including a chromosome-level assembly (Hoencamp et al. 2021). Exhaustive studies have been conducted on tardigrade genomes including intense debate on horizontal gene transfer (Arakawa 2022 for review). However, there is a near complete lack of data on Onychophora, with no published genomes and only one highly fragmented draft assembly publicly available in NCBI (GenBank: GCA_003024985.2). There have been several dozen onychophoran transcriptomes generated to date but of only modest quality. Benchmarking Universal Single-copy Orthologs (BUSCO) scores for these transcriptomes range from 6.65% to 82.62% with a mean of 46%. To rectify the lack of genomic resources for this phylum, we present here the first high-quality, annotated onychophoran genome from *Epiperipatus broadwayi*.

## Results and Discussion

### Sequencing and Assembly

In line with estimates predicted with flow cytometry and Feulgen image analysis densitometry (Mora et al. 1996; Jeffery et al. 2012), we estimated the genome size to be 5.6Gb with kmer frequencies. Using long-read sequence data from Oxford Nanopore Technologies (ONT) (supplementary Table S1, Supplementary Material online) and long-range information from Hi-C, we were able to assemble the most contiguous and complete genome for the phylum to date. These data produced a final scaffolded assembly with a total length of 5.6Gb and a scaffold N50 of 5.3Mb (Table 1). Using the Arthropoda OrthoDB (Kriventseva et al. 2019), BUSCO analysis found 89.5% completeness (Table 1). In comparison, the publicly available genome of *Euperipatoides rowelli* is only 40.5% complete.

### Genome Gigantism

The genomes of Onychophora are known examples of genome gigantism (Mora et al. 1996; Jeffery et al. 2012) and this phenomenon seems to be correlated with several key features of their genome organization. We identified 70.92% (3.97Gb) of the *Epiperipatus broadwayi* genome as repetitive (fig. 1) which is some of the highest proportions of repeats in an assembled invertebrate genome. The average intron length of *Epiperipatus broadwayi* is 5X higher than those of most arthropods and exceeds those of many vertebrates (supplementary Table S2, Supplementary Material online). Both repeat content and intron size inflation are correlated with genome size and responsible for giant genomes within Panarthropoda (Wang 2014; Verlinden 2020) and among Metazoa (Nowoshilow 2018; Meyer 2021). These large introns and extensive repetitive elements likely contribute to the large estimated genome size in all Onychophora studied to date (Mora et al. 1996; Jeffery et al. 2012).

### Gene Expansion

The annotation pipeline predicted a total of 46,891 genes (57,420 transcripts). Orthofinder found a large fraction of genes ($n = 27,304$), in the genome of *Epiperipatus broadwayi* with no orthologs in other lineages. Only 1,506 of these genes were identified as putatively TE or TE-derived. Additionally, we were able to find expression of around 21% of these transcripts (>5 transcripts per million) but at a lower expression compared to transcripts with orthologs found in other species (supplementary fig. S1, Supplementary Material online). Additionally, half of the unmatched orthogroups were represented by at least one contig in either the genomes of two peripatopsids: *Euperipatoides rowelli* or *Peripatoides* sp. The incomplete nature of the transcriptome and two additional genomes (BUSCO scores of 67.7%, 40.5%, and 37%, respectively) and the Devonian divergence between peripatids and peripatopsids (Baker et al. 2021) could account for the missing orthogroups. It is unclear whether these genes are truly specific to Onychophora or were undetectable with current methods. It is possible that high divergence rates in these genes could confound the deep divergence between this

**Table 1**

Genome summary statistics from QUAST and BUSCO results of the scaffold level assembly of *Epiperipatus broadwayi*. Scaffold level assembly statistics for *Euperipatoides rowelli* (GenBank: GCA_003024985.2), the only publicly available onychophoran genome, are shown for comparison

| Species | Coverage and Sequencing Technology | Assembly Algorithm | Total Length (bp) | Number of Scaffolds | Scaffold N50 (bp) | BUSCO |
|---|---|---|---|---|---|---|
| *Euperipatoides rowelli* | 140× Illumina | AllPathsLG | 2,681,849,960 | 311,309 | 14,341 | 40.5% *C*:410 [*S*:383, *D*:27], *F*:336, *M*:267, *n*:1013 |
| *Epiperipatus broadwayi* | 26× ONT 100× Illumina 60× Hi-C | FLYE + SALSA2 | 5,601,407,943 | 18,588 | 5,322,020 | 89.5% *C*:907 [*S*:880, *D*:27], *F*:69, *M*:37, *n*:1013 |

lineage and its sister group, making the identification of homologs in other phyla difficult. As this is the only high-quality onychophoran genome sequenced to date, the sequencing and assembly of more onychophoran species across both families will help elucidate the true gene content of these genomes.

### Onychophoran Hemocyanins

The increase in body size of metazoans during the Cambrian, along with later terrestrialization events necessitated the evolution of an efficient circulatory system to transport oxygen through large bodies. One key molecule implicated in the growth of body size and in the colonization of land in two of the most successful phyla of animals is the oxygen carrier hemocyanin. To test the utility of our new genome, we searched for putative hemocyanins using standard gene mining methods. Phylogenetic reconstruction of hemocyanins revealed a clade of onychophoran sequences sister group to crustacean hemocyanins (supplementary fig. S2*a*, Supplementary Material online). The three arthropod classes represented in the analysis were found to be monophyletic but the relationships among the classes were spurious, a result expected from single gene phylogenies of Cambrian divergences. We recovered eight full transcripts with the presence of key amino acid residues involved in the oxygen-transporting mechanism of hemocyanin (supplementary fig. S2*b*, Supplementary Material online) (Hazes et al. 1993; Kusche et al. 2002). The presence of a single clade of onychophoran hemocyanins implies radiation of hemocyanin subunits in Onychophora independent from that in arthropods. This suggests an ancestral hemocyanin in the ancestor of onychophorans and arthropods that subsequently diversified in each of those lineages possibly driven by increased body size (Burmester 2002; Kusche et al. 2002), or improvements to hemocyanin stability, cooperativity, and regulation (Rehm et al. 2012)

## Conclusion

The genomes of Onychophora had previously proved difficult to assemble due to their large sizes and high repeat content. Leveraging long-read sequencing technology, high coverage short-read sequencing, and chromosome conformation capture sequencing, we present the first annotated draft genome for the phylum. The discovery of a previously unknown diversity of putative hemocyanins in this new genome has already demonstrated its utility in understanding the evolution of key characteristics in panarthropods. Holding a key phylogenetic position as a sister group to Arthropoda, sequencing the onychophoran genome is an important step in understanding the radiation of the largest phylum of animals. Additionally, unresolved questions regarding the biogeography, reproduction, and phylogenetics of this little understood group of animals can be tackled with the help of new genomic resources. The genome of *Epiperipatus broadwayi* is exceptionally repeat dense, which, along with intron elongation, contributes to its large size. This is the first step in generating genomic resources for this phylum, the last among the panarthropods to enter the genomic era.

## Materials and Methods

### DNA Extraction and Sequencing

Genomic DNA (gDNA) was extracted from the trunk tissue of a single individual (https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:143930) using a high salt protocol and repaired with PreCR Repair Mix (New England Biolabs) followed by a chloroform cleanup. DNA quality was assessed by using a Nanodrop spectrophotometer (Thermo Fisher Scientific), Qubit fluorometer (Thermo Fisher Scientific), and Tapestation gDNA Screen tape (Agilent).

Two libraries were prepared for separate sequencing runs on the MinION (Oxford Nanopore Technologies). One was sheared by pipetting and cleaned using the Circulomics short-read eliminator kit (Circulomics) and the other was directly cleaned using the Circulomics short-read eliminator XL kit (Circulomics). Libraries were prepped using the SQK-LSK109 kit (Oxford Nanopore Technologies). Flow cells were cleaned using the Flow Cell Wash Kit (Oxford Nanopore Technologies) and reloaded for a total of three loads per library. A larger sequencing runs on DNA from the same individual, sheared to 29 kb using Covaris g-tube (Covaris), was prepped and conducted on the PromethION (Oxford Nanopore Technologies) at the Bauer Core Facility

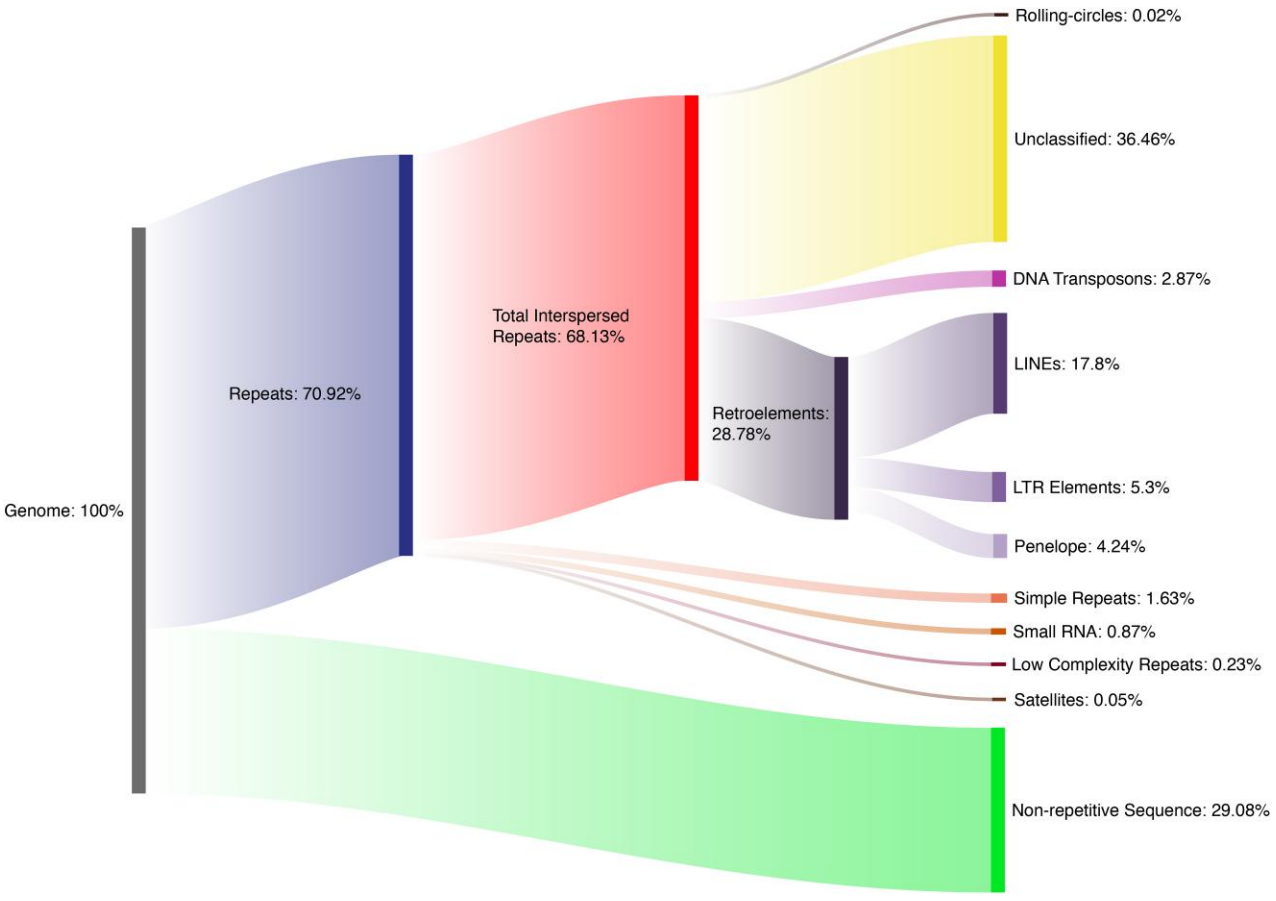


**Fig. 1.**—Repetitive element composition in the genome of *Epiperipatus broadwayi*. Class percentages can be higher than the sum of subclasses due to uncertainty in classification and excluded subclasses. Other discrepancies among the percentages are due to short, unmasked regions between adjacent simple repeats which are annotated as a single stretch.

at Harvard University. Whole-genome shotgun and Hi-C libraries were prepared using the same individual using the Kapa Hyper-Plus library preparation kit (Roche) and Arima Hi-C kit (Arima Genomics) combined with the Kapa Hyper-Prep kit (Roche) respectively. Both libraries sequenced for 150 bp paired-end reads in an Illumina NovaSeq S4 flow cell at the Bauer Core Facility at Harvard University.

## Assembly

Unique 21mers were counted in raw Illumina reads using Jellyfish (Marcais and Kingsford 2011) to estimate genome size. Nanopore data were base called using Guppy v.4.5.2. Raw Nanopore reads were assembled using Flye with default settings (Kolmogorov et al. 2019). Both Nanopore and Illumina read were used to polish the assembly using medaka and HyPo as a final polish (Kundu et al. 2019). After processing with TrimGalore (Krueger 2021), Hi-C reads were used for scaffolding with SALSA2 (Ghurye et al. 2019) following mapping with the Arima-Hi-C mapping pipeline. Scaffolds were screened for contaminants using Blobtools (Laetsch and Blaxter 2017) and scaffolds of dubious origin were manually checked using BLAST (Altschul et al. 1990). Haplotigs were removed using Purge Haplotigs (Roach et al. 2018). The genome assembly workflow can be viewed in supplementary figure S3*a*, Supplementary Material online. Completeness was assessed using BUSCO (Simão et al. 2015; Waterhouse 2018) and the Arthropoda OrthoDB v10 (Kriventseva et al. 2019).

## Annotation

Transposable elements (TEs) were identified and masked using RepeatModeler (Flynn et al. 2020) and RepeatMasker (Smit et al. 2021). The masked assembly was then annotated with BRAKER2 (Bruna et al. 2021) using three databases: a custom Onychophora peptide database from the translated transcriptomes (supplementary Table S3, Supplementary Material online), the Arthropoda OrthoDB v10 (Kriventseva et al. 2019), and the untranslated transcriptome of *Epiperipatus broadwayi* (GenBank: SRX10007847). The three predictions were combined using TSEBRA (Gabriel et al. 2021). Functional annotations were added to the final protein predictions using the online Orthologous Matrix (OMA) browser (Altenhoff et al.

2015) and eggNOG (Cantalapiedra et al. 2021). The full annotation workflow can be viewed in supplementary figure S3*b*, Supplementary Material online.

### Gene Expansion Analysis

Gene family expansion analysis was conducted with Orthofinder (Emms and Kelly 2019) using genomes of major panarthropod phyla (supplementary Table S3, Supplementary Material online). Due to the extensive expansion of several gene families, the predicted proteins were compared to the Pfam database (Mistry 2021) using Hmmer v.3.2.1 (Eddy 2011) to check for TEs or transposable-element-associated proteins. Orthogroups with at least 25% of transcripts hitting known TE-associated domains were considered a TE-derived orthogroup. Kallisto v0.46.1 (Bray et al. 2016) was used to quantify transcript abundances using RNA (GenBank: SRX10007847). To check for the presence of the putative onychophoran-specific orthogroups in other available genomes, DIAMOND blastx (Buchfink et al. 2021) was run with the *Euperipatoides rowelli* genome from NCBI (GenBank: GCA_003024985.2) and a newly sequenced genome of *Peripatoides* sp. (unpublished, https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:144603) using *Epiperipatus broadwayi* as a database.

### Onychophoran Hemocyanins

Representative arthropod hemocyanin protein fasta files from GenBank (supplementary Table S3, Supplementary Material online) were used to identify putative hemocyanins in the genome with blastp v2.12.0 (Altschul et al. 1990) and Hmmer v.3.2.1 (Eddy 2011). The resulting transcripts were then used for reciprocal BLAST searches against the NCBI nr database. Additional sequences of arthropod phenoloxidases were added as outgroups to the analysis (supplementary Table S3, Supplementary Material online) (Burmester 2001; Kusche et al. 2002). All sequences were combined and the resulting matrix was aligned using MAFFT (Katoh and Standley 2013). Gaps, divergent sequences, and short sequences were removed using CIAlign v1.0.17 (Tumescheit et al. 2022). Model testing, phylogenetic tree reconstruction, and branch support assessment were conducted using IQ-TREE (Nguyen et al. 2015; Hoang et al. 2017; Kalyaanamoorthy et al. 2017).

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

Specimens were collected under permits for Trinidad & Tobago (Trinidad permit A No. 1176; B No. 000573; Tobago Special Game License without number). Ligia Benavides, Julia Cosgrove, and Gustavo Hormiga participated in the fieldwork. Christopher Laumer provided his modified salting protocol for DNA extraction. Parts of sequencing and library preparation were done at the Bauer Core Facility at Harvard University. The computations in this paper were run on the FASRC Cannon cluster supported by the Factul of Arts and Sciences Division of Science Research Computing Group at Harvard University. This work was supported by internal funds from the Museum of Comparative Zoology, by a Faculty of Arts and Sciences Dean's Fund award; and by the National Science Foundation (grant number 2154245) "Collaborative Research: PurSUiT: Understanding the Neotropical Velvet Worms (Onychophora, Peripatidae, Neopatida), a Cretaceous Radiation of Terrestrial Panarthropods". This work was published by a grant from the Wetmore Colles fund. Two anonymous reviewers and the Associate editor provided comments that improved this paper.

## Data Availability

New sequences and the assembled genome are deposited in the Sequence Read Archive (BioProject PRJNA866281, BioSample SAMN30158680, SRA Accession Numbers SRR20852068–SRR20852072, and GenBank JAQFVV000000000. Nucleotide fasta, peptide fasta, detailed methods, Blobtools, and annotation outputs, along with the alignment and newick tree file for the hemocyanin analysis are deposited in Harvard Dataverse: https://doi.org/10.7910/DVN/IKUCMY.

## References

Altenhoff AM, et al. 2015. The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res. 43(Database issue):D240–D249.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Anderson DT 1973. Onychophorans. In Embryology and phylogeny in annelids and arthropods. Oxford: Pergamon.

Arakawa K. 2022. Examples of extreme survival: tardigrade genomics and molecular anhydrobiology. Annu Rev Anim Biosci. 10:17–37.

Baer A, et al. 2017. Mechanoresponsive lipid-protein nanoglobules facilitate reversible fibre formation in velvet worm slime. Nat Commun. 8(1):974.

Baer A, Schmidt S, Mayer G, Harrington MJ. 2019. Fibers on the fly: multiscale mechanisms of fiber formation in the capture slime of velvet worms. Integr Comp Biol. 59(6):1690–1699.

Baker CM, Buckman-Young RS, Costa CS, Giribet G. 2021. Phylogenomic analysis of velvet worms (Onychophora) uncovers an evolutionary radiation in the neotropics. Mol Biol Evol. 38(12):5391–5404.

Benkendorff K, Beardmore K, Gooley AA, Packer NH, Tait NN. 1999. Characterization of the slime gland secretion from the peripatus, *Euperipatoides kanangrensis* (Onychophora: Peripatopsidae). Comparative Biochemistry and Physiology Part B 124:457–465.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 34(5):525–527.

Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with

GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics Bioinformatics 3(1):lqaa108.
Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 18(4): 366–368.
Burmester T. 2001. Molecular evolution of the arthropod hemocyanin superfamily. Mol Biol Evol. 18(2):184–195.
Burmester T. 2002. Origin and evolution of arthropod hemocyanins and related proteins. J Comp Physiol. B 172(2):95–107.
Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 38(12):5825–5829.
Dunn CW, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452(7188):745–749.
Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7(10):e1002195.
Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238.
Flynn JM, et al. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 117-(17):9451–9457.
Gabriel L, Hoff KJ, Bruna T, Borodovsky M, Stanke M. 2021. TSEBRA: transcript selector for BRAKER. BMC Bioinf. 22(1):566.
Ghurye J, et al. 2019. Integrating hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 15(8):e1007273.
Giribet G, et al. 2018. The 'peripatos' in eurogondwana?—lack of evidence that southeast Asian onychophorans walked through Europe. Invertebr Syst. 32(4):842.
Giribet G, Edgecombe GD. 2020. The invertebrate tree of life. Princeton, NJ; Oxford: Princeton University Press.
Haritos VS, et al. 2010. Harnessing disorder: onychophorans use highly unstructured proteins, not silks, for prey capture. Proc Biol Sci. 277(1698):3255–3263.
Hazes B, et al. 1993. Crystal structure of deoxygenated *Limulus polyphemus* subunit II hemocyanin at 2.18 Å resolution: clues for a mechanism for allosteric regulation. Protein Sci. 2:597–619.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2017. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522ka.
Hoencamp C, et al. 2021. 3D Genomics across the tree of life reveals condensin II as a determinant of architecture type. Science 372-(6545):984–989.
Jeffery NW, Oliveira IS, Gregory TR, Rowell DM, Mayer G. 2012. Genome size and chromosome number in velvet worms (Onychophora). Genetica 140(10–12):497–504.
Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. Modelfinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14(6):587–589.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37(5): 540–546.
Kriventseva EV, et al. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47(D1):D807–D811.
Krueger F. 2021. Trim Galore! (Version v.0.6.7). Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Kundu R, Casey J, Sung W. 2019. Hypo: super fast & accurate polisher for long read genome assemblies. bioRxiv https://doi.org10.1101/2019.12.19.882506
Kusche K, Ruhberg H, Burmester T. 2002. A hemocyanin from the onychophora and the emergence of respiratory proteins. Proc Natl Acad Sci U S A. 99(16):10545–10548.
Laetsch DR, Blaxter ML. 2017. Blobtools: interrogation of genome assemblies. F1000Research 6:1287.
Laumer CE, et al. 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. Proc Royal Soc B: Biol Sci. 286(1906):20190831.
Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6): 764–770.
Mayer G, Franke FA, Treffkorn S, Gross V, Oliveira IS. 2015. Onychophora. In: Wanninger A, editors. Evolutionary developmental biology of invertebrates 3: ecdysozoa I: non-tetraconata. Vienna: Springer. p. 53–98.
Meyer A, et al. 2021. Giant lungfish genome elucidates the conquest of land by vertebrates. Nature 590(7845):284–289.
Mistry J, et al. 2021. Pfam: the protein families database in 2021. Nucleic Acids Res. 49(D1):D412–D419.
Monge-Nájera J. 1995. Phylogeny, biogeography and reproductive trends in the onychophora. Zool J Linn Soc. 114:21–60.
Mora M, Herrera A, León P. 1996. The genome of *Epiperipatus biolleyi* (peripatidae), a costa rican onychophoran. Rev Biol Trop. 44(1): 153–157.
Murienne J, Daniels SR, Buckley TR, Mayer G, Giribet G. 2014. A living fossil tale of pangaean biogeography. Proc Royal Soc B Biol Sci. 281(1775):20132648.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.
Nowoshilow S, et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. Nature 554(7690):50–55.
Oliveira IS, Read VM, Mayer G. 2012. A world checklist of onychophora (velvet worms), with notes on nomenclature and status of names. Zookeys 211:1–70.
Rehm P, Pick C, Borner J, Markl J, Burmester T. 2012. The diversity and evolution of chelicerate hemocyanins. BMC Evol Biol. 12:19.
Roach MJ, Schmidt SA, Borneman AR. 2018. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinf. 19(1):460.
Rota-Stabelli O, et al. 2010. Ecdysozoan mitogenomics: evidence for a common origin of the legged invertebrates, the Panarthropoda. Genome Biol Evol. 2:425–440.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Smit AFA, Hubley S, Green P. 2021. RepeatMasker (Version 4.1.2). Retrieved from http://repeatmasker.org
Tumescheit C, Firth AE, Brown K. 2022. CIAlign: a highly customisable command line tool to clean, interpret and visualise multiple sequence alignments. PeerJ 10:e12983.
Verlinden H, et al. 2020. First draft genome assembly of the desert locust, *Schistocerca gregaria*. F1000Research 9:775.
Wang X, et al. 2014. The locust genome provides insight into swarm formation and long-distance flight. Nat Commun. 5:2957.
Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 35(3): 543–548.

**Associate editor**: Prof. Cristina Vieira