G3, 2023, 13(2), jkac336
https://doi.org/10.1093/g3journal/jkac336
Advance Access Publication Date: 21 December 2022
Genome Report

Genome assembly of the acoel flatworm *Symsagittifera roscoffensis*, a model for research on body plan evolution and photosymbiosis

Pedro Martinez [1,2,*] Kirill Ustyantsev,[3] Mikhail Biryukov [4] Stijn Mouton,[3] Liza Glasenburg,[3] Simon G. Sprecher,[5] Xavier Bailly,[6] Eugene Berezikov [3,*]

[1]Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain
[2]Institut Català de Recerca i Estudis Avançats (ICREA), Barcelona 08193, Spain
[3]European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen 9700AD, The Netherlands
[4]Institute of Cytology and Genetics SB RAS, Novosibirsk 630090, Russia
[5]Department of Biology, University of Fribourg, Chemin du Musee 10, 1700 Fribourg, Switzerland
[6]Station Biologique de Roscoff, Multicellular Marine Models (M3) team, FR2424, CNRS/Sorbonne Université—Place Georges Teissier, 29680 Roscoff, France

*Corresponding author: Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain. Email: pedro.martinez@ub.edu; *Corresponding author: European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen 9700AD, The Netherlands. Email: e.berezikov@umcg.nl

Abstract

Symsagittifera roscoffensis is a well-known member of the order Acoela that lives in symbiosis with the algae *Tetraselmis convolutae* during its adult stage. Its natural habitat is the eastern coast of the Atlantic, where at specific locations thousands of individuals can be found, mostly, lying in large pools on the surface of sand at low tide. As a member of the Acoela it has been thought as a proxy for ancestral bilaterian animals; however, its phylogenetic position remains still debated. In order to understand the basic structural characteristics of the acoel genome, we sequenced and assembled the genome of aposymbiotic species *S. roscoffensis*. The size of *this* genome was measured to be in the range of 910–940 Mb. Sequencing of the genome was performed using PacBio Hi-Fi technology. Hi-C and RNA-seq data were also generated to scaffold and annotate it. The resulting assembly is 1.1 Gb large (covering 118% of the estimated genome size) and highly continuous, with N50 scaffold size of 1.04 Mb. The repetitive fraction of the genome is 61%, of which 85% (half of the genome) are LTR retrotransposons. Genome-guided transcriptome assembly identified 34,493 genes, of which 29,351 are protein coding (BUSCO score 97.6%), and 30.2% of genes are spliced leader trans-spliced. The completeness of this genome suggests that it can be used extensively to characterize gene families and conduct accurate phylogenomic reconstructions.

Keywords: Acoela, Xenacoelomorpha, symbiogenesis, photosymbiosis, genome evolution

Introduction

Acoel flatworms (order Acoela) are members of the phylum Xenacoelomorpha, which also include the clades Nemertodermatida and Xenoturbellida (Philippe et al. 2011). The acoels are represented by approximately 400 described species, almost all of which are marine (Hejnol et al. 2009; Achatz et al. 2013). They exhibit a remarkable anatomical diversity, with many having salient characteristics such as an association with photosymbionts or extensive regenerative abilities. *Symsagittifera roscoffensis*, a species with the aforementioned properties, is 1 of the best-studied species of the Acoela (Fig. 1). Photosymbiotic adults are abundant along most of the Atlantic coast of Europe (from Wales to Gibraltar), easy to collect, and live in an obligatory relationship with the algae *Tetraselmis convolutae* (Bailly et al. 2014). As a member of the Acoela, a lineage considered to be an early offshoot of the Bilateria (but see Cannon et al. 2016; Philippe et al. 2019 for alternative views), it has also been used as models of ancestral bilaterians. Moreover, understanding the genomic characteristics of *S. roscoffensis* is of special relevance, as it can shed light on, for instance, symbiogenesis, regenerative processes, and, of course, the phylogenetic position of the Acoela.

In the recent past, we and others generated the first draft genome of *S. roscoffensis* (Philippe et al. 2019) but most contigs were extremely small (N50 < 5 kb). Here we present a new draft of the *S. roscoffensis* genome, generated using PacBio Hi-Fi technology and scaffolded with Hi-C data, which significantly increased genome assembly continuity (scaffold N50 = 1.04 Mb).

Methods and materials

Preparation of aposymbiotic animals

Aposymbiotic animals were used as a source of genomic DNA in order to avoid sequences of microalgae symbionts. Gravid animals were collected at low tide from beaches in the areas of Roscoff and Carantec, Brittany, France and were transported to

Received: October 26, 2022. **Accepted:** December 17, 2022

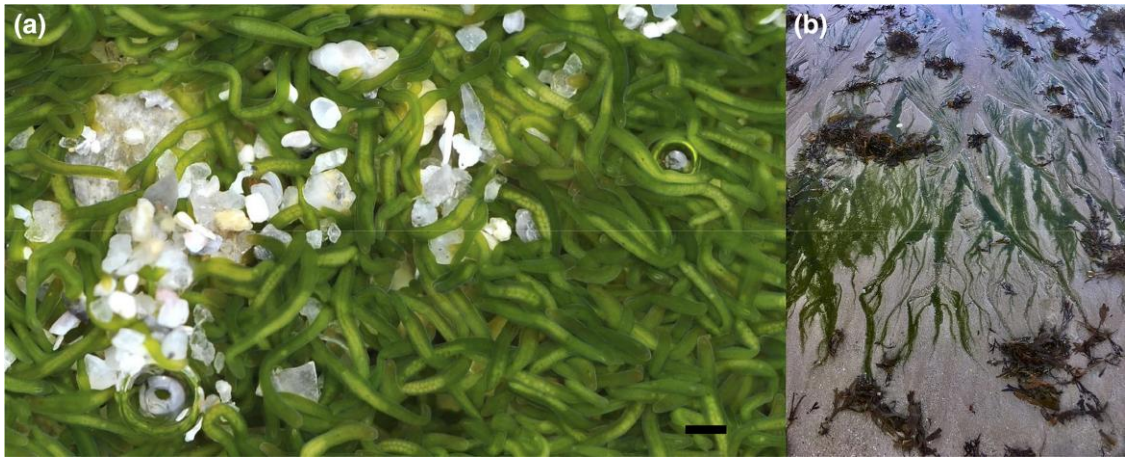

Fig. 1. Specimens of the acoel *S. roscoffensis* in their natural habitat. a) Adult (gravid) *S. roscoffensis*. Credit Wilfried Thomas/Station Biologique de Roscoff. b) Biotope. Pools of adult specimens at low tide in a Brittany beach (France). Scale bar in panel a) is 1 mm.

laboratory, where most of them spontaneously spawned. Hatched juveniles were stored in RNAlater or snap-frozen in liquid nitrogen.

Genome size measurement

The genome size was determined by a flow cytometry approach (Hare and Johnston 2011) as previously described for *Macrostomum lignano* flatworms (Wudarski et al. 2017). Nuclei were isolated from symbiotic adults and from juveniles without symbionts. The *M. lignano* NL12 line (Wudarski et al. 2017) and human fibroblasts were used as references. Nuclei were stained with propidium iodide and fluorescence was measured on a BD FacsCanto II Cell Analyzer.

Pacific biosciences Hi-Fi genome sequencing

Genomic DNA was extracted using MagAttract High Molecular Weight DNA extraction kit. The preparation and sequencing of the library was performed by GenomeScan B.V. (Leiden, The Netherlands) using Sequel II Sequencing Kit 2.0 and 8M SMRT Cell. Reads were processed with the ccs tool v.6.2.0 and filtered using HiFiAdapterFilt v. 2.0.0 (Sim et al. 2022).

Hi-C library construction and sequencing

The preparation and sequencing of Hi-C library was performed by Arima Genomics (San Diego, USA) using snap-frozen animals. The library was prepared using the Arima-HiC+ kit and the Arima Library Prep Module and sequenced on an Illumina HiSeq X instrument.

RNA library construction and sequencing

RNA was isolated with Qiagen RNeasy Micro Kit according to the manufacturer's protocol, except that the DNase I treatment step was omitted. The RNA-seq library was constructed according to Smart-3SEQ protocol (Foley et al. 2019) and sequenced on an Illumina NextSeq 500 instrument.

De novo transcriptome assembly

De novo transcriptome assembly SYMROS200831 was generated using a ReCAP pipeline (Grudniewska et al. 2016) from public whole transcript RNA-seq data (SRR5760179 and SRR8506641). Reads were normalized to 30× coverage and assembled into contigs using Trinity v.2.11.0 (Grabherr et al. 2011), remapped using Bowtie v.2.3.4.3 (Langmead and Salzberg 2012) and reassembled using CAP3 v. 12/21/07 (Huang and Madan 1999).

Genome assembly and evaluation

PacBio Hi-Fi data were assembled with FALCON/FALCON-Unzip v.1.8.1/v.1.3.7 (Chin et al. 2016), Flye 2.9 (Kolmogorov et al. 2019), HiCanu v.2.2 (Nurk et al. 2020), Hifiasm v.0.16.1 (Cheng et al. 2021), IPA v.1.8.0 (Sovic 2022), Peregrine v.0.1.5.3 (Chin and Khalak 2019), Raven v.1.8.1 (Vaser and Šikić 2021), and wtdbg2 v.2.5 (Ruan and Li 2020); using parameters default for each assembler, and deduplicated by purge_dups v.1.2.5 (Guan et al. 2020). The quality of the assemblies was evaluated by mapping de novo transcriptome assembly SYMROS200831 to the genome assemblies with GMAP v.2021-08-25 (Wu and Watanabe 2005) and calculating the fraction of transcripts that map to a given assembly and the fraction of eukaryotic BUSCO models (v.2.0) present (Simão et al. 2015).

Hi-C reads were mapped to the deduplicated peregrine assembly by BWA v.0.7.17-r1188 (Li and Durbin 2009) and processed by the Arima Genomics mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline). Scaffolding was performed by SALSA2 v.2.3 (Ghurye et al. 2019). Paired-end RNA-seq reads (SRR5760179 and SRR8506641) were mapped to the scaffolds with HISAT v.2.2.1 (Kim et al. 2019) and further scaffolding was performed by P_RNA_scaffolder (Zhu et al. 2018). Gap closing was performed with LR_gapcloser v.1.1 (Xu et al. 2019) using initial PacBio Hi-Fi reads. Final polishing was done by pilon v.1.24 (Walker et al. 2014) with RNA-seq reads to fix frameshifts in coding sequences.

Mitochondrial genome assembly and annotation

The mitochondrial genome was reconstructed by performing tblastn (Camacho et al. 2009) searches in genome assemblies generated by different assemblers using 11 protein-coding sequences from the *Isodiametra pulchra* mitochondrial genome (NC_034948.1). Two, 99.5% identical, candidates contigs were identified and further analyzed by self-blast. A single 100% identical terminal repeat (length 535) was identified and used for circularization. The resulting mitochondrial genome was annotated using MITOS2 web server (Donath et al. 2019).

Repeat analysis

Tandem repeats were annotated with Tandem Repeat Finder 4.10.0 (Benson 1999). A de novo library of classified repetitive element models was created using RepeatModeler 2.0.3 (Flynn et al. 2020). Homology-based annotation of long terminal repeat (LTR) retrotransposons was done using the Domain-Associated Retrotransposon Search (DARTS) algorithm (Biryukov and Ustyantsev 2022). RepeatModeler- and LTR retrotransposon-derived libraries were merged and used as input for RepeatMasker 4.1.2-p1 (Tempel 2012).

Gene prediction and annotation

Gene annotation was performed using TBONE pipeline (Wudarski et al. 2017). RNA-seq data SRR5760179 and SRR8506641 and Smart-3SEQ data generated in this study were mapped to the genome assembly with HISAT v.2.2.1 (Kim et al. 2019) and initial gene models were constructed by Scallop v0.10.5 (Shao and Kingsford 2017) and StringTie v2.2.1 (Kovaka et al. 2019). De novo transcriptome assembly SYMROS200831 was mapped to the genome assembly with GMAP v.2021-08-25 (Wu and Watanabe 2005). All gene models were merged by gffread v.0.12.7 (Pertea and Pertea 2020) and further processed to identify trans-spliced genes.

Trans-splicing leader sequence was determined through the analysis and mapping of the de novo transcriptome assembly SYMROS200831 to the genome assembly. Transcripts, in which first 10–50 nucleotides are not mapped to the genome, were identified and the respective nonmapped 5′ sequences were extracted from the transcripts. The most abundant sequences were manually aligned to each other and the trans-splicing leader sequence GCCTAATTGTTGTGATAAACTTATTAAATAGA was reconstructed. The structure of the spliced leader (SL) RNA gene was determined by mapping the SL sequence to the genome assembly using blastn (Camacho et al. 2009) and examining matching genomic regions for canonical SL RNA folding using RNAfold web server (Gruber et al. 2008).

Reads containing trans-splicing sequence were extracted from RNA-seq data, trimmed and mapped to the genome assembly with HISAT v.2.2.1 (Kim et al. 2019). The resulting wiggle files were used to identify genomic peaks corresponding to trans-splicing locations. Similarly, peaks corresponding to polyadenylation sites at 3′ end gene boundaries were identified by mapping reads from Smart-3SEQ RNA-seq libraries. The generated trans-splicing and polyadenylation signals were used to refine gene boundaries and separate trans-spliced genes. Open reading frames were predicted by TransDecoder v.5.5.0 (Haas et al. 2013). To remove redundancy, for each genomic locus a single representative transcript was selected and included into a subset called "core genes."

Results and discussion

Genome assembly and evaluation

In order to distinguish the genomes of *S. roscoffensis* and its microsymbiotic algae, we used cultured aposymbiotic juveniles without microsymbionts. Using a flow cytometry approach, the genome size of *S. roscoffensis* is estimated to be in the range of 910–940 Mb (Supplementary Fig. 1). We sequenced the genome to 20× coverage with Pacific Biosciences Hi Fidelity reads (1.29 mln ccs reads, mean length 15.7 kb) and assembled the data with 8 different genome assemblers (Supplementary Table 1). For the evaluation of the assemblies, we examined: the assembly size, N50 contig length, the fraction of transcripts from de novo transcriptome assembly mapping to the genome, and the number of gene models from the Eukaryotic BUSCO subset identified. We focused specifically on the Eukaryotic BUSCOs because almost all genes from this subset are expected to be present in an Acoel genome. Since genome sequencing was performed on a population of animals obtained directly from a natural habitat, it is expected that a substantial level of heterozygosity is present in the sequencing data, leading to large assemblies with under-collapsed heterozygous regions. Indeed, FALCON, Flye, Hifiasm, IPA, and Peregrine produced redundant assemblies of 1.2–2.5 Gb in size (Supplementary Table 1). HiCanu and Raven over-collapsed the assemblies (737 and 555 Mb), while Wtdbg2 generated an assembly closest in size to the measured genome size (935 Mb). Wtdbg2 also produced the highest N50 length (140.3 kb) among the tested assemblers. Moreover, the fraction of de novo transcripts mapped and BUSCO models identified was the highest for the Peregrine assembly (Supplementary Table 1). After deduplication of the redundant assemblies with purge_dups (Guan et al. 2020), the Peregrine assembly appeared to be substantially better compared to all other tested assemblies, producing a N50 size of 197.6 kb and the lowest fraction of missing de novo and BUSCO transcripts, while its assembly size of 1.1 Gb is only ~18% larger than the measured genome size (Supplementary Table 1). Therefore, Peregrine assembly was used for further scaffolding and gap closing. We deliberately choose an under-collapsed rather than an over-collapsed assembly in order to maximally retain gene content, although this means that the assembly does contain some regions that represent diverged alleles and not true genomic duplications.

Table 1. Characteristics of genome assembly SymRos_1_5.

	Contigs	Scaffolds
Total number	7,843	3,460
Total length (bp)	1,101,399,379	1,103,025,803
Average length (bp)	140,431	318,794
Shortest (bp)	12,747	12,747
Longest (bp)	1,836,468	8,003,794
N50 (bp)	237,875	1,039,899
L50	1,417	287
GC content (%)	36.7	36.7
Coding genes		29,351
Noncoding genes		5,142
Number of SL trans-spliced genes		10,433
Average transcript length (kb)		1.78
Longest transcript (kb)		53.5
Average gene length (kb)		9.83
Average number of introns		3.2
Average intron length (kb)		2.5
Eukaryotic BUSCOs ($n = 303$)		296 (97.6%)
Complete and single-copy BUSCOs		188 (62%)
Complete and duplicated BUSCOs		108 (35.6%)
Fragmented BUSCOs		3 (1.0%)
Missing BUSCOs		4 (1.4%)

For genome scaffolding, we generated 388 mln Illumina read pairs (~100× genome coverage) from a Hi-C library. Scaffolding was performed by SALSA2 (Ghurye et al. 2019), followed by P_RNA_scaffolder (Zhu et al. 2018) with RNA-seq reads, which substantially improved assembly continuity (3,460 scaffolds, N50 = 1039.9 kb, Table 1). Gaps were closed by LR_gapcloser (Xu et al. 2019) followed by assembly polishing with pilon (Walker et al. 2014), reducing the number of contigs from 8,943 to 7,843 and improving N50 contig size from 197.6 to 237.9 kb (Table 1). The mitochondrial genome was reconstructed from PacBio Hi-Fi genome assemblies and is 99% identical to the published sequence (Mwinyi et al. 2010).

Repeat annotation

The genome is highly repetitive, with transposable elements and simple repeats comprising more than 60% of its sequence, of which 85% are LTR retrotransposons (Supplementary Table 2).

Gene annotation

To annotate genes, we used TBONE pipeline (Wudarski et al. 2017), which takes into account potential effects of SL trans-splicing present in flatworms (Ustyantsev and Berezikov 2021). By analyzing de novo transcriptome assembly SYMROS20083, we determined that SL trans-splicing is also present in *S. roscoffensis* (Supplementary Fig. 2). The genome-guided transcriptome assembly SymRos_1_5_RNA.v1 generated by TBONE pipeline contains 34,493 genes, of which 29,351 are protein-coding and 5,142 are noncoding (Table 1). The number of SL trans-spliced genes is 10,433, comprising 30.2% of all genes. The transcriptome assembly contains 296 out of 303 eukaryotic BUSCO gene models, or 97.6% (Table 1).

Preliminary analysis confirmed the presence of gene families (Homeobox classes, bHLHs, GPCRs and Wnts; with their rich complements) described in previous papers (Moreno et al. 2009; Gavilán et al. 2016; Brauchle et al. 2018). This, again, attests the quality and usefulness of the genome assembly. At the same time, the number of duplicated BUSCO genes is quite high at 35.6% (Table 1). Some of these duplications might be true, since partial genome duplications have been reported in flatworms (Zadesenets et al. 2017), but these duplications can be also explained by the fact that we used an under-collapsed assembly for scaffolding. False gene duplications are a known and difficult to address issue in genome assemblies, stemming from the inability of genome assembly algorithms to discriminate between haplotype paralogs and homologs in highly heterozygous regions (Ko et al. 2022). Thus, the *S. roscoffensis* genome assembly reported here should be used with caution when analyzing potential gene family expansions.

Characteristics of the *S. roscoffensis* genome and comparison with other xenacoelomorphs

Here, by combining PacBio Hi-Fi sequencing with Hi-C scaffolding, we generated a highly continuous and complete assembly SymRos_1_5 with N50 scaffold size of 1.04 Mb and BUSCO score of 97.4. The karyotype of *S. roscoffensis* is $2n = 20$ (Moreno et al. 2009). Despite the availability of Hi-C data, the assembly is still far from chromosome-level, which can be attributed to the highly repetitive nature of the genome and high level of heterozygosity in the population of animals used.

Based on flow cytometry data the genome size of *S. roscoffensis* is in the range of 910–940 Mb, which is comparable to that of other acoel genomes, *Hofstenia miamia* (950 Mb) (Gehrke et al. 2019) and *Praesagittifera naikaiensis* (654 Mb) (Arimoto et al. 2019). The assembled genome is larger than the measured genome size, likely due to remaining heterozygous regions not purged from the assembly.

The GC content of *S. roscoffensis* genome is 36.7% (Table 1), thus it is an AT-rich genome, similar to other xenacolomorphs [43% GC content in *Xenoturbella bocki* (Schiffer et al. 2022), 39.1% GC content in *P. naikaiensis* (Arimoto et al. 2019)].

In the assembled genomes of acoels, the content of repetitive sequences is high and varies from 53% in *H. miamia* (Gehrke et al. 2019) to 61% in *S. roscoffensis* and 70% in *P. naikaiensis* (Arimoto et al. 2019), with the major prevalence of LTR retrotransposons in all of them. In contrast, the xenoturbellid *X. bocki* has only 25% of its genome in repeats (Schiffer et al. 2022).

We annotated 34,493 genes in *S. roscoffensis*, which is higher than those reported for *H. miamia* (~22,000) or *X. bocki* (~15,000). This variability can be explained by differences in annotation pipelines used to identify genes, with the TBONE pipeline used here more inclusive for nonconserved, noncoding, repetitive, and low-expressed genes.

Some organisms, including flatworms, have SL trans-splicing, in which sequence from 1 RNA molecule (SL) is spliced to 5′ ends of different mRNAs (Lasda and Blumenthal 2011). We identified that 30.2% of the genes in *S. roscoffensis* undergo such SL trans-splicing (Table 1), which is similar to the number of trans-spliced genes in the flatworm *M. lignano* (Ustyantsev and Berezikov 2021).

Data availability

All raw sequencing data have been deposited in the NCBI Sequence Read Archive (accession codes SRR20990873–SRR20990875) and can be accessed with BioProject No. PRJNA867535. The genome assembly has been deposited at DDBJ/ENA/GenBank under the accession JANVAR000000000. The annotated genome is available at http://gb.macgenome.org.

Supplemental material available at G3 online.

Acknowledgments

We would like to acknowledge Brenda Gavilán and Sergio Melero for helping us with the collection of hatchlings.

Funding

The work in P. Martinez's laboratory was funded by the Ministerio de Ciencia, Innovación y Universidades, Spain (project number: PID2021-124415NB-I00). The work in E. Berezikov's laboratory was supported by the Dutch Research Council Open Competition XS grant (file number OCENW.XS21.2.051). The work of M. Biryukov on repeat annotation was supported by the Russian State Budget project FWNR-2022-0016. S. G. Sprecher was supported by Swiss National Science Foundation grant 310030_188471 and IZCOZ0_182957. X. Bailly was funded by the Functional Genomics Joint Research Activities of the ASSEMBLE Plus Program (Association of European Marine Biological Laboratories Expanded).

Conflicts of interest

None declared.

Literature cited

Achatz JG, Chiodin M, Salvenmoser W, Tyler S, Martinez P. The Acoela: on their kind and kinships, especially with nemertodermatids and xenoturbellids (*Bilateria incertae sedis*). Org Divers Evol. 2013;13(2):267–286. doi:10.1007/s13127-012-0112-4.

Arimoto A, Hikosaka-Katayama T, Hikosaka A, Tagawa K, Inoue T, *et al.* A draft nuclear-genome assembly of the acoel flatworm *Praesagittifera naikaiensis*. Gigascience. 2019;8(4):giz023. doi:10.1093/gigascience/giz023.

Bailly X, Laguerre L, Correc G, Dupont S, Kurth T, *et al.* The chimerical and multifaceted marine acoel *Symsagittifera roscoffensis*: from

photosymbiosis to brain regeneration. Front Microbiol. 2014;5:498. doi:10.3389/fmicb.2014.00498.

Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–580. doi:10.1093/nar/27.2.573.

Biryukov M, Ustyantsev K. DARTS: an algorithm for domain-associated retrotransposon search in genome assemblies. Genes (Basel). 2022;13(1):9. doi:10.3390/genes13010009.

Brauchle M, Bilican A, Eyer C, Bailly X, Martínez P, *et al.* Xenacoelomorpha survey reveals that all 11 animal homeobox gene classes were present in the first bilaterians. Genome Biol Evol. 2018;10(9):2205–2217. doi:10.1093/gbe/evy170.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, *et al.* BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421. doi:10.1186/1471-2105-10-421.

Cannon JT, Vellutini BC, Smith J, Ronquist F, Jondelius U, *et al.* Xenacoelomorpha is the sister group to Nephrozoa. Nature. 2016;530(7588):89–93. doi:10.1038/nature16520.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–175. doi:10.1038/s41592-020-01056-5.

Chin C-S, Khalak A. Human genome assembly in 100 minutes. bioRxiv 705616. https://doi.org/10.1101/705616.

Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016;13(12):1050–1054. doi:10.1038/nmeth.4035.

Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, *et al.* Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019;47-(20):10543–10552. doi:10.1093/nar/gkz833.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, *et al.* Repeatmodeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020;117(17):9451–9457. doi:10.1073/pnas.1921046117.

Foley JW, Zhu C, Jolivet P, Zhu SX, Lu P, *et al.* Gene expression profiling of single cells from archival tissue with laser-capture microdissection and Smart-3SEQ. Genome Res. 2019;29(11):1816–1825. doi:10.1101/gr.234807.118.

Gavilán B, Perea-Atienza E, Martínez P. Xenacoelomorpha: a case of independent nervous system centralization? Philos Trans R Soc B Biol Sci. 2016;371(1685):20150039. doi:10.1098/rstb.2015.0039.

Gehrke AR, Neverett E, Luo YJ, Brandt A, Ricci L, *et al.* Acoel genome reveals the regulatory landscape of whole-body regeneration. Science. 2019;363(6432):eaau6173. doi:10.1126/science.aau6173

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, *et al.* Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 2019;15(8):e1007273. doi:10.1371/journal.pcbi.1007273.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–652. doi:10.1038/nbt.1883.

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA Websuite. Nucleic Acids Res. 2008;36:W70–W74. doi:10.1093/nar/gkn188.

Grudniewska M, Mouton S, Simanov D, Beltman F, Grelling M, *et al.* Transcriptional signatures of somatic neoblasts and germline cells in *Macrostomum lignano*. Elife. 2016;5:e20607. doi:10.7554/eLife.20607.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36(9):2896–2898. doi:10.1093/bioinformatics/btaa025.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–1512. doi:10.1038/nprot.2013.084.

Hare EE, Johnston JS. Genome size determination using flow cytometry of propidium iodide-stained nuclei. Methods Mol Biol. 2011;772:3–12. doi:10.1007/978-1-61779-228-1_1.

Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, *et al.* Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc R Soc B Biol Sci. 2009;276(1677):4261–4270. doi:10.1098/rspb.2009.0896.

Huang X, Madan A. CAP3: a DNA sequence assembly program. Genome Res. 1999;9(9):868–877. doi:10.1101/gr.9.9.868.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–915. doi:10.1038/s41587-019-0201-4.

Ko BJ, Lee C, Kim J, Rhie A, Yoo DA, *et al.* Widespread false gene gains caused by duplication errors in genome assemblies. Genome Biol. 2022;23(1):205. doi:10.1186/s13059-022-02764-1.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–546. doi:10.1038/s41587-019-0072-8.

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20(1):278. doi:10.1186/s13059-019-1910-1.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359. doi:10.1038/nmeth.1923.

Lasda EL, Blumenthal T. Trans-splicing. Wiley Interdiscip Rev RNA. 2011;2(3):417–434. doi:10.1002/wrna.71.

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324.

Moreno E, Nadal M, Baguñà J, Martínez P. Tracking the origins of the bilaterian Hox patterning system: insights from the acoel flatworm *Symsagittifera roscoffensis*. Evol Dev. 2009;11(5):574–581. doi:10.1111/j.1525-142X.2009.00363.x.

Mwinyi A, Bailly X, Bourlat SJ, Jondelius U, Littlewood DTJ, Podsiadlowski L. The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of *Symsagittifera roscoffensis*. BMC Evol Biol. 2010;10(1):309. doi:10.1186/1471-2148-10-309.

Nurk S, Walenz BP, Rhie A, Vollger MR, Logsdon GA, *et al.* Hicanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. 2020;30(9):1291–1305. doi:10.1101/gr.263566.120.

Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Res. 2020;9:304. doi:10.12688/f1000research.23297.1.

Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, *et al.* Acoelomorph flatworms are deuterostomes related to Xenoturbella. Nature. 2011;470(7333):255–258. doi:10.1038/nature09676.

Philippe H, Poustka AJ, Chiodin M, Hoff KJ, Dessimoz C, *et al.* Mitigating anticipated effects of systematic errors supports sister-group relationship between Xenacoelomorpha and Ambulacraria. Curr Biol. 2019;29(11):1818–1826.e6. doi:10.1016/j.cub.2019.04.009.

Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020;17(2):155–158. doi:10.1038/s41592-019-0669-3.

Schiffer PH, Natsidis P, Leite DJ, Robertson H, Lapraz F, *et al.* The slow evolving genome of the xenacoelomorph worm *Xenoturbella bocki*.

bioRxiv 2022.06.24.497508. https://doi.org/10.1101/2022.06.24.497508.

Shao M, Kingsford C. Accurate assembly of transcripts through phase-preserving graph decomposition. Nat Biotechnol. 2017; 35(12):1167–1169. doi:10.1038/nbt.4020.

Sim SB, Corpuz RL, Simmonds TJ, Geib SM. Hifiadapterfilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. BMC Genomics. 2022;23(1):157. doi:10.1186/s12864-022-08375-1.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM, *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31(19):3210–3212. doi:10.1093/bioinformatics/btv351.

Sovic I. IPA HiFi Genome Assembler. https://github.com/PacificBiosciences/pbipa. 2022.

Tempel S. Using and understanding RepeatMasker. Methods Mol Biol. 2012;859:29–51. doi:10.1007/978-1-61779-603-6_2.

Ustyantsev KV, Berezikov EV. Computational analysis of spliced leader trans-splicing in the regenerative flatworm *Macrostomum lignano* reveals its prevalence in conserved and stem cell related genes. Vavilov J Genet Breed. 2021;25(1):101–107. doi:10.18699/VJ21.012.

Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci. 2021;1(5):332–336. doi:10.1038/s43588-021-00073-4.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11): e112963. doi:10.1371/journal.pone.0112963.

Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005; 21(9):1859–1875. doi:10.1093/bioinformatics/bti310.

Wudarski J, Simanov D, Ustyantsev K, de Mulder K, Grelling M, *et al.* Efficient transgenesis and annotated genome sequence of the regenerative flatworm model *Macrostomum lignano*. Nat Commun. 2017;8(1):2120. doi:10.1038/s41467-017-02214-8.

Xu G-C, Xu T-J, Zhu R, Zhang Y, Li S-Q, Wang H-W. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. Gigascience. 2019;8(1):giy157. doi:10.1093/gigascience/giy157.

Zadesenets K, Ershov N, Berezikov E, Rubtsov N. Chromosome evolution in the free-living flatworms: first evidence of intrachromosomal rearrangements in karyotype evolution of *Macrostomum lignano* (Platyhelminthes, Macrostomida). Genes (Basel). 2017; 8(11):298. doi:10.3390/genes8110298.

Zhu B-H, Xiao J, Xue W, Xu G-C, Sun M-Y, Li J-T. P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. BMC Genomics. 2018;19(1):175. doi:10.1186/s12864-018-4567-3.

Communicating editor: M. Nowrousian