

A Chromosome-level Assembly and Functional Genomic Resources for the Model Annelid *Capitella teleta*

Billie E. Davies ¹, Paul Gonzalez ², Abhinav Sur ³, Jingcheng Wei ¹, Tom Frankish ¹, Jimena Montagne ¹, Allan M. Carrillo-Baltodano ¹, Kero Guynes ^{1,4}, Yan Liang ^{1,5}, Rory D. Donnellan ¹, R. Travis Moreland ², Sumeeta Singh², Suiyuan Zhang ², Reynold Yu ², Tyra G. Wolfsberg ², Néva P. Meyer ⁶, Elaine C. Seaver ^{7,*}, Andreas D. Baxevas ^{2,*}, José M. Martín-Durán ^{1,*}

¹School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4NS, UK

²Center for Genomics and Data Science Research, Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

³Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

⁴Blizard Institute, Barts and the London School of Medicine and Dentistry, QMUL, London E1 2AT, UK

⁵Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, Cambridge, UK

⁶Biology Department, Clark University, Worcester, MA, USA

⁷Whitney Laboratory for Marine Bioscience, University of Florida, Saint Augustine, FL 32080, USA

*Corresponding authors: E-mails: seaver@whitney.ufl.edu; andy@mail.nih.gov; chema.martin@qmul.ac.uk.

Accepted: February 16, 2026

Abstract

The polychaete *Capitella teleta* is a primary model for evolutionary developmental biology, comparative genomics, conservation, and ecotoxicology. Although it was the first polychaete genome sequenced, the original assembly is outdated by modern standards. Here, we combine long-read and short-read sequencing with Hi-C chromatin conformation capture to assemble chromosome-level nuclear and mitochondrial genomes of the laboratory strain of *C. teleta*. This reference assembly accurately reflects the expected genome size (~243.6 Mb) and contains a highly complete, evolutionarily conserved gene repertoire. Notably, the nuclear and mitochondrial genomes are heavily rearranged, indicating a decoupling between gene family repertoire and chromosomal evolution. The analyses of developmental time courses of bulk and single-cell RNA-seq, ATAC-seq, and EM-seq data using the new reference assembly resulted in a significant improvement in quality, enabling the identification of new cell-type-specific gene markers. Finally, we generated a publicly available genome browser that ensures these resources comply with FAIR principles. Our study provides state-of-the-art genomic resources for *C. teleta*, addressing a pressing community need and opening new research opportunities in animal and genome evolution.

Key words: *Capitella teleta*, Annelida, Spiralia, genome browser, mitochondrial genome, single-cell RNA-seq.

Introduction

Capitella teleta (Blake et al. 2009; formerly known as *Capitella* sp. I) is a marine polychaete (Annelida) widely used as a model in developmental and regenerative biology, comparative genomics, conservation biology,

and ecotoxicology (Fig. 1a and b) (Blake et al. 2009). Initially cultured as a laboratory strain more than 50 years ago, *C. teleta* is among the best-studied annelids (Seaver 2016). Its highly stereotyped cleavage has enabled the generation of a comprehensive fate map (Meyer et al. 2010, Meyer and Seaver 2010), and the

Significance

Capitella teleta is a leading spiralia model, but its reference genome has become outdated by modern standards. Here, we combine long-read sequencing and Hi-C scaffolding to generate a high-quality, chromosome-level assembly and annotate a complete, slow-evolving gene repertoire. We further provide a complete mitochondrial genome and demonstrate that the improved gene annotations significantly enhance the resolution of functional genomic datasets, particularly for single-cell RNA-seq analyses. By integrating these data into a publicly accessible genome browser, we provide a centralized resource to facilitate comparative genomic studies in this key annelid model.

year-round availability of embryos in laboratory cultures, synchronized development, and developmental staging system have greatly facilitated the establishment of *C. teleta* as a tractable experimental system (Seaver 2022). Published functional genomic methods, such as CRISPR/Cas9 mutagenesis, morpholino knock-down, exogenous RNA expression, and transcriptomic profiling, as well as experimental techniques such as blastomere and juvenile microinjections, microsurgery, and single-blastomere deletions, have been successfully applied (Carrillo-Baltodano and Meyer 2017, Klann and Seaver 2019, Neal et al. 2019, Lanza and Seaver 2020, Seaver and de Jong 2021, Martín-Zamora et al. 2023, Guynes et al. 2024, Liang et al. 2025). Consequently, the genome of *C. teleta* was the first polychaete genome and one of the first spiralian genomes to be sequenced (Simakov et al. 2013). While existing genomic data have greatly facilitated studies, the field would benefit from highly contiguous whole-genome sequence data to bring the molecular study of *C. teleta* to current gold standards.

Results and Discussion

A Chromosome-level Genome Assembly

To generate a chromosome-level genome assembly of *C. teleta*, we sequenced ~103Gb Oxford Nanopore long reads, ~58.4 Gb of short Illumina reads (Table S1), and ~9 Gb of HiC data. Matching the karyotype of *C. teleta* (Grassle et al. 1987), 92% of the assembly length is grouped into 10 pseudochromosome scaffolds (Fig. 1c; Fig. S1a and b), although chromosome 3 remains divided into two large scaffolds. The resulting assembly, at 277.13 Mb, is closer to the estimated genome size of *C. teleta* (243.6 Mb), exhibits low heterozygosity (0.39% to 0.41%), is highly complete (96.37% of the short reads map to the assembly), and is 47 Mb shorter than the original assembly (Fig. 1d; Fig. S1c). Despite this difference, the original and new genome assemblies align well (Fig. S2a). This chromosome-level assembly exhibited a modest improvement in genome completeness (metazoan_odb10: 97% [92.2% single-copy and 4.7% duplicated] vs. 94.7% [90.3% single-

copy and 4.4% duplicated] completeness; Fig. 1d). Repeat elements now represent 40.71%, an increase from 31.76% in the original version (Simakov et al. 2013) (Fig. 1e; Table S2). However, the age distribution of transposable elements remains equivalent across both assemblies (Fig. 1f). *De novo* functional annotation predicted 26,591 protein-coding genes (Fig. 1d), 5,798 fewer than in the original assembly (Simakov et al. 2013), which is consistent with a smaller assembly length and the removal of redundant haplotypes. Yet, both annotations are broadly similar, as indicated by the high proportion of complete BUSCO genes in both versions, as well as by the 17,797 genes (66.93%) in the new annotation that have a best reciprocal BLAST hit to a gene in the original assembly (Fig. S2b). The increase in assembly contiguity resulted in longer gene and transcript lengths, improved UTR annotations (Fig. 1g), and more genes with assigned biological functions (eg 67.1% of transcripts have a BLASTX hit in the new annotation compared to 59.8% in the old assembly; Fig. S2c and d). Thus, our new chromosome-level assembly for *C. teleta* provides a substantially enhanced, high-quality reference genome that will be invaluable for future functional and comparative studies.

Gene Family Evolution

We performed gene-family evolution analyses to identify gene gains and losses (Tables S3 and S4). *Capitella teleta* has retained a comparable number of ancestral Eumetazoan and Bilaterian orthogroups similar to other slow-evolving annelids, such as *O. fusiformis*, and even humans (Fig. S3a and b) (Martín-Zamora et al. 2023), which is also reflected in the higher proportion of orthologs (including one-to-one orthologs) with the human genome (Fig. S3c). A Gene Ontology (GO) term enrichment analysis (Table S5) indicated that most of the lost gene families are involved in metabolism, while expanded gene families are involved in developmental processes, neurodevelopment, and the immune response (Fig. S3d and e; Table S6). Consistent with previous analyses (Simakov et al. 2013, Nath et al. 2025), we also identified 214 genes grouped into five expanded orthogroups containing at least one gene annotated

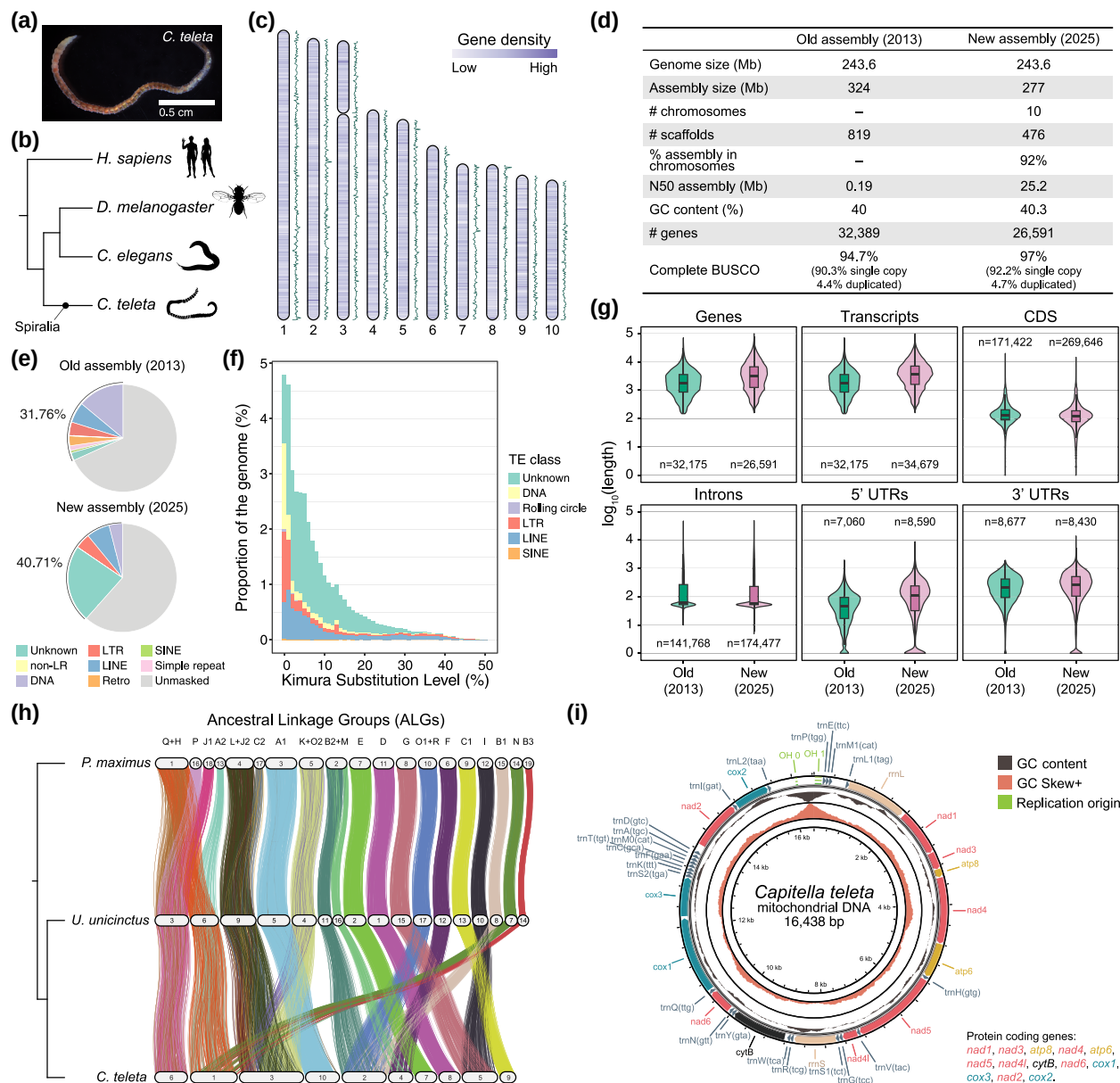


Fig. 1. A chromosome-level genome assembly and updated annotation of the annelid *Capitella teleta*. a) An adult *C. teleta* specimen. Scale bar, 0.5 cm. b) Phylogenetic tree showing the evolutionary position of *C. teleta* (a spiralian) relative to common model organisms. c) Ideogram of the 10 pseudochromosomes in the new assembly, with shading indicating gene density and the green line indicating transposable elements (TEs) density. d) Table comparing key metrics of the old (2013; [Simakov et al. 2013]) and new (2025; this study) genome assemblies, highlighting significant improvements in contiguity (N50 increased from 0.19 Mb to 25.2 Mb) and contiguity (92% of the assembly is anchored to pseudochromosomes). e) Pie charts comparing the annotated repeat content of the old and new assemblies. The new assembly identifies a larger proportion of the genome as repetitive elements (31.76% vs. 40.71%). f) Transposable element (TE) landscape of the new assembly. The abundance of different TE classes is plotted against their Kimura substitution level, indicating bursts of TE activity (peak at low divergence) that are progressively removed. g) Violin plots comparing the length distributions of annotated genes and gene features between the old and new assemblies. The new assembly resolves significantly longer genes, transcripts, and introns, while the lengths of coding sequences (CDS) and untranslated regions (UTRs) remain comparable. Sample sizes (*n*) for each feature are indicated on the plots. h) Conserved synteny between the annelids *Urechis unicinctus* and *C. teleta*, and the mollusc *Pecten maximus* (as a close representative of the ancestral state). Unlike *U. unicinctus* and other polychaetes, *C. teleta* has experienced multiple fusions and fissions of the ancestral linkage groups. Lines connect orthologous genes, and colored blocks represent ancestral linkage groups. i) Gene map of the 16,438 bp mitochondrial genome of *C. teleta* (lab strain). Genes are shown on the outer ring. Inner rings depict GC content (black) and GC skew (orange). Replication origins are shown in light green. *Capitella teleta*'s mitochondrial genome includes 13 protein-coding genes, two ribosomal RNAs and 23 tRNAs.

as a GPCR, two of which were shared with *Urechis unicinctus* (the sister lineage to *C. teleta*) (Weigert et al. 2014) and may represent ancestral expansions (Tables S5 and S6). Finally, ancestral linkage group conservation in *C. teleta* revealed 10 fusions, as well as the fission of one chromosomal arrangement (H+Q) present in the last common annelid ancestor (Fig. 1h; Fig. S4a–d). The fission of H+Q and its subsequent fusion with A2 and P+J1, respectively, is also present in *U. unicinctus* (Lewin et al. 2024) and may represent a shared synapomorphy for the clade rather than *C. teleta*-specific divergences. Therefore, *C. teleta*, like *O. fusiformis*, is a slow-evolving annelid with a well-conserved gene repertoire but numerous chromosomal rearrangements (Martín-Zamora et al. 2023, Lewin et al. 2024).

The Mitochondrial Genome

Capitella teleta's mitogenome has two replication origins, a generally positive GC skew, and, as in other annelid mitochondrial genomes (Weigert et al. 2016, Struck et al. 2023), a complete gene set composed of 13 coding genes, two rRNAs, and 22 tRNAs (including two tRNAs for methionine, M0 and M1) (Fig. 1i) that has experienced extensive syntenic rearrangements compared to other Capitellidae and annelid lineages (Fig. S5). Although most genes (8/13; 61.5%) use the standard ATG start codon, there are alternative start codons, such as ATT (3/13; 23.1%), ATA (1/13; 7.7%; in *nad6*) and the relatively rare TTA (1/13; 7.7%; in *cox1*). The species description of *C. teleta* reported the holotype's *cox1* sequence (Blake et al. 2009), which appears to be profoundly divergent but still closely related to the one reported here (Figs. S6 and S7; Supplementary File S1). Indeed, the phylogenetic analysis of the publicly available *cox1* sequences revealed a complex phylogenetic pattern for the genus *Capitella*. While *Capitella capitata* and *C. neociculata* likely represent species complexes (Grassle and Grassle 1976), *C. teleta* (cosmopolitan, including the holotype and the *cox1* sequence reported here) is a well-defined monophyletic clade (Figs. S7 and S8). These results underscore the need for a comprehensive reassessment of the genus *Capitella* that incorporates additional mitochondrial and nuclear genes alongside morphological descriptions.

Functional Genomic Resources

We remapped and re-analyzed previously generated *C. teleta* bulk RNA-seq, ATAC-seq, and EM-seq datasets (Martín-Zamora et al. 2023, Guynes et al. 2024, Liang et al. 2025) using our new chromosome-level genome assembly and annotations (Fig. 2a). The high-resolution time course of bulk RNA-seq samples captured the expression of 34,333 transcripts grouped into ten stage-

specific clusters of temporally coregulated genes (Fig. 2b and c) that are enriched in GO and KEGG terms consistent with the main developmental processes occurring at their respective life cycle stages (Figs. S9 and S10). The chromosome-level assembly also improved read mapping for both ATAC-seq and EM-seq datasets, capturing an increased number of reproducible open chromatin regions across the five highly correlated ATAC-seq data points (44,368 vs. 45,897 peaks in the old and new genome assemblies, respectively) (Fig. 2d and e) (Martín-Zamora et al. 2023). Consistent with previous analyses (Martín-Zamora et al. 2023), these ATAC-seq peaks are distributed relatively evenly across the genome (Fig. 2e), become more abundant as development progresses (Fig. 2f), and exhibit dynamic accessibility during *C. teleta* embryogenesis (Fig. 2g and h). Finally, the increased number of mapped EM-seq reads considerably improved the detection of 5mC levels, supporting the erosion of the global DNA methylome in the adult (Fig. 2i) (Guynes et al. 2024). Taken together, our analyses reinforce the benefits of a chromosome-level assembly in epigenomic studies.

Single-cell RNA-seq (scRNA-seq) Analyses

scRNA-seq has been applied to *C. teleta* to explore cell-type composition in early-stage larvae (Sur and Meyer 2021). To test whether the new assembly improved downstream scRNAseq computational analysis, we combined and re-analyzed the available datasets, revealing an increase in the number of cells captured and genes detected, but a lower read depth per cell and a lower median number of UMIs/cell when using the new annotation compared to the old assembly (Fig. 2j; Fig. S11a–c). We identified nine clusters from 12,745 cells and 11 clusters from 17,115 cells when mapping to the old and new annotations, respectively (Fig. 2j; Figs. S12 to S22). Notably, fewer technical clusters (ie lacking cluster-specific genes) were identified using the chromosome-level assembly (Fig. 2j; Fig. S11d). Among the clusters that could be unambiguously matched between the old and new analyses, we identified genes that were not detected or exhibited noisy expression in the old dataset but showed cell-type-specific expression in the new one (Fig. S11e). Indeed, the new scRNA-seq analysis identified novel cell-type markers and previously undescribed cellular subpopulations among mesodermal and ectodermal cells. For example, *hyd1n*, a component of motile cilia, is a new marker of the ciliary band cluster and is expressed in the larval ciliary bands (Fig. 2k; Fig. S23a). Cells expressing *abcc9*, a marker of mammalian mesodermal pericytes, are also expressed in the larval mesoderm of *C. teleta* (Fig. 2k; Fig. S23b). Likewise, the new dataset captured four

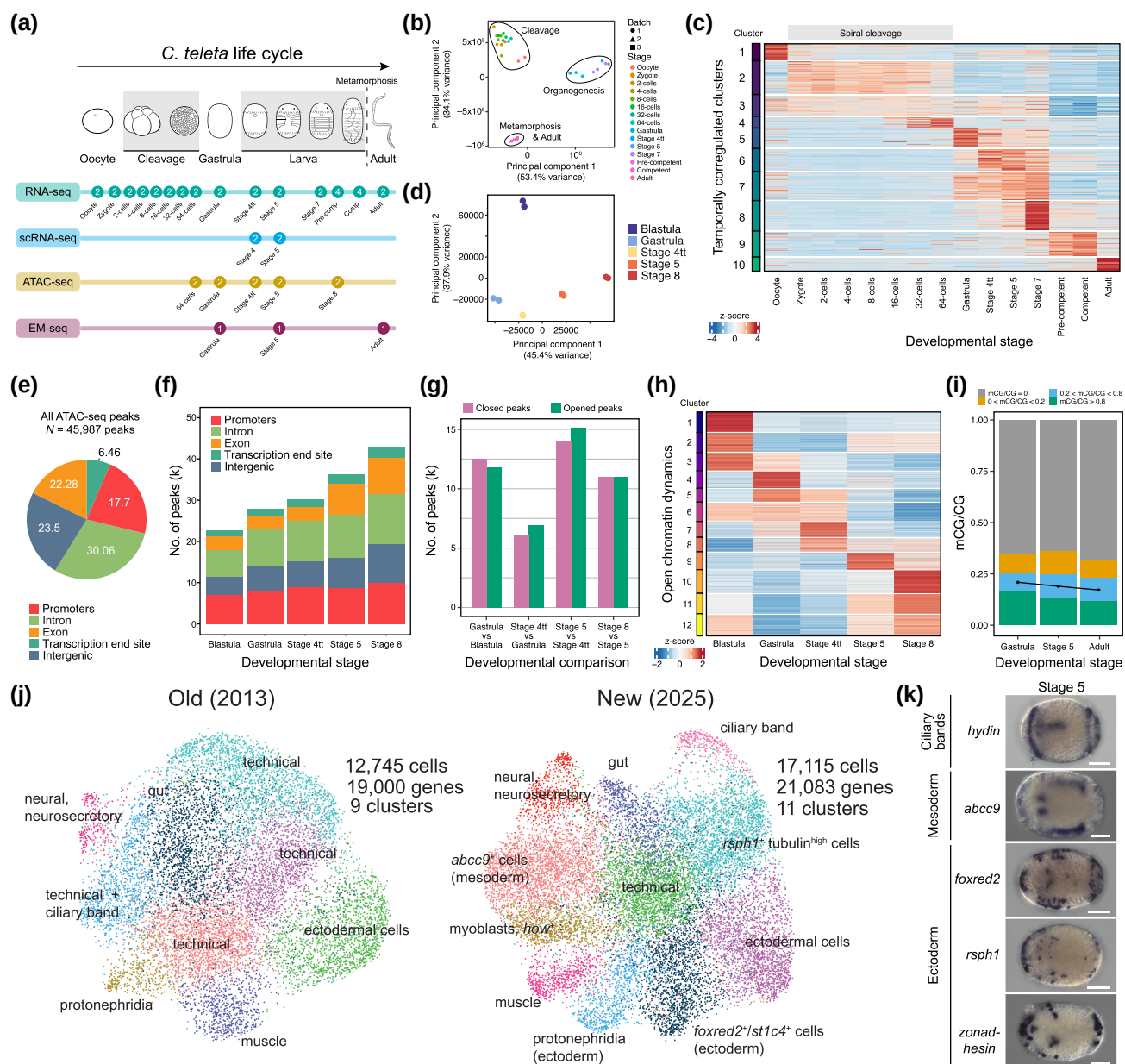


Fig. 2. Multi-omic profiling reveals the regulatory dynamics of *C. teleta* development. **a)** Schematic of the *C. teleta* life cycle, with timelines indicating the stages sampled for RNA-seq, single-cell RNA-seq (scRNA-seq), ATAC-seq, and Enzymatic Methyl-seq (EM-seq). **b)** A Principal Component Analysis (PCA) of the bulk RNA-seq time course reveals three large transcriptomic groups as the life cycle progresses. **c)** A heatmap of soft *k*-means clustered bulk RNA-seq data reveals clusters of genes with coordinated expression patterns, indicating distinct temporal waves of gene activation and repression throughout the life cycle. **d)** PCA of bulk ATAC-seq samples confirms that each developmental stage has a unique chromatin accessibility profile. **e)** Consensus ATAC-seq peaks distribution is broadly similar across intergenic, intronic, exonic, and promoter regions. **f)** The total number of accessible peaks increases as development progresses, while the genomic annotation of these peaks remains constant. **g)** Bar plots showing dynamic opening and closing of chromatin, with the largest reshaping of the regulatory landscape happening between stages 4tt and 5, at the beginning of organogenesis. **h)** A heatmap of soft *k*-means clustered peaks based on their accessibility reveals groups of temporally coregulated regulatory elements. **i)** Bar plots indicating the proportions of DNA methylation levels at three time points of the life cycle, supporting global demethylation in adulthood. **j)** UMAP visualizations of single-cell RNA-seq data from stage 4 and 5 larvae, analyzed with the old (Simakov et al. 2013) and new (this study) genome annotations. The new annotation increases the number of identified cells (from 12,745 to 17,115) and resolves additional distinct cell clusters (from 9 to 11), including the separation of ectodermal and mesodermal subpopulations. **k)** Whole mount in situ hybridization of the cell-type-specific markers *hyd1n*, *abcc9*, *foxred2*, *rsph1* and *zonadhesin* in stage 5. Larval images are shown in ventral view with anterior to the left. Scale bars are 50 μ m.

transcriptionally distinct ectodermal subpopulations, as confirmed by whole-mount in situ hybridization of the flavoprotein *foxred2*, the axonemal component *rsph1*, and the adhesion protein *zonadhesin* (Fig. 2k; Fig. S23c–e). Therefore, the new chromosome-level assembly and updated annotations detect previously missed cell type specific markers, thereby improving cell type identification.

A New Genome Browser

We have developed the Capitella Genome Project Portal (<https://research.nhgri.nih.gov/capitella>), a centralized resource for genome annotations and additional biological information, featuring an interactive browser that allows users to visualize genomic data, and a custom interface for performing BLAST queries.

Materials and Methods

Genome Sequencing and Annotation

gDNA was extracted from snap-frozen adult *C. teleta* using the MagAttract HMW DNA Kit (Qiagen, 67563) according to the manufacturer's recommendations with slight modifications (Table S1). Nanopore sequencing was performed on r9.4.1 and r10.4.1 pores (Table S1), and base-called with Guppy in super accurate (SUP) mode. Passed reads were merged and filtered with Nanofilt (v.2.8.0) (De Coster et al. 2018) to select reads greater than 3.8 kb and assembled with Canu (v2.2) (Koren et al. 2017) using “batOptions=-dg 3 -db 3 -dr 1 -ca 500 -cp 50”. The assembly was polished using long reads greater than 1 kb and short Illumina reads with HyPo (v.1.0.3) (Kundu et al. 2019). Purge Haplotigs (v.1.1.2) (Roach et al. 2018) with parameters “-l 65, -m 145, -h 300” was used to reconstruct a high-quality haploid reference assembly, which was decontaminated with BlobTools (v1.1) (Laetsch and Blaxter 2017). BUSCO (v.5.4.4) and QUAST (v.4.6.1) were used to assess assembly completeness and quality (Gurevich et al. 2013; Simão et al. 2015). For chromosome-scale scaffolding, ~60 M Hi-C reads from a single individual using the Proximo Hi-C (Phase Genomics) kit were mapped with BWA-MEM (v.0.7.17) and processed with YaHS (v.1.2a) (Zhou et al. 2023). Repeat elements and protein-coding genes were annotated following an established pipeline (Martín-Zamora et al. 2023). To assess the completeness of this annotation, we ran BUSCO (v.5.4.4) (Simão et al. 2015) in proteome mode and used Trinotate (v.4.0.0) (Bryant et al. 2017) and PANTHER (v.2.2) (Thomas et al. 2003) to infer functional annotations.

Gene Family Evolution

Orthogroups were reconstructed from non-redundant proteomes using OrthoFinder (v.2.5.5) (Emms and

Kelly 2019), MMseqs2 (v15.6f452) (Steinegger and Söding 2017) and IQ-TREE (v1.6.12) (Minh et al. 2020). Published code (Martin-Duran et al. 2021) was used to infer gene family gains and losses, and gene expansions. Gene Orthology enrichment was performed with topGO (v.2.60.1) (Alexa and Rahnenfuhrer 2025) and a Fisher's exact test. Ancestral linkage groups were inferred from de novo annotations with MetaEuk (v.6-a5d39d9) (Levy Karin et al. 2020) and rbhXpress (v.1.2.3), and plotted with macrosyntR (v.0.2.19) (El Hilali and Copley 2023).

Mitochondrial Genome

Capitella teleta's mitochondrial genome was obtained by BLASTN searches of *Notomastus* sp. (NCBI LC661358.1) mitochondrial genome against a de novo assembled transcriptome (Martín-Zamora et al. 2023), annotated with MITOS 2 (Bernt et al. 2013) with default settings and the invertebrate mitochondrial genetic code (option 5), and manually curated with ORFfinder. To reconstruct a *cox1* gene phylogeny, sequences were aligned with Clustal Omega (Sievers et al. 2011), trimmed using TrimAl (v1.5.0) (Capella-Gutiérrez et al. 2009), and a maximum-likelihood analysis was performed using IQ-TREE2 (v2.3.6) (Minh et al. 2020) and 5000 ultrafast bootstrap replications. The consensus tree was viewed and edited using FigTree v1.4.4.

Functional Genomics

Bulk RNA-seq datasets were batch corrected and normalized using SVA (v. 3.46.0) (Leek et al. 2012), Limma (v.3.54.2) (Ritchie et al. 2015) and DESeq2 (v.1.38.3) (Love et al. 2014). Mfuzz (v.2.58.0) (Kumar and Futschik 2007) was used to cluster the data into ten clusters (optimal number inferred via the elbow method). Transcripts with no expression at any developmental stage (346 out of 34,679) were discarded. ATAC-seq data were mapped with NextGenMap (v.0.5.5) (Sedlazeck et al. 2013), and peaks were called, clustered, and annotated as previously described (Martín-Zamora et al. 2023). EM-seq reads were quality filtered and aligned using Bismark (v0.24.0) (Krueger and Andrews 2011). Methylation was called using the bismark_methylation_extractor tool, and downstream methylation analyses were performed with custom R scripts (Guynes et al. 2024).

Single-Cell RNA-seq Analyses

Published datasets (Sur and Meyer 2021) were processed using 10× Genomics Cell Ranger (v8.0.0) (Zheng et al. 2017) with standard parameters. Single-cell transcriptomes were analyzed using Seurat (v5.1.0) (Hao et al. 2024). Cells that had a low (≤ 300) or high ($\geq 1,500$) number of detected features and

UMIs were removed. Expression was then log-normalized and scaled, and the top 2,000 variable genes were used for visualization. A Uniform Manifold Approximation and Projection (UMAP) was calculated using 25 nearest neighbors and the 14 most significant PCs. For clustering, we used the Leiden approach on the top 14 PCs. Marker genes for individual clusters were identified using ROC and Wilcoxon Rank Sum Tests, and subclusters with fewer than two differentially expressed genes were merged.

Gene Expression Analyses

Colorimetric whole-mount in situ hybridization on larvae was performed using an established protocol (Carrillo-Baltodano and Meyer 2017). Representative larvae were imaged using a Leica DMRA2 upright microscope with an Infinity5 camera (Lumenera), and focus-stacking software (Helicon Focus Lite) was used to combine the stack into a single, fully focused image.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank past and present members of the Martín-Durán lab for their help and contributions to establishing *C. teleta* in the lab, as well as the technical support staff at the Department of Biology at Queen Mary University of London. We also thank Philipp Rescheneder, Sean McKenzie, and Vania Costa from Oxford Nanopore Technologies for their assistance with long-read sequencing in *C. teleta*. This study utilized computational resources from the High-Performance Computing Facility Apocrita at Queen Mary University of London.

Author Contributions

E.C.S., A.D.B., and J.M.M.D. designed and conceived the study. B.E.D., E.C.S., and P.G. conducted data generation. B.E.D., P.G., and J.M.M.D. conducted genome assembly and annotation. B.E.D., J.W., and J.M.M.D. performed genome assembly comparisons, gene family evolutionary analyses, and macrosynteny reconstruction. T.F. and J.W. assembled, annotated, and analyzed the mitochondrial genome. B.E.D. and R.D.D. performed bulk transcriptomic analyses. Y.L. and J.W. analyzed ATAC-seq data, and K.G. analyzed EM-seq data. A.S. performed scRNA-seq analyses, and A.M.C.B. and J.M. performed gene expression validation. R.T.M., S.S., S.Z., R.Y., T.G.W., and A.D.B. conceptualized and developed

the *Capitella* Genome Project Portal. All authors contributed to drafting the manuscript.

Funding

This work was funded by grants to J.M.M.-D. from the European Union Horizon 2020 Framework Programme under the European Research Council (ERC Starting Grant agreement number 801669), the Biotechnology and Biological Sciences Research Council (BB/T008709/1 and BB/Y004221/1) and the Leverhulme Trust (RPG-2023-121), as well as the China Scholarship Council doctoral fellowship to J.W. (CSC NO. 202306330025). This research was supported in part by the Intramural Research Program of the National Institutes of Health [ZIA HG000140 to A.D.B.]. The contributions of the NIH authors are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Newly generated data have been deposited in DDBJ/ENA/GenBank under the project accession PRJEB101787. DNA and protein sequence downloads, functional annotation of gene models, and multi-omic data are available on the *Capitella* Genome Project portal (<https://research.nhgri.nih.gov/capitella>).

Literature Cited

- Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package. 2025.
- Bernt M, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013;69:313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>.
- Blake JA, Grassle JP, Eckelbarger KJ. *Capitella teleta*, a new species designation for the opportunistic and experimental *Capitella* sp I, with a review of the literature for confirmed records. *Zoosymposia*. 2009;2:25–53. <https://doi.org/10.11646/zoosymposia.2.1.6>.
- Bryant DM, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep*. 2017;18:762–776. <https://doi.org/10.1016/j.celrep.2016.12.063>.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
- Carrillo-Baltodano AM, Meyer NP. Decoupling brain from nerve cord development in the annelid *Capitella teleta*: insights into

- the evolution of nervous systems. *Dev Biol.* 2017;431:134–144. <https://doi.org/10.1016/j.ydbio.2017.09.022>.
- De Coster W, D'hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- El Hilali S, Copley RR. 2023. macrosyntr: drawing automatically ordered Oxford grids from standard genomic files in R [preprint]. *bioRxiv* 525673. <https://doi.org/10.1101/2023.01.26.525673>.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Grassle J, Grassle JF. Sibling species in the marine pollution indicator *Capitella* (polychaeta). *Science.* 1976;192:567–569. <https://doi.org/10.1126/science.1257794>.
- Grassle JP, Gelfman CE, Mills SW. Karyotypes of *Capitella* sibling species, and several species in the related genera *Capitellides* and *Capitomastus* (Polychaeta). *Bull Biol Soc Wash.* 1987;77:88.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Guynes K, et al. Annelid methylomes reveal ancestral developmental and aging-associated epigenetic erosion across Bilateria. *Genome Biol.* 2024;25:204. <https://doi.org/10.1186/s13059-024-03346-z>.
- Hao Y, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol.* 2024;42:293–304. <https://doi.org/10.1038/s41587-023-01767-y>.
- Klann M, Seaver EC. Functional role of pax6 during eye and nervous system development in the annelid *Capitella teleta*. *Dev Biol.* 2019;456:86–103. <https://doi.org/10.1016/j.ydbio.2019.08.011>.
- Koren S, et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 2017;27:722–736. <https://doi.org/10.1101/gr.215087.116>.
- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics.* 2011;27:1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.
- Kumar L, Futschik ME. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation.* 2007;2:5–7. <https://doi.org/10.6026/97320630002005>.
- Kundu R, Casey J, Sung W-K. 2019. Hypo: super fast & accurate polisher for long read genome assemblies [preprint]. *bioRxiv* 882506. <https://doi.org/10.1101/2019.12.19.882506>.
- Laetsch D, Blaxter M. BlobTools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Res.* 2017;6:1287. <https://doi.org/10.12688/f1000research.12232.1>.
- Lanza AR, Seaver EC. Functional evidence that Activin/Nodal signaling is required for establishing the dorsal-ventral axis in the annelid *Capitella teleta*. *Development.* 2020;147:dev189373. <https://doi.org/10.1242/dev.189373>.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28:882–883. <https://doi.org/10.1093/bioinformatics/bts034>.
- Levy Karin E, Mirdita M, Soding J. MetaEuk-sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome.* 2020;8:48. <https://doi.org/10.1186/s40168-020-00808-x>.
- Lewin TD, Liao JJ, Luo YJ. Annelid comparative genomics and the evolution of massive lineage-specific genome rearrangement in Bilaterians. *Mol Biol Evol.* 2024;41:msae172. <https://doi.org/10.1093/molbev/msae172>.
- Liang Y, Wei J, Kang Y, Carrillo-Baltodano AM, Martín-Durán JM. Cell fate specification modes shape transcriptome evolution in the highly conserved spiral cleavage. *EMBO Rep.* 2025;26:5088. <https://doi.org/10.1038/s44319-025-00569-4>.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Martin-Duran JM, et al. Conservative route to genome compaction in a miniature annelid. *Nat Ecol Evol.* 2021;5:231–242. <https://doi.org/10.1038/s41559-020-01327-6>.
- Martín-Zamora FM, et al. Annelid functional genomics reveal the origins of bilaterian life cycles. *Nature.* 2023;615:105–110. <https://doi.org/10.1038/s41586-022-05636-7>.
- Meyer NP, Boyle MJ, Martindale MQ, Seaver EC. A comprehensive fate map by intracellular injection of identified blastomeres in the marine polychaete *Capitella teleta*. *Evodevo.* 2010;1:8. <https://doi.org/10.1186/2041-9139-1-8>.
- Meyer NP, Seaver EC. Cell lineage and fate map of the primary somatoblast of the polychaete annelid *Capitella teleta*. *Integr Comp Biol.* 2010;50:756–767. <https://doi.org/10.1093/icb/icq120>.
- Minh BQ, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Nath R, Panda B, Rakesh S, Krishnan A. Lineage-specific class-A GPCR dynamics reflect diverse chemosensory adaptations in lophotrochozoa. *Mol Biol Evol.* 2025;42:msaf042. <https://doi.org/10.1093/molbev/msaf042>.
- Neal S, De Jong DM, Seaver EC. CRISPR/CAS9 mutagenesis of a single *r-opsin* gene blocks phototaxis in a marine larva. *Proc Biol Sci.* 2019;286:20182491. <https://doi.org/10.1098/rspb.2018.2491>.
- Ritchie ME, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47. <https://doi.org/10.1093/nar/gkv007>.
- Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics.* 2018;19:460. <https://doi.org/10.1186/s12859-018-2485-7>.
- Seaver EC. Annelid models I: *Capitella teleta*. *Curr Opin Genet Dev.* 2016;39:35–41. <https://doi.org/10.1016/j.gde.2016.05.025>.
- Seaver EC. Sifting through the mud: a tale of building the annelid *Capitella teleta* for EvoDevo studies. *Curr Top Dev Biol.* 2022;147:401–432. <https://doi.org/10.1016/bs.ctdb.2021.12.018>.
- Seaver EC, De Jong DM. Regeneration in the segmented Annelid *Capitella teleta*. *Genes (Basel).* 2021;12:1769. <https://doi.org/10.3390/genes12111769>.
- Sedlazeck FJ, Rescheneder P, Von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics.* 2013;29:2790–2791. <https://doi.org/10.1093/bioinformatics/btt468>.
- Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539. <https://doi.org/10.1038/msb.2011.75>.
- Simakov O, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature.* 2013;493:526–531. <https://doi.org/10.1038/nature11696>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat*

- Biotechnol. 2017;35:1026–1028. <https://doi.org/10.1038/nbt.3988>.
- Struck TH, Golombek A, Hoesel C, Dimitrov D, Elgetany AH. Mitochondrial genome evolution in Annelida-A systematic study on conservative and variable gene orders and the factors influencing its evolution. Syst Biol. 2023;72:925–945. <https://doi.org/10.1093/sysbio/syad023>.
- Sur A, Meyer NP. Resolving transcriptional states and predicting lineages in the annelid *Capitella teleta* using single-cell RNAseq. Front Ecol Evol. 2021;8:2020. <https://doi.org/10.3389/fevo.2020.618007>.
- Thomas PD, et al. PANTHER: a library of protein families and sub-families indexed by function. Genome Res. 2003;13: 2129–2141. <https://doi.org/10.1101/gr.772403>.
- Weigert A, et al. Illuminating the base of the annelid tree using transcriptomics. Mol Biol Evol. 2014;31:1391–1401. <https://doi.org/10.1093/molbev/msu080>.
- Weigert A, et al. Evolution of mitochondrial gene order in Annelida. Mol Phylogenet Evol. 2016;94:196–206. <https://doi.org/10.1016/j.ympev.2015.08.008>.
- Zheng GX, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049. <https://doi.org/10.1038/ncomms14049>.
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023;39:btac808. <https://doi.org/10.1093/bioinformatics/btac808>.
- Associate editor:** John Wang