DATA NOTE

A draft nuclear-genome assembly of the acoel flatworm Praesagittifera naikaiensis

Asuka Arimoto,†, Tomoe Hikosaka-Katayama, Akira Hikosaka, Kuni Tagawa, Toyoshige Inoue, Tatsuya Ueki, Masa-aki Yoshida, Miyuki Kanda, Eiichi Shoguchi, Kanako Hisata and Noriyuki Satoh

1Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Okinawa 904-0495, Japan, 2Natural Science Center for Basic Research and Development, Gene Science Division, Hiroshima University, 1-4-2 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan, 3Division of Human Sciences, Graduate School of Integrated Arts and Sciences, Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8521, Japan, 4Marine Biological Laboratory, Graduate School of Science, Hiroshima University, 2445 Mukaishima, Onomichi, Hiroshima 722-0073, Japan, 5Department of Biological Science, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan, 6Marine Biological Science Section, Education and Research Center for Biological Resources, Faculty of Life and Environmental Science, Shimane University, 194 Kamo, Okinoshima-cho, Oki, Shimane 685-0024, Japan and 7DNA Sequence Section, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Okinawa 904-0495, Japan

Abstract

Background: Acoels are primitive bilaterians with very simple soft bodies, in which many organs, including the gut, are not developed. They provide platforms for studying molecular and developmental mechanisms involved in the formation of the basic bilaterian body plan, whole-body regeneration, and symbiosis with photosynthetic microalgae. Because genomic information is essential for future research on acoel biology, we sequenced and assembled the nuclear genome of an acoel, Praesagittifera naikaiensis. Findings: To avoid sequence contamination derived from symbiotic microalgae, DNA was extracted from embryos that were free of algae. More than 290x sequencing coverage was achieved using a combination of Illumina (paired-end and mate-pair libraries) and PacBio sequencing. RNA sequencing and Iso-Seq data from embryos, larvae, and adults were also obtained. First, a preliminary 17-kilobase pair (kb) mitochondrial genome was assembled, which was deleted from the nuclear sequence assembly. As a result, a draft nuclear genome assembly was 656 Mb in length, with a scaffold N50 of 117 kb and a contig N50 of 57 kb. Although 70% of the assembled sequences were likely...
Background

Acoels are small, very simple, planaria-like animals lacking a coelom, a gut, and a circulatory system. Traditional taxonomy positioned the Acoela as the most basal order of the phylum Platyhelminthes [1]. Recent analyses using molecular data, however, have suggested that acoels are members of a new phylum, the Xenacoelomorpha, together with nemedodermites and xenoturbellids [2–4]. However, whether Xenacoelomorpha is a monophyletic taxon, whether xenacoelomorphs are basal to all other bilaterians, and whether they have close affinity to ambulacrians are matters of debate [2–4]. Nonetheless, acoels are pivotal to understanding the origins and evolution of simple bilaterians. Keywords: acoel; Praesagittifera naikaiensis; draft nuclear genome

Sampling and Sequencing

Biological materials

The marine acoel worm *P. naikaiensis* is 2–3 mm in length (Fig. 1A) [9]. Members of this species are easily found at seashores of the Seto Inland Sea, Japan, especially during the early summer season (Fig. 1B). Adults contain symbiotic microalgae, *Tetraselmis* species, which are integrated during juvenile growth (Fig. 1C). Adults were collected at the seashore near the Marine Biological Laboratory of Hiroshima University and maintained in aquaria in the laboratory on a 12-h light/12-h dark photoperiod. Naturally laid eggs were collected and cultured for embryogenesis (Fig. 1C). Embryos were free of symbiotic microalgae. After washing embryos with filtered seawater, genomic DNA was extracted from them using the phenol/chloroform extraction method. Embryos, juveniles, and adults were sampled for RNA sequencing. Total RNA extraction was performed using TRIzol Reagent (ThermoFisher, MA, USA, 15596-026) and an RNeasy mini Kit (Qiagen, Hilden, Germany, 74104).

Library preparation and sequencing

DNA

All sequencing libraries were constructed according to the manufacturers’ standard protocols. Briefly, for the Illumina platform, polymerase chain reaction (PCR)-free, paired-end libraries were prepared using an Illumina TruSeq DNA PCR-Free LT Library Prep Kit (Illumina, CA, USA, FC-121-3001). Four mate-pair libraries were prepared using a Nextera Mate Pair Library Prep Kit (Illumina, FC-132-1001) (Supplementary Table 1).

For the PacBio platform, a DNA library was prepared using the manufacturer’s 10-kb (kilobase pair) template preparation protocol. A SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, CA, USA, 100-259-100) was used for PacBio library preparation. The long-read DNA library was sequenced using a PacBio RSII sequencer employing P6-C4 chemistry (Pacific Biosciences, 100-372-700) with 360-min movie lengths. A total of 52 SMRT Cells were sequenced for long-read DNA library.

RNA

An RNA sequencing (RNA-seq) library was prepared using a TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, RS-122-2101). The library was sequenced using the Illumina HiSeq 2500 platform (Supplementary Table 1).

Complementary DNAs for Iso-Seq libraries were prepared using a SMARTer PCR cDNA Synthesis Kit (Clontech, CA, USA, 634925). The SageELF size selection system (Sage Science, MA, USA) was used following the manufacturer’s standard protocol (Supplementary Table 1). A SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, 100-259-100) was used for Iso-Seq library preparation. The library was sequenced using a PacBio RSII sequencer employing P6-C4 chemistry (Pacific Biosciences, 100-372-700) with 360-min movie lengths. A total of 8 SMRT Cells were sequenced for the Iso-Seq RNA library.

Assembly of mitochondrial and nuclear genomes

Adapter sequences in PCR-free and mate-pair Illumina reads were removed with Trimmomatic 0.36 (Trimmomatic, RRID:SCR_011848) [10] and NextClip 1.3.1 (NextClip, RRID:SCR_005465) [11], respectively. Low-quality (<Q20) inserts were removed using Sickle 1.33 (Sickle, RRID:SCR_006800) [12] after adapter cleanup. Reads that lacked a corresponding pair were discarded.

Mitochondrial genome assembly

To distinguish mitochondrial genome sequences from the nuclear genome assembly, we first assembled the mitochondrial genome. To this end, a mitochondrial 16S ribosomal RNA of *Symsagittifera roscoffensis* (accession No. NC_014578) [7] was used to collect PacBio long reads of *P. naikaiensis* sequences using BLAST+ 2.3.0 [13] with the “dc-megablast” option.

Collected reads longer than 1 kb and shorter than 12 kb were assembled using sprai 0.9.9.19 [14] with default settings. Circularity of assembled contigs was checked automatically in the sprai assembly pipeline.

Nuclear genome assembly

*P. naikaiensis* mitochondrial sequences were mapped onto trimmed reads using BWA 0.7.12 (BWA, RRID:SCR_010910) [15],
Figure 1: The acorn worm P. naikaiensis. (A) An adult, dorsal view. Anterior, top and posterior, bottom. Green dots throughout the entire body are symbiotic green algae. Two eggs are seen in the center of the worm. (B) An enormous number of adults gathering on the sandy seashore, resembling dark masses. (C) Embryogenesis. A newly laid egg within the eggshell (i), a 4-cell stage embryo (ii), a gastrula (iii), a flattened-stage embryo (iv), newly hatched aposymbiotic algae (v), and a symbiotic juvenile with symbiotic algae (vi). (D) A peripheral region of an adult worm showing symbiotic microalgae, Tetraselmis species. Scale = 50 μm in (A), (C), (D); 10 cm in (B).

Figure 2: A screen shot of the genome browser of Praesagittifera naikaiensis.

and those read pairs that mapped onto the mitochondrial sequences were excluded from the data set. PacBio long reads were also mapped against the mitochondrial genome using BLASR (BLASR, RRID:SCR_000764) (commit version: 5.3.574e1c2) [16]. Only unmapped or cleaned-up Illumina and PacBio reads were assembled using the MaSuRCA (MaSuRCA, RRID:SCR_010691) assembler 3.2.2 [17].

Putative heterozygous and/or polymorphic sequences that remained in the assembled genome were merged as homozygous sequences using redundans 0.13c [18]. Gaps in the homozygous genome were filled using PBjelly (PBjelly, RRID:SCR_012091) in PBSuite 15.8.24 [19]. After gap closing, BESST 2.2.6 [20] and LINKS 1.8.5 [21] were used to perform scaffolding with Illumina and PacBio reads, respectively. Scaffolds were polished using Racon (commit version: 083444) [22] with PacBio long reads.

PacBio Iso-Seq reads were mapped onto scaffolds using GMAP (GMAP, RRID:SCR_008992) version 2017-08-15 [23], and then L_RNA_scaffolder [24] was used to concatenate scaffolds based on the results of Iso-Seq read mapping. Scaffolds were polished using Pilon 1.22 (Pilon, RRID:SCR_014731) [25] with PCR-free Illumina reads used for the MaSuRCA assembly. BUSCO 3.0.2 (Benchmarking Universal Single-Copy Orthologs, RRID:SCR_015008) [26] with a metazaoan data set was used to evaluate the polished final genome assembly.

Genome size estimation

PCR-free, paired-end reads used for genome assembly were analyzed. K-mers in the data set were counted with Jellyfish 2.2.3 (Jellyfish, RRID:SCR_005491) [27] (Supplementary Figure 1). The genome size of P. naikaiensis was estimated from obtained k-mer frequencies using GenomeScope web tools [28].
Transcriptome assembly, gene prediction, and gene annotation

Adapter sequences and low-quality (<Q30) reads in the resulting RNA-seq paired-end data were removed using Trimmomatic (Trimmomatic, RRID:SCR_011848). Cleaned reads were assembled using Trinity 2.1.1 (Trinity, RRID:SCR_013048) [31] with default settings and the strand-specific option. In addition, genome-guided transcriptome assembly was performed. RNA-seq reads were mapped onto the genome using STAR 2.5.2a (STAR, RRID:SCR_018599) [32] and then mapped reads were assembled using Trinity. De novo assembled Illumina transcriptome and PacBio Iso-Seq sequences were mapped onto the genome using minimap-2 version 2.6 [33] with the “-ax splice” option. These mapping results and the genome-guided assembly of illumina RNA-seq reads were integrated based on genome sequences using PASA 2.2.0 (PASA, RRID:SCR_014656) [34]. Putative full-length transcripts having both a 5′ and a 3′ untranslated region were detected using TransDecoder 5.0.2 [35]. These full-length transcripts were used as a training set for gene prediction. De novo transcriptome assembly of a data set containing 15 xenacoelomorphs (Supplementary Table 2) was also performed, following the procedure described above to create similarity hints for gene prediction. Assembled sequences of other acellos were translated into protein sequences using TransDecoder and then mapped against the P. naikaiensis genome using Exonerate 2.2.0 (Exonerate, RRID:SCR_016088) [36]. A final set of gene models reflecting hint information was generated with AUGUSTUS 3.2.1 (Augustus, RRID:SCR_008417) [37]. Gene models were annotated using BLAST searches (E-value cutoff of 10⁻⁵) against the NCBI RefSeq protein database release 88. Protein domains in gene models were detected using HMMER 3.1b2 (Hmmer, RRID:SCR_005305) [38] and Pfam-A 31.0 under default settings, except for an E-value cutoff of 10⁻³.

A draft assembly

Mitochondrial genome

The complete, closed circular mitochondrial genome of P. naikaiensis was recovered from genome-sequencing data. The mitochondrial genome is 17,787 base-pairs long and contains 12 protein-coding genes, small and large ribosomal RNAs, and 22 predicted transfer RNAs (Supplementary Figure 2). When cox1 was positioned at the start of the genome on the “positive” strand, 8 protein-coding genes were found in the same strand while nad2, cyt b, and nad5 were found on the “negative” strand (Supplementary Figure 2). Both ribosomal RNAs were found separately on the positive strand. Although the number of mitochondrial genes of P. naikaiensis is comparable to that of the previously reported mitochondrial genomes of Archaphanostoma ylauæ [8], the order of genes in the genomes was quite different between them.

Nuclear genome

K-mer analysis showed that the P. naikaiensis genome constitutes ~654 Mb (Table 1; Supplementary Figure 1). Illumina paired-end and mate-pair, and PacBio reads provided 204x and 221x, and 73x coverage of the estimated genome, respectively (Supplementary Table 1 and Supplementary Figure 1). The assembly appeared to plateau during both scaffolding and contig formation (Supplementary Figure 3). As a result, the draft assembly comprised 656 Mb (Table 1), very close to the estimated genome size. The scaffold N50 reached 117 kb, and 12 scaffolds were >500 kb in length (Table 1; Supplementary Table 3). In-

---

**Table 1: Genome assembly statistics for P. naikaiensis**

| Genome feature                  | Value  |
|---------------------------------|--------|
| Estimated genome size*          | 654.1 Mb |
| Assembled genome size           | 656.1 Mb |
| Scaffolds (≥500 bp)             |        |
| No.                             | 12,072 |
| N50                             | 117 kb |
| Contigs (≥500 bp)               |        |
| No.                             | 24,071 |
| N50                             | 57 kb  |
| Gaps                            | 1.66%  |
| Repetitive sequences            | 69.8%  |
| Guanine + cytosine content      | 39.1%  |
| Predicted protein-coding genes  | 22,143 |
| (loci)                          |        |
| Genes with transcript support   | 99%    |
| Mean transcript length          | 2447 nucleotides |
| Mean exon frequency per gene    | 5.7    |
| BUSCO analysis                  |        |
| Complete                        | 748/978 (76.5%) |
| Fragmented                      | 37/978 (3.8%) |

*Estimated by k-mer analysis of Illumina PCR-free reads as shown in Supplementary Figure 1.

**Table 2: Repetitive sequences in the P. naikaiensis genome assembly**

| Class               | %   |
|---------------------|-----|
| DNA transposons      | 12.2|
| MULE                | 5.7 |
| Maverick            | 4.5 |
| hAT                 | 0.9 |
| Others              | 4.9 |
| Retrotransposons     | 41.5|
| LTR                 | 35.7|
| Gypsy               | 28.2|
| Copia               | 1.5 |
| Pao                 | 1.1 |
| Others              | 4.9 |
| Long interspersed nuclear elements | 4.8 |
| CR1                 | 1.7 |
| CRE                 | 1.0 |
| L2                  | 0.9 |
| Others              | 1.2 |
| Short interspersed nuclear elements | 1.0 |
| Others              | 2.2 |
| RNA                 | 0.03|
| Low complexity       | 0.07|
| Satellite            | 0.4 |
| Simple repeat        | 1.7 |
| Unclassified         | 20.4|
| Total (excluding overlapped sequences) | 69.8 |

Repeat analysis

Repetitive sequences in the assembled genome were identified using RepeatModeler 1.0.11 (RepeatModeler, RRID:SCR_015027) [29] and RepeatMasker 4.0.7 (RepeatMasker, RRID:SCR_012954) [30].
Table 3: Numbers of putative transcriptional regulator genes in the P. naikaiensis genome

| Accession ID | Description         | No. of genes |
|--------------|---------------------|--------------|
| PF00010      | HLH                 | 20           |
| PF00046      | Homeobox            | 62           |
| PF00096      | zf-C2H2             | 73           |
| PF00104      | Hormone_recep       | 14           |
| PF00105      | zf-C4               | 20           |
| PF00157      | Pou                 | 3            |
| PF00170      | bZIPI              | 13           |
| PF00178      | Ets                 | 13           |
| PF00250      | Fork_head           | 11           |
| PF00292      | PAX                 | 5            |
| PF00319      | SRF-TF              | 2            |
| PF00320      | GATA                | 7            |
| PF00505      | HMG_box             | 12           |
| PF00554      | RHD                 | 1            |
| PF00853      | Runt                | 1            |
| PF00870      | PS3                 | 1            |
| PF00907      | T-box               | 4            |
| PF01388      | ARID                | 4            |
| PF01530      | zf-C2HC             | 2            |
| PF02376      | CUT                 | 3            |
| PF03299      | TF_AP-2             | 1            |
| PF05044      | Prox1               | 1            |
| PF07527      | Hairy_orange        | 11           |
| PF07716      | bZIP_2              |             |

Transcriptomes

Transcriptome data, especially those from PacBio Iso-Seq long reads, provided a set of high-quality RNA data (Supplementary Table 1). The mean length of transcriptomes was 2,447 nucleotides, and the mean number of exons per gene was 5.7 (Table 1).

Gene modeling

Gene modeling of the P. naikaiensis genome produced 22,143 protein-coding genes (Table 1). As mentioned above, we obtained a set of high-quality RNA data. As a result, 99% of gene models were substantiated by the transcriptomes (Table 1).

BUSCO analysis indicated that 76.5% and 3.8% of them were supported as complete and fragmented genes, respectively (Table 1).

Gene annotation

Gene families predicted by RefSeq (BLAST), Pfam (HMMER), and PANTHER (HMMER) were 15,294, 13,225, and 17,384 in number, respectively (Supplementary Figure 5). Using Pfam-supported families, we examined the number of gene families. Table 3 shows numbers of putative transcription regulator genes in the P. naikaiensis genome. The 2 most abundant families were zinc.
finger (C2H2 type) and homeobox domain-containing genes, with 73 and 62 members, respectively. Twenty each were annotated to the helix-loop-helix and zinc finger (C4 types) families. Although more detailed analysis is required, the *P. naikaiensis* genome appears to contain numbers of transcription regulator genes comparable to those of other bilaterian genomes.

A similar analysis was carried out on putative signaling molecule genes (Table 4). The largest gene family was tyrosine kinase, represented by 316 genes. In addition, epidermal growth factor (EGF)-like domain genes, G-protein \(\alpha\) subunit genes, and regulator of G-protein signaling genes numbered 28, 31, and 16, respectively (Table 4).

### Genome browser

A genome browser was established for the assembled sequences using the JBrowser 1.12.3 [39]. Its URL is [40] (Fig. 2). The gene annotations from the protein domain search and BLAST search have similarly been shown on the site.

### Availability of supporting data and materials

Genomic and transcriptomic sequence reads have been deposited in the DDBJ sequence read archive under accession No. PRJDB7329. All data are also available from the GigaScience GigaDB repository [41].

### Additional files

**Supplementary Table 1:** Sequence data summary.

**Supplementary Figure 1:** K-mer analysis and genome size estimation of *Prasagittifera naikaiensis* genomic DNA reads.

**Supplementary Table 2:** Xenacoelomorph dataset used for gene prediction.

**Supplementary Figure 2:** A preliminary circular assembly of the mitochondria genome of *Prasagittifera naikaiensis*.

**Supplementary Figure 3:** Accumulation of assembled sequences (contigs, blue and scaffolds, red) reaching over 600 Mb.

**Supplementary Table 3:** Summary of the *Prasagittifera naikaiensis* genome assembly.

**Supplementary Figure 4:** GC content of the *Prasagittifera naikaiensis* genome.

**Supplementary Figure 5:** *Prasagittifera naikaiensis* gene annotation.

### Abbreviations

BUSCO: Benchmarking Universal Single-Copy Orthologs; EGF: epidermal growth factor; kb: kilobase pair; LTR: long terminal repeat; Mb: megabase pair; MULE: mutator-like transposable element; PCR: polymerase chain reaction; RNA-seq: RNA sequencing; TE: transposable element.
Competing interests
The authors declare that they have no competing interests.

Funding
This work was funded by Okinawa Institute of Science and Technology Internal Funds to the Marine Genomics Unit (N.S.). This work was also supported by a JSPS grant (No. 17K0753S) and a research grant from the Research Institute of Marine Invertebrates (Japan) to A.H.

Author contributions
N.S., T.H.K., A.H., K.T., A.A., M.A.Y., and T.U. conceived and supervised the project. T.H.K. and T.I. collected the majority of samples. M.K. and A.A. performed sequencing. N.S., T.H.K., A.H., K.T., A.A., M.A.Y., and T.U. conceived and supervised the project. T.H.K. and T.I. collected the majority of samples. M.K. and A.A. performed sequencing. N.S., A.A., and T.H.K. prepared the manuscript, and all authors approved the final manuscript.

Acknowledgments
We are grateful to Dr. Steven D. Aird for his technical editing of the manuscript.

References
1. Hyman LH. The Invertebrates: Platyhelminthes and Rhyncocoea; the Acoelomate Bilateria. New York: McGraw-Hill, 1951.
2. Ruiz-Trillo I, Riuort M, Littlewood DTJ, et al. Acoel flatworms: earliest extant bilaterian metazoa, not members of Platyhelminthes. Science 1999;283:1919–23.
3. Philippe H, Brinkmann H, Copley RR, et al. Acoelomorph flatworms are deuterostomes related to Xenoturbella. Nature 2011;470:255–8.
4. Cannon JT, Vellutini BC, Smith J, 3rd, et al. Xenacoelomorpha is the sister group to Nephrozoa. Nature 2016;530:89–93.
5. Srivastava M, Mazza-Curll KL, van Wolswinkel JC, et al. Whole-body acoel regeneration is controlled by Wnt and Bmp-Admp signaling. Curr Biol 2014;24:1107–13.
6. Ruiz-Trillo I, Riuort M, Fournade HM, et al. Mitochondrial genome data support the basal position of Acoelomorpha and the polyphyley of the Platyhelminthes. Mol Phylogenet Evol 2004;33:321–32.
7. Mwinyi A, Bailly X, Boulat SJ, et al. The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of Symaphagittfera roscoffensis. BMC Biol 2010;10:309.
8. Robertson HE, Lapraz F, Egger B, et al. The mitochondrial genomes of the acoelomorph worms Paratomella rubra, Isodiametra pulchra and Archaphanostoma yluve. Sci Rep 2017;12:1847.
9. Hikosaka-Katayama T, Hikosaka A. Artificial rearing system for Praesagittifera naikaiensis (Acoela, Acoelomorpha)[in Japanese]. Bull Grad Sch Integr Arts Sci Hiroshima Univ I Stud Hum Sci 2015;10:17–23.
10. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
11. Leggett RM, Clavijo BJ, Clissold L, et al. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. Bioinformatics 2014;30:566–8.
12. Sickle. https://github.com/najoshi/sickle, Accessed 6 October 2017.
13. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.
14. sprai = single pass read accuracy improver. http://zombie.cb.k.u-tokyo.ac.jp/sprai/index.html, Accessed 12 September 2016.
15. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013; arXiv:1303.3997v2 [q-bio.GN].
16. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): algorithm and theory. BMC Bioinformatics 2012;13:238.
17. Zimin AV, Marçais G, Puiu D, et al. The MaSuRCA genome assembler. Bioinformatics 2013;29:2669–77.
18. Przyshcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 2016;44:e113.
19. English AC, Richards S, Han Y, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 2012;7:e47768.
20. Sahlin K, Chikhi R, Arvestad L. Assembly scaffolding with PE-contaminated mate-pair libraries. Bioinformatics 2016;32:1925–32.
21. Warren RL, Yang C, Vandervalk BP, et al. LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience 2015;4:35.
22. Vaser R, Sović I, Najarajan N, et al. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 2017;27:737–46.
23. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 2005;21:1859–75.
24. Xue W, Li JT, Zhu YP, et al. LRNA scafholder: scaffolding genomes with transcripts. BMC Genomics 2013;14:604.
25. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9:e112963.
26. Waterhouse RM, Seppey M, Simão FA, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 2018;35:543–8.
27. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 2011;27:764–70.
28. Vurture GW, Sedlazeck FJ, Nattestad M, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 2017;33:2202–4.
29. RepeatModeler. http://www.repeatmasker.org/RepeatModeler, Accessed 13 February 2018.
30. RepeatMasker. http://www.repeatmasker.org, Accessed 13 February 2018.
31. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644–52.
32. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
33. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34:3094–100.
34. Haas BJ, Delcher AL, Mount SM, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 2003;31:5654–66.
35. TransDecoder. https://github.com/TransDecoder/TransDecoder, Accessed 6 January 2018.
36. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics.
2005;6:31.
37. Stanke M, Diekhans M, Baertsch R, et al. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 2008;24:637–44.
38. HMMER. http://hmmer.org., Accessed 4 March 2015
39. Skinner ME, Uzilov AV, Stein LD, et al. JBrowse: a next-generation genome browser. Genome Res 2009;19:1630–8.
40. Praesagittifera naikaiensis ver. 1.0. http://mari
41. Arimoto A, Hikosaka-Katayama T, Hikosaka A, et al. Supporting data for "A draft genome assembly of the acoel flatworm Praesagittifera naikaiensis." GigaScience Database 2019. http://dx.doi.org/10.5524/100564.