< Mol. Ecol. Resources >-revised

GENOMIC RESOURCES NOTE

The genome of the freshwater monogonont rotifer *Brachionus calyciflorus*

Hui-Su Kim[1], Bo-Young Lee[1], Jeonghoon Han[1], Chang-Bum Jeong[1], Dae-Sik Hwang[1], Min-Chul Lee[1], Hye-Min Kang[1], Duck-Hyun Kim[1], Hee-Jin Kim[2], Spiros Papakostas[3], Steven A.J. Declerck[4], Ik-Young Choi[5], Atsushi Hagiwara[2,6], Heum Gi Park[7], and Jae-Seong Lee[1,*]

[1] Department of Biological Science, College of Science, Sungkyunkwan University, Suwon 16419, South Korea
[2] Graduate School of Fisheries and Environmental Sciences, Nagasaki University, Nagasaki 852-8521, Japan
[3] Department of Biology, University of Turku, Turku, Finland
[4] Department of Aquatic Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box 50, 6700AB, Wageningen, The Netherlands
[5] Department of Agriculture and Life Industry, Kangwon National University, Chuncheon 24341, South Korea
[6] Organization for Marine Science and Technology, Nagasaki University, Nagasaki 852-8521, Japan
[7] Department of Marine Resource Development, College of Life Sciences, Gangneung-Wonju National University, Gangneung 25457, South Korea

Abstract

Monogononta is the most speciose class of rotifers, with more than 2000 species. The monogonont genus *Brachionus* is widely distributed at a global scale, and a few of its species are commonly used as ecological and evolutionary models to address questions related to aquatic ecology, cryptic speciation, evolutionary ecology, the evolution of sex, and ecotoxicology. With the importance of *Brachionus* species in many areas of research, it is remarkable that the genome has not been characterized. This study aims to address this lacuna by presenting, for the first time, the whole genome assembly of the freshwater species *Brachionus calyciflorus*. The total length of the assembled genome was 129.6 Mb, with 1,041 scaffolds. The N50 value was 786.6 kb and the GC content was 24%. A total of 16,114 genes were annotated with repeat sequences, accounting for 21% of the assembled genome. This assembled genome may form a basis for future studies addressing key questions on the evolution of monogonont rotifers. It will also provide the necessary molecular resources to mechanistically investigate ecophysiological and ecotoxicological responses.

Key words: Monogonont rotifer, genome, molecular ecotoxicology, evolution

Introduction

The phylum Rotifera comprises a group of primary freshwater/seawater metazoans comprising four classes: Seisonidea, Bdelloidea, Monogononta, and Acanthocephala. Monogononta represents the most speciose rotifer class, with 1,570 species (Segers 2008). Monogonont rotifers inhabit a wide variety of aquatic and moist habitats, and are ubiquitous

*Correspondence: jslee2@skku.edu

among the plankton of freshwater, brackish, and marine waters, in mosses, and in other wet habitats (Wallace et al. 2006). Monogonont rotifers are suitable model organisms, due to a peculiar combination of morphological and physiological characteristics such as small body size (typically ranging from ~100 to 360 μm), ease of culturing, cyclic parthenogenesis, a strong capacity for rapid population growth, and a high sensitivity to various toxic substances (Snell et al. 1991; Snell &, Carmona 1995; Hagiwara et al. 1997; Preston et al. 2000; Hagiwara et al. 2007).

The monogonont genus *Brachionus* is widely distributed and plays an important role in aquatic ecosystems (Arndt 1993; Dahms et al. 2011; Won et al. 2017). In several *Brachionus* species, the ecology, morphology, and reproduction biology have been well studied (Hagiwara et al. 1995; Sha et al. 2015). Indeed, several *Brachionus* species have been used as model systems for studies on aquatic ecology (Gilbert & Walsh 2005), speciation biology (Gomez et al. 2002; Papakostas et al. 2016), rapid evolutionary adaptation (Declerck et al. 2015; Declerck & Papakostas 2017), the evolution of sex (Fussmann et al. 2003; Stelzer & Snell 2003; Snell et al. 2006; Smith & Snell 2012), population dynamics (Ortells et al. 2000, 2003; Yoshinaga et al. 2003), and ecotoxicology (Snell & Persoone 1989a,b; Snell & Janssen 1995; Snell et al. 2003; Kim et al. 2013; Jeong et al. 2016; Won et al. 2016). More specifically, the freshwater rotifer *Brachionus calyciflorus* has been used for many studies on the effects of toxicants, endocrine disruptors, and gradients of temperature and salinity (Snell et al. 1991; Snell &, Carmona 1995; Preston et al. 2000). Recent evidence suggests that *B. calyciflorus* is a species complex of at least four cryptic species (Papakostas et al. 2016).

Despite the importance of monogonont rotifers in many research areas, the genomic resources are limited, with the exception of the strictly asexual bdelloid rotifer *Adineta vaga*, for which the genome assembly is available (Flot et al. 2013). The characteristics of the *A. vaga* genome are highly unique, with degenerated tetraploidy with anciently duplicated segments; abundant gene conversion to limit deleterious mutations; and expansion of gene families that are involved in resistance to oxidation, carbohydrate metabolism, and defense against the activity of transposable elements (Mark Welch & Meselson 2000; Mark Welch et al. 2008; Flot et al. 2013). However, the monogonont rotifer *B. calyciflorus* is a cyclical parthenogenetic organism and is therefore considered a suitable model for the study of the evolution of sex (Serra & Snell 2009).

The whole genome information of *B. calyciflorus* will be very useful for understanding the evolutionary relationships between two different reproductive modes (i.e. the mictic and amictic cycles) and for revealing the molecular mechanisms of the response to environmental stressors. In this study, we present a *de novo* assembly of the genome of the freshwater rotifer *B. calyciflorus* with gene annotation.

Materials and methods

Rotifer culture

Resting eggs of *B. calyciflorus* were collected in Zwartenhoek, The Netherlands (52.0263N and 4.18355E). The eggs were hatched and neonates were used to establish clonal lines. Clonal lines were screened with restriction fragment length polymorphism (RFLP) analysis of amplified Internal Transcribed Spacer 1 (*ITS1*) and identified as putative species 'C' according to Papakostas et al. (2016). *B. calyciflorus* rotifers were reared and maintained at the aquarium facility of the Department of Biological Science, Sungkyunkwan University (Suwon, South Korea).

B. calyciflorus **(Fig. 1)** was reared in freshwater containing penicillin (Sigma-Aldrich; final concentration 100 units/L) and streptomycin (Sigma-Aldrich; final concentration 100 μg/L) to minimize contaminants. The culture temperature was 25℃ with a photoperiod of light:dark (LD) 12:12 h. The green alga *Chlorella vulgaris* (strain KMCC FC-012, Busan, South Korea) were used as a live diet (approximately 6×10^4 cells/mL). The species

identification of *B. calyciflorus* was confirmed by morphological analysis and sequencing of the mitochondrial DNA gene *CO1*, as suggested by Hwang et al. (2013). Prior to DNA extraction, the water was refreshed every 2 to 3 h for each 12 h period, to allow the rotifers to consume any remaining *Chlorella* and excrete their gut contents. All animal handling and experimental procedures were approved by the Animal Welfare Ethical Committee and the Animal Experimental Ethics Committee of Sungkyunkwan University (Suwon, South Korea).

Preparation of genomic DNA and sequencing libraries

For genomic DNA isolation, adult rotifers (approximately 4000 individuals) were homogenized in three volumes of DNA extraction buffer (100 mM NaCl; 10 mM Tris-Cl, pH 8.0; 25 mM ethylenediamine-tetraacetic acid [EDTA]; 0.5% sodium dodecyl sulfate [SDS]; 100 μg/ml proteinase K; and 1 μg/ml RNase) using a Teflon homogenizer, and incubated in a water bath at 55℃ overnight. The incubated sample was subjected to phenol/chloroform and chloroform extraction, after which genomic DNA was precipitated with isopropanol and 0.2x volume of 10 M ammonium acetate, followed by centrifugation at 9000 rpm for 10 min. After washing the pellet with 70% ethanol, the genomic DNA was resuspended in TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0). The quality and quantity of the genomic DNA were analyzed using a QIAxpert system (Qiagen, Hilden, Germany) and agarose gel electrophoresis with visual inspection, respectively. Whole genome sequencing of *B. calyciflorus* was performed using the Illumina HiSeq 2500 platform (Illumina, USA). Three PE libraries (PE300, PE500, and PE800) were constructed for the initial contigs assembly, and three MP libraries (MP2kb, MP5kb, and MP10kb) were used for scaffold construction. The sequencing reaction was performed in 251 bp and 151 bp for the PE libraries and the MP libraries, respectively. Construction of genomic sequencing libraries and all sequencing processes were performed at the National Information Center for Educational Media (NICEM; Seoul, South Korea) according to the manufacturer's instructions.

Whole genome assembly

The preprocessing pipeline of raw sequence reads included sequence adaptor removal, quality trimming, error correction, duplicate removal, and contaminant removal. Sequencing adaptor removal and quality trimming for the PE libraries were performed using Trimmomatic v0.33 (Bolger et al. 2014). For quality trimming, all sequence reads with a Phred score below 20 were removed. Duplicated raw reads were removed using FastUniq (Xu et al. 2012), which identifies the duplicates by comparing sequences between read pairs. Error correction in the trimmed reads was conducted using BBTools (http://jgi.doe.gov/data-and-tools/bbtools/). After preprocessing of the raw sequences, GenomeScope (Vurture et al. 2017) was used to estimate the overall characteristics of the *B. calyciflorus* genome, including k-mer analysis prior to the assembly of the sequence reads. For the MP libraries, sequencing adaptors and junction adaptors were removed using Skewer (Jiang et al. 2014). Quality trimming, error correction, and duplicate removal were carried out with the same procedures as those used for the PE libraries.

One of the challenges in genome assembly when using the sequences generated from whole bodies of organisms is the elimination of the contaminants from the culture environment. Contaminant removal in this study was performed in two steps. First, we collected contaminant sequence data from the NCBI RefGen database and constructed a customized contaminant database (https://www.ncbi.nlm.nih.gov/refseq/about/prokaryotes/). All sequence reads mapping to sequences in the database were removed using BBTools. Second, we performed preliminary assembly without parameter optimization to reduce the complexity of the dataset. Blobology (Kumar et al. 2013) was used to screen the contaminants. Based on the Blobology plot analysis, contigs of contaminants were confirmed, and all sequence reads mapping to the contaminant contigs were removed from the final

dataset. *De novo* assembly, including scaffold construction and gap closing, was performed based on multiple k-mer values automatically optimized by the Platanus assembler v1.2.4 (Kajitani et al. 2014). Allelic relationships among scaffolds were reconstructed using HaploMerger v2 (Huang et al. 2012, Huang et al. 2017) with repeat-masked assembled sequences. HaploMerger2 includes SSPACE v3.0 (Marten et al. 2010) and GapCloser as the scaffold constructor and the gap closer, respectively. CEGMA v2.5 (Parra et al. 2017) and BUSCO v3.0 (Simao et al. 2015) were used to assess the completeness of the *B. calyciflorus* genome.

RNA-seq library construction and sequencing analysis

The transcriptome sequences used in this study were obtained from RNA isolated from adult *B. calyciflorus* rotifers; the isolate was filtered with a 90 μm sieve. In brief, total RNA was extracted from rotifers, using TRIZOL® reagent (Molecular Research Center, Inc., Cincinnati, OH, USA) according to the manufacturer's instructions. Co-purified genomic DNA was removed by treatment with DNase I (Sigma, St. Louis, MO, USA). Total RNA was quantified by measuring the absorbance at 230, 260, and 280 nm with a spectrophotometer (QIAxpert®). To ensure that the obtained RNA was of suitable quality for constructing sequencing libraries, *18S*/*28S* ribosomal RNA integrity and band ratios were determined using a Bioanalyzer (Agilent 2100; Santa Clara, CA, USA). Complementary DNA synthesis, RNA-seq library construction, and sequencing reactions were performed at the NICEM using an Illumina HiSeq2500 instrument. All procedures were performed according to the manufacturer's instructions. After sequencing analysis, quality control checks were performed according to the same procedures as those used for the PE libraries. *In silico* normalization was performed using a 50% depth of the cleaned sequence reads and *de novo* transcriptome assembly was performed with Trinity assembler v2.4.0 (Grabherr et al. 2011). The open reading frames (ORFs) of the assembled transcripts were identified using TransDecoder v3.0.1 (http://transdecoder.github.io). Reference guiding of the assembled transcripts was performed using HiSat2 v2.0.5 (https://ccb.jhu.edu/software/hisat2/) and Cufflinks v2.2.1 (Trapnell et al. 2010). The reference index was built with the final assembly and the cleaned reads were aligned using HiSat2; transcriptome assembly was conducted using Cufflinks.

Genome annotation

The repeat elements in the *B. calyciflorus* genome were identified prior to gene prediction. Specifically, repeats and transposable elements (TEs) were annotated using RepeatModeler v1.0.10 (http://www.repeatmasker.org/), RepeatMasker v4.0.7 (http://www.repeatmasker.org/), and TEclass v2.1.3 (Abrusán et al. 2009). A *de novo* repeat library was constructed using RepeatModeler, and an unclassified repeat library was classified with the latest GIRI repbase (http://www.girinst.org) using TEclass. Kimura distances between genome copies and the consensus library TE were calculated using Repeat Landscape in RepeatMasker (Chalopin et al. 2015), to compare the distribution of TE copies in the *B. calyciflorus* genome with that of the bdelloid rotifer *A. vaga*.

The initial gene model was predicted using SNAP (Korf, 2004), GeneMark-ES (Her-Hovhannisyan et al. 2008), and AUGUSTUS v.3.2.1 (Stanke et al., 2008). The MAKER v.2.3.1 pipeline (Holt and Yandel, 2011) was used to annotate the assembled genome. To obtain additional evidence for these genes, RNA-Seq transcripts, expressed sequencing tags (ESTs), and genes from closely related species were mapped to the assembled genome using Exonerate v2.2.0 (Slater & Birney 2005). Non-coding RNAs (ncRNAs) were annotated using two programs: tRNAscan-SE v1.23 (http://lowelab.ucsc.edu/tRNAscan-SE/) was used to annotate tRNAs, and Infernal v1.1.2 (http://eddylab.org/infernal/) with the Rfam v12.2 database was used to search for all other RNAs. For functional annotation of the predicted genes, BLAST analyses were performed against an NCBI non-redundant database and

against a UniRef90 database from UniProt. Gene Ontology (GO) analysis was performed with InterProScan using Blast2GO_cli v1.1.5 (https://www.blast2go.com/). After genome annotation, we compared the orthologous genes of *B. calyciflorus* with the genomes of other closely related species, including those of the bdelloid rotifer *A. vaga* (GenBank Accession no. **GCA_000513175.1**) (Flot et al. 2013), the water flea *Daphnia pulex* (GenBank Accession no. **GCA_000187875.1**) (Colbourne et al., 2011), the spiralian *Helobdella robusta* (Simakov et al. 2013) (GenBank Accession no. **GCA_000326865.1**), and the oyster *Crassostrea gigas* (GenBank Accession no. **GCA_000297895.1**) (Zhang et al. 2012), using OrthoVenn (Wang et al. 2015).

Construction of a genome browser and a local BLAST database

We constructed a browser for the *B. calyciflorus* genome. The genome browser was implemented with J-browse and WebApollo for convenient access and maintenance. We also constructed an accessible local *B. calyciflorus* BLAST database using SequenceServer 1.0.9 (http://www. sequenceserver.com/).

Results

De novo* genome assembly of *B. calyciflorus

Information regarding the raw sequence reads generated by the HiSeq 2500 instrument and the cleaned reads that passed QC is summarized in **Table 1**. We obtained a total of 814,055,900 sequence reads from the PE and MP libraries, yielding a total sequence length of 137,451,315,174 bp. After preprocessing the raw reads, including the removal of contaminants **(Suppl. Fig. 1)**, 358,418,698 (44%) of the total sequences remained, with a total sequence length of 44,834,410,032 (32.6%). K-mer analysis (default k=21) using GenomeScope estimated the *B. calyciflorus* genome size at 128,529,354 bp from the PE800 library **(Fig. 2)**, which is approximately 87.6% of the genome size (0.15 pg) as measured by flow cytometry (Stelzer 2011). The final assembled *B. calyciflorus* genome is summarized in **Table 2**. In the final assembly, 1,041 scaffolds (>1 kb) were included, with a total length of 129,636,934 bp. The N50 was 786,674 bp and the GC content was 24.24%. CEGMA determined that 94.35% of the 248 genes comprising the essential eukaryotic core of genes were present and BUSCO determined that 88% of the complete single-copy genes based on a metazoan model set were present **(Table 3)**.

Genome annotation

From the assembled sequences, repeat sequences were identified in the genome of *B. calyciflorus* **(Table 4)**. Repeat sequences accounted for 21.01% of the genome, and the DNA transposons (13.02%) were the most abundant repeat type. For genome annotation, a total of 74,502 genes were predicted based on the information of the initial gene model **(Supp. Table S1)**. Transcriptome information obtained by RNA-seq provides important evidence for understanding gene structure, which is important for gene prediction. We developed transcriptome data from the whole body using the RNA-seq technique. A total of 13,361,474 sequence reads were used for transcriptome assembly and gene prediction **(Table 5)**. After *de novo* transcriptome assembly of the cleaned sequences, we obtained 48,480 contigs with an N50 value of 1.3 kb **(Suppl. Table S2)**. After manual curation of structural annotations, we confirmed the final set of 16,114 annotated genes in the *B. calyciflorus* genome **(Table 6)**. A total of 1,063 tRNAs were also identified **(Suppl. Table S3),** and 11,563 genes were functionally annotated.

Comparative analysis with other species

We compared the repeat elements in the rotifer genomes of *B. calyciflorus* and *A. vaga* **(Fig. 3)**. While DNA transposons were the most abundant type in the *B. calyciflorus* genome, LTRs (8.32%) were the most abundant type found in *A. vaga* **(Suppl. Table S5)**. Interestingly, SINEs were not identified in *B. calyciflorus*. Using the annotated genomes, we constructed orthologous gene clusters of *B. calyciflorus* by comparing with the genomes of the bdelloid rotifer *A. vaga* (Flot et al. 2013), the water flea *D. pulex* (Colbourne et al. 2011), the spiralian *H. robusta* (Simakov et al. 2013), and the oyster *C. gigas* (Zhang et al. 2012). The *B. calyciflorus* genome contained 7,287 orthologous genes. In total, 3,435 gene families were shared among all five species, and 782 genes were *B. calyciflorus* specific **(Fig. 4)**. *B. calyciflorus* shared 5,581 (76.59%) gene families with *A. vaga,* 5,015 (68.82%) with *H.robusta*, 5,175 (71.02%) with *C. gigas*, and 4,776 (65.54%) with *D. pulex* **(Fig. 4)**. We constructed a J browser (http://tigriopus.synology.me:8080/apollo/5/jbrowse/index.html) for the *B. calyciflorus* genome, which displays all information regarding the final assembly and gene annotation. The genome browser also provides all information regarding the coding genes, including evidence tracks for structural annotation based on RNA-Seq and interspecific comparisons **(Suppl. Fig. 2)**.

Discussion

We have developed the whole genome assembly of the freshwater monogonont rotifer *B. calyciflorus* based on three PE libraries and three MP libraries, using the Illumina HiSeq2500 platform. After preprocessing the raw reads, we were able to use 44.0% of the sequence reads for genome assembly **(Table 1)**, which appeared to be quite a low amount. Since the introduction of NGS for whole genome sequencing, genome studies of non-model species have increased. However, for many non-model species of interest, challenges remain in isolating the contaminants from sequence reads, due to the presence of food organisms or parasites (Kumar et al. 2013). Since we used the whole bodies for genome sequencing, it was crucial to remove the contaminant sequences from *B. calyciflorus* sequence reads. Due to the organism's small size, absolute isolation of the contaminant-free genome of *B. calyciflorus* is extremely difficult, despite the antibiotic treatment and gut contents removal being performed prior to DNA extraction. Therefore, many cleaned sequence reads were removed to eliminate contamination by other organisms, which resulted in the usage of a low amount (44.0%) of the sequence reads **(Suppl. Fig. 1)**.

For *de novo* assembly of the *B. calyciflorus* genome, we tested several sequence assemblers to select the best assembler. Platanus v1.2.4 (Kajitani et al. 2014) and HaploMerger2 (Huang et al. 2012b; Huang et al. 2017) were specially designed for highly heterozygous diploid genomes, and ALLPATHS-LG (Gnerre et al. 2011) is a *de novo* assembler that is useful for large genomes. Platanus v1.2.4 assembles the sequence reads into contigs based on de Bruijn graphs with automatically optimized k-mer size (Kajitani et al. 2014), while HaploMerger2 provides an automated pipeline for streamlining the post-assembly refinement operations for polymorphic diploid assemblies based on a LASTZ-chainNet approach (Huang et al. 2012; Huang et al. 2017). To select the most suitable assembler, QUAST (Gurevich et al. 2013) was used as a tool to evaluate the efficiencies of the assemblers for *B. calyciflorus* genome assembly **(Suppl. Fig. 3 and Suppl. Table S4)**. As seen in **Suppl. Fig 3**, the cumulative contig length curve indicated that the combination of Platanus and HaploMerger2 showed the highest quality of assembly, compared to the quality using ALLPATHS-LG (Gnerre et al. 2011) or Platanus only. The statistics values in **Suppl. Table S4** also supported these results.

We found that the two rotifer species *B. calyciflorus* and *A. vaga* **(Table 4 and Suppl. Table S5)** showed very different patterns of TE distribution. SINEs are non-coding transposable elements present at high frequencies in various eukaryotic genomes. Interestingly, the genome of *B. calyciflorus* had no SINE sequences. Most unicellular

eukaryotes and *Drosophila* species are also known to lack SINEs in their genomes. Indeed, it was speculated that certain properties of host genomes, such as small genome size, were associated with the failure to maintain SINEs after emergence (Kramerov & Vassetzky 2011). The *Drosophila* genome is relatively small (~120 Mb) (Adams et al. 2000), which can affect the mechanisms counteracting mobile element expansion (Kramerov & Vassetzky 2011). Since the genome size of *B. calyciflorus* is similar to that of *Drosophila*, this speculation is likely to be applied to this rotifer genome. Furthermore, in *A. vaga*, only 0.02% of the genome consists of SINEs, despite the genome being twice the size of the *B. calyciflorus* genome **(Suppl. Table S5)**. The Kimura distance **(Fig. 3)** showed that the overall TE distribution in *B. calyciflorus* was more recently diverged than those in *A. vaga*.

Conclusion

We assembled and annotated the genome of the freshwater monogonont rotifer *B. calyciflorus*, which is a potential invertebrate model species for evolution and ecotoxicology. The estimated genome length is 129.6 Mb, based on 1,041 scaffolds with an N50 of 786.6 kb and 16,114 annotated genes. This genome assembly of *B. calyciflorus* may provide a framework for studies to address important questions on the evolution of monogonont rotifers.

Availability of supporting data

The datasets supporting the results of this article are available in NCBI. The raw sequencing reads are available at SRA **(SRR6027262-SRR6027267)** and the genome assembly data have been deposited at GenBank under accession no. **NTPY01000000**. The J browser for the *B. calyciflorus* genome is accessible at http://tigriopus.synology.me:8080/apollo/5/jbrowse/index.html.

Competing interests

The authors declare no competing interests.

Author contributions

H-SK, B-YL, and J-SL designed the experiments, analyzed the data, and wrote the manuscript. JH, C-BJ, D-SH, M-CL, H-MK, D-HK, and H-JK performed the experiments. JK, SP, SAJD, AH, and J-SL discussed the experiments and worked through potential problems during their execution.

Acknowledgments

We thank the editor (Dr. Sébastien Renaut) and two anonymous reviewers for their valuable advice on the manuscript. In addition, we thank the rotifer research community for sharing many ideas, and we also thank Prof. Terry Snell (Georgia Institute of Technology, GA, USA) for his advice regarding rotifer genome work. We also thank Wei Zhang for collection, species identification, and provisioning of resting eggs. This work was supported by a grant from the Marine Biotechnology Program (PJT200620, "Genome analysis of marine organisms and development of functional application"), funded by the Ministry of Oceans and Fisheries of Korea, and was also supported by a grant from the National Research Foundation (2015R1D1A1A01059) funded to Heum Gi Park.

References

Adams, MD, Celniker SE, Holt RA et al. (2000) The genome sequence of *Drosophila*

melanogaster. *Science*, **287**, 2185-2195.
Abrusán G, Grundmann N, DeMester L et al. (2009) TEclass: A tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics*, **25**, 1329–1330.
Arndt H (1993) Rotifers as predators on components of the microbial web (bacteria, heterotrophic flagellates, ciliates) – a review. *Hydrobiologia*, **255**, 231–246.
Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
C. elegans sequencing consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, **282**, 2012-2018.
Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ et al. (2011) The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**, 555-561.
Dahms H-U, Hagiwara A, Lee J-S (2011) Ecotoxicology, ecophysiology, and mechanistic studies with rotifers. *Aquat. Toxicol.*, **101**, 1–12.
Declerck SA, Malo AR, Diehl S et al. (2015) Rapid adaptation of herbivore consumers to nutrient limitation: Eco-evolutionary feedbacks to population demography and resource control. *Ecol. Lett.* **18**, 553-562.
Declerck SAJ, Papakostas S (2017) Monogonont rotifers as model systems for the study of micro-evolutionary adaptation and its eco-evolutionary implications. *Hydrobiologia*, **796**, 131-144.
Flot JF, Hespeels B, Li X et al. (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature*, **500**, 453-457.
Fussmann GF, Ellner SP, Hairston NG Jr. (2003) Evolution as a critical component of plankton dynamics. *Proc. Roy. Soc. B-Biol. Sci.*, **270**, 1015–1022.
Gilbert JJ, Walsh EJ (2005) *Brachionus calyciflorus* is a species complex: Mating behavior and genetic differentiation among four geographically isolated strains. *Hydrobiologia*, **546**, 257–265.
Gnerre S, Maccallum I, Przybylski D et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA*, **108**, 1513-1518.
Gomez A, Serra M, Carvalho GR, Lunt DH (2002) Speciation in ancient cryptic species complexes: Evidence from the molecular phylogeny of *Brachionus plicatilis* (Rotifera). *Evolution*, **56**, 1431–1444.
Grabherr MG, Haas BJ, Yassour M et al.(2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.*, **29**, 644–652.
Gurevich A, Saveliev V, Vyahhi N et al. (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, **29**, 1072-1075.
Hagiwara A (2007) Development of rotifer strains with useful traits for rearing fish larvae. *Aquaculture*, **268**, 44–52
Hagiwara A, Balompapueng MD, Munuswamy N, Hirayama K (1997) Mass production and preservation of the resting eggs of the euryhaline rotifer *Brachionus plicatilis* and *B. rotundiformis*. *Aquaculture*, **155**, 223–230.
Hagiwara A, Kotani T, Snell TW, Assava-Aree M, Hirayama K (1995) Morphology, reproduction, genetics, and mating behavior of a small, tropical marine *Brachionus* strain (Rotifera). *J. Exp. Mar. Biol. Ecol.*, **194**, 25–37.
Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, **12**, 491.
Huang S, Chen Z, Huang G et al. (2012) HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. *Genome Res.*, **22**, 1581–1588.
Huang S, Kang M, Xu A (2017) HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, **33**, 2577-2579.
Hwang D-S, Dahms H-U, Park H-G, Lee J-S (2013) A new intertidal *Brachionus* and

intrageneric phylogenetic relationships among *Brachionus* as revealed by allometry and *CO1-ITS1* gene analysis. *Zool. Stud.* **52**, 13.
Jeong C-B, Won E-J, Kang H-M, Lee M-C, Hwang D-S, Hwang U-K, Zhou B, Souissi S, Lee S-J, Lee J-S (2016) Microplastic size-dependent toxicity, oxidative stress induction, and p-JNK and p-p38 activation in the monogonont rotifer (*Brachionus koreanus*). *Environ. Sci. Technol.*, **50**, 8849-8857.
Jiang H, Lei R, Ding S-W et al. (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, **15**, 182.
Kajitani R, Toshimoto K, Noguchi H et al. (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.*, **24**, 1384–1395.
Kim R-O, Kim B-M, Jeong C-B, Nelson DR, Lee J-S, Rhee J-S (2013) Expression pattern of entire cytochrome P450 genes and response of defensomes in the benzo[α]pyrene-exposed monogonont rotifer *Brachionus koreanus*. *Environ. Sci. Technol.*, **47**, 13804–13812.
Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics*, **5**, 59.
Kramerov DA, Vassetzky NS (2011) Origin and evolution of SINEs in eukaryotic genomes. *Heredity*, **107**, 487-495.
Kumar S, Jones M, Koutsovoulos G et al. (2013) Blobology: exploring raw genome data for contaminants symbionts, and parasites using taxon-annotated GC coverage plots. *Front. Genet.*, **4**, 1-12.
Marcais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**, 764–770.
Mark Welch D, Mark Welch J, Meselson M (2008) Evidence for degenerate tetaploidy in bdelloid rotifers. *Proc. Natl. Acad. Sci. USA*, **105**, 5145-5149.
Mark Welch D, Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, **288**, 1211-1215.
Marten B, Christiaan VH, Hans JJ (2010) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, **27**, 578-579.
Ortells R, Snell TW, Gomez A, Serra M (2000) Patterns of genetic differentiation in resting egg banks of a rotifer species complex in Spain. *Arch. Hydrobiol.*, **149**, 529–551
Ortells R, Gomez A, Serra M (2003) Coexistence of cryptic rotifer species: ecological and genetic characterisation of *Brachionus plicatilis*. *Freshw. Biol.*, **48**, 2194–2202
Papakostas S, Michaloudi E, Proios K et al. (2016) Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: Evidence from a rotifer cryptic species complex. *Syst. Biol.*, **65**, 508-524.
Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061–1067.
Preston BL, Snell TW, Robertson TL et al. (2000) Use of freshwater rotifer *Brachionus calyciflorus* in screening assay for potential endocrine disruptors. *Environ. Toxicol. Chem.*, **19**, 2923–2928.
Segers H (2008) Global diversity of rotifers (Rotifera) in freshwater. *Hydrobiologia*, **595**, 49-59
Serra M, Snell TW (2009) Sex loss in rotifers. In: Schön I, Martens K, Van Dijk P (eds) Lost sex. The evolutionary biology of parthenogenesis. Springer Academic, Dordrecht. pp 281–294.
Sha J, Wang Y, Chen H et al. (2015) Using population demographic parameters to assess impacts of two polybrominated diphenyl ethers (BDE-47, BDE-209) on the rotifer *Brachionus plicatilis*. *Ecotoxicol. Environ. Saf.*, **119**, 106–115.
Simakov O, Marletaz, F, Cho SJ et al. (2013) Insights into bilaterian evolution from three spiralian genomes. *Nature*, **493**, 526-531.
Simao FA, Waterhouse RM, Ioannidis P et al. (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.
Slater GS, Birney E (2005) Automated generation of heuristics for biological sequence

comparison. *BMC Bioinformatics*, **6**, 31.
Smith HA, Snell TW (2012) Rapid evolution of sex frequency and dormancy as hydroperiod adaptations. *J. Evol. Biol*., **25**, 2501–2510.
Snell TW, Janssen CR (1995) Rotifers in ecotoxicology: a review. *Hydrobiologia*, **313**, 231–247.
Snell TW, Kubanek JM, Carter WE et al. (2006) A protein signal triggers sexual reproduction in *Brachionus plicatilis* (Rotifera). *Mar. Biol.*, **149**, 763–773.
Snell TW, Brogdon SE, Morgan MB (2003) Gene expression profiling in ecotoxicology. *Ecotoxicology*, **12**, 475–483.
Snell TW, Carmona MJ (1995) Comparative toxicant sensitivity of sexual and asexual reproduction in the rotifer *Brachionus calyciflorus*. *Environ. Toxicol. Chem.*, **14**, 415-420.
Snell TW, Moffat BD, Janssen C et al. (1991) Acute toxicity tests using rotifers: IV. Effects of cyst age, temperature, and salinity on the sensitivity of *Brachionus calyciflorus*. *Ecotoxicol. Environ. Saf*., **21**, 308–317.
Snell TW, Persoone G (1989a) Acute toxicity bioassays using rotifers. I. A test for brackish and marine environments with *Brachionus plicatilis*. *Aquat. Toxicol.*, **14**, 65-80.
Snell TW, Persoone G (1989b) Acute toxicity tests using rotifers. II. A freshwater test with *Brachionus rubens*. *Aquat. Toxicol*., **14**, 81-92.
Stanke M, Diekhans M, Baertsch R et al. (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, **24**, 637–644.
Stelzer CP (2011) A first assessment of genome size diversity in monogonont rotifers. *Hydrobiologia*, **662**, 77-82.
Stelzer C-P, Snell TW (2003) Induction of sexual reproduction in *Brachionus plicatilis* (Monogononta, Rotifera) by a density-dependent chemical cue. *Limnol. Oceanogr.,* **48**, 939–943.
Ter-Hovhannisyan V, Lomsadze A, Chernoff YO et al. (2008) Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res.*, **18**, 1979–1990.
Trapnell C, Williams BA, Pertea G et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.*, **28**, 511–515.
Vurture G, Sedlazeck F., Nattestad M et al. (2017) GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, **33**, 2202-2204.
Wallace RL, Snell TW, Ricci C et al. (2006) Rotifera 1: Biology, Ecology and Systematics. Backhuys Publishers, Leiden
Won E-J, Kim R-O, Kang H-M, Kim H-S, Han J, Lee YH, Hwang D-S, Hwang U-K, Zhou B, Lee S-J, Lee J-S (2016) Adverse effects, expression of the *Bk-CYP3045C1* gene, and activation of the ERK signaling pathway in the water accommodated fraction-exposed rotifer. *Environ. Sci. Technol.*, **50**, 6025-6035.
Won E-J, Han J, Kim D-H, Dahms H-U, Lee J-S. (2017) Rotifers in ecotoxicology. In: Hagiwara A, Yoshinaga T (Eds.) Rotifers: Aquaculture, Ecology, Gerontology, and Ecotoxicology. Springer, Tokyo. pp. 149-176.
Xu H, Luo X, Qian J et al. (2012) FastUniq: A fast *de novo* duplicates removal tool for paired short reads. *PLoS One*, **7**, e52249
Yoshinaga T, Kaneko G, Kinoshita S et al. (2003) The molecular mechanisms of life history alterations in a rotifer: A novel approach in population dynamics. *Comp. Biochem. Physiol. B*, **136**, 715–722.
Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X et al. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, **490**, 49-54.

Table 1. Summary of the sequencing libraries and sequencing reads generated using the HiSeq2500 platform

		Raw data		Cleaned reads (%)				Coverage	SRA
Library	**Type**	**No. reads**	**Total length (bp)**	**No. reads**		**Total length (bp)**		**(x)**	**Accession no.**
PE300	Paired-end	240,840,750	34,387,258,879	75,761,556	(31.5%)	9,863,935,114	(28.7%)	75.8	SRR6027264
PE500	Paired-end	139,325,772	32,617,152,975	35,470,504	(25.5%)	6,520,133,476	(20.0%)	50.1	SRR6027265
PE800	Paired-end	71,591,200	17,005,298,876	57,032,684	(79.7%)	11,156,881,506	(65.6%)	85.8	SRR6027266
MP2kb	Mate pair	77,678,406	11,436,553,275	70,091,194	(90.2%)	6,373,312,334	(55.7%)	49	SRR6027267
MP5kb	Mate pair	148,798,992	21,946,433,889	65,375,618	(43.9%)	5,936,920,528	(27.1%)	45.6	SRR6027262
MP10kb	Mate pair	135,820,780	20,058,616,280	54,687,142	(40.3%)	4,983,227,074	(24.8%)	38.3	SRR6027263
Total		814,055,900	137,451,314,174	358,418,698	(44.0%)	44,834,410,032	(32.6%)	345.9	

Table 2. Summary statistics of the assembled *B. calyciflorus* genome

Statistic	Value
Number of scaffolds	1,041
Length of scaffolds (bp)	129,636,934
N50 (bp)	786,674
Largest scaffold (bp)	3,647,490
Gap (%)	6.41
GC content (%)	24.24

Table 3. Assessment of assembly completeness

Program	Category	Percentage
CEGMA	Complete	94.35%
	Partial	96.77%
	Complete Single-copy	88.00%
	Complete Duplicated-copy	2.70%
	Fragmented	2.00%
	Missing	7.30%

Table 4. Composition of repetitive sequences in the *B. calyciflorus* genome

	Class	Length	Percent
DNA	DNA/CMC-EnSpm	69,973	0.05%
	DNA/CMC-Transib	22,315	0.02%
	DNA/Ginger	167,736	0.13%
	DNA/hAT	71,555	0.06%
	DNA/hAT-Ac	762,876	0.59%
	DNA/hAT-Tip100	54,519	0.04%
	DNA/Maverick	651,704	0.50%
	DNA/Merlin	100,786	0.08%
	DNA/MULE-MuDR	1,087,150	0.84%
	DNA/MULE-NOF	77,074	0.06%
	DNA/PiggyBac	68,888	0.05%
	DNA/TcMar-Pogo	106,344	0.08%
	DNA/TcMar-Sagan	158,102	0.12%
	DNA/TcMar-Tc1	815,564	0.63%
	DNA/TcMar-Tc2	30,506	0.02%
	DNA/other	12,684,806	9.76%
LINE	LINE/CR1	894,288	0.69%
	LINE/I	37,746	0.03%
	LINE/Jockey	282,338	0.22%
	LINE/L1-Tx1	141,136	0.11%
	LINE/L2	552,761	0.43%
	LINE/L2-Hydra	73,523	0.06%
	LINE/LOA	15,128	0.01%
	LINE/Penelope	290,429	0.22%
	LINE/Proto2	167,379	0.13%
	LINE/RTE-BovB	196,565	0.15%
	LINE/other	1,202,319	0.92%
LTR	LTR/Copia	79,420	0.06%
	LTR/DIRS	42,391	0.03%
	LTR/Gypsy	1,588,974	1.22%
	LTR/other	887,494	0.68%
etc	rRNA	235	0.00%
	Satellite	62,494	0.05%
	Simple_repeat	3,317,781	2.55%
	Unknown	551,095	0.42%
	Total	**27,313,394**	**21.01%**

Table 5. Summary of RNA-Seq reads

Statistic	Value
Number of genes	16,114
Length of genes (bp)	26,187,845
Average length (bp)	1,625
Largest gene (bp)	51,513
GC content (%)	30.84

Table 6. Gene annotation statistics for the assembled *B. calyciflorus* genome

	Raw reads		Cleaned reads	
Tissue	No. of Reads	Read Length (bp)	No. of Reads	Read Length (bp)
Whole body	60,072,864	14,167,735,623	13,361,474	2,700,231,100

Figure legends

Fig. 1. External morphology of *B. calyciflorus*.

Fig. 2. Graph of the k-mer distribution (K=21) generated from the PE800 library using GenomeScope. The big peak at the coverage of ~ 60 in the graph is the homozygous portion of the genome, which accounts for the strands of the DNA having identical 21-mers. The smaller shoulder to the left of the peak corresponds to the heterozygous portion of the genome, which accounts for the strands of the DNA having different 21-mers. If the genome is highly heterozygous, the height of the shoulder peak would be closer to that of the homozygous peak.

Fig. 3. Analysis of the TE copy divergence in two different rotifers (A) *B. calyciflorus* and (B) *A. vaga*, based on the Kimura distance. The Y axis represents the genome coverage for each type of TE (DNA transposon, SINE, LINE, LTR retrotransposons), and the X axis represents the k-value.

Fig. 4. Venn diagram of the orthologous clusters from five invertebrate species, using OrthoVenn.

Fig. 1

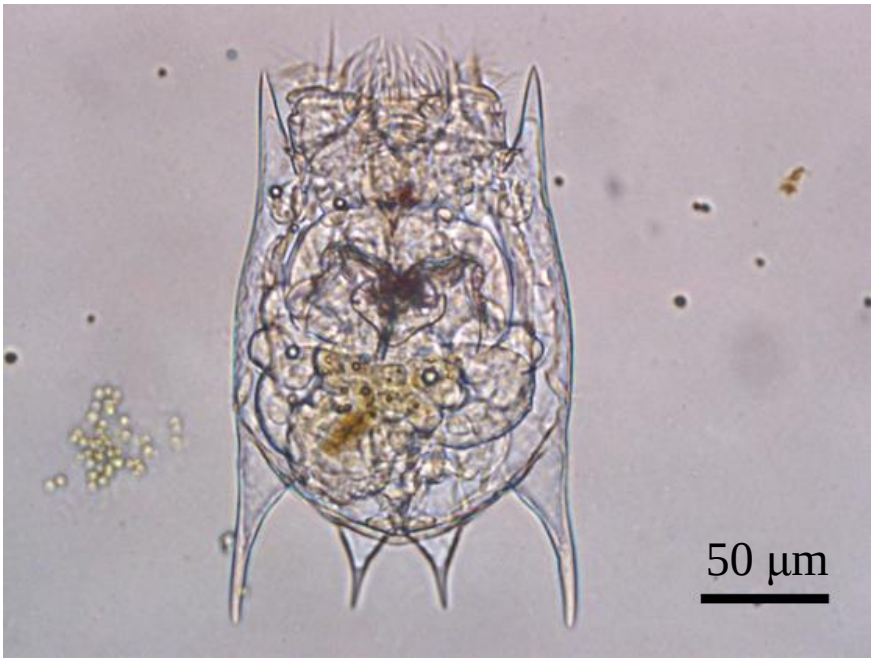
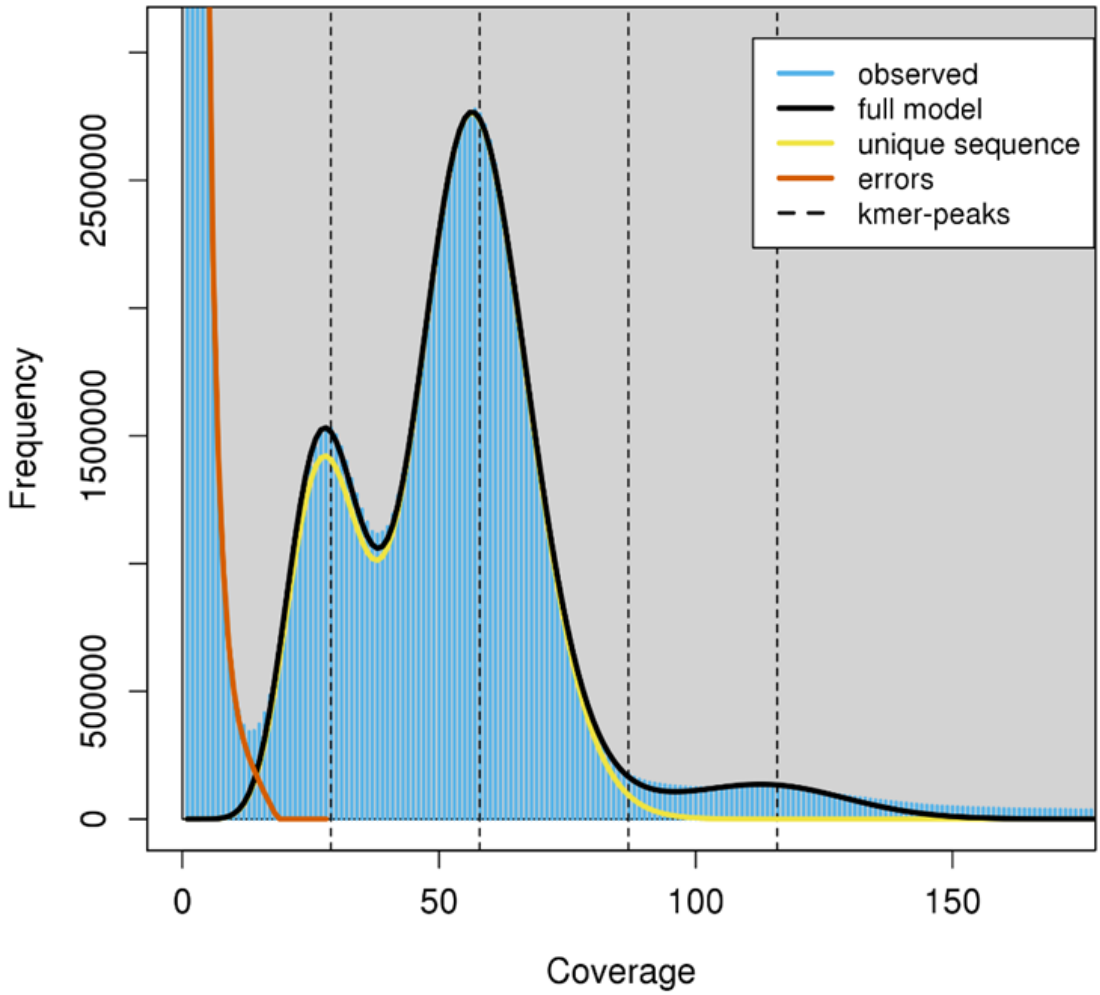
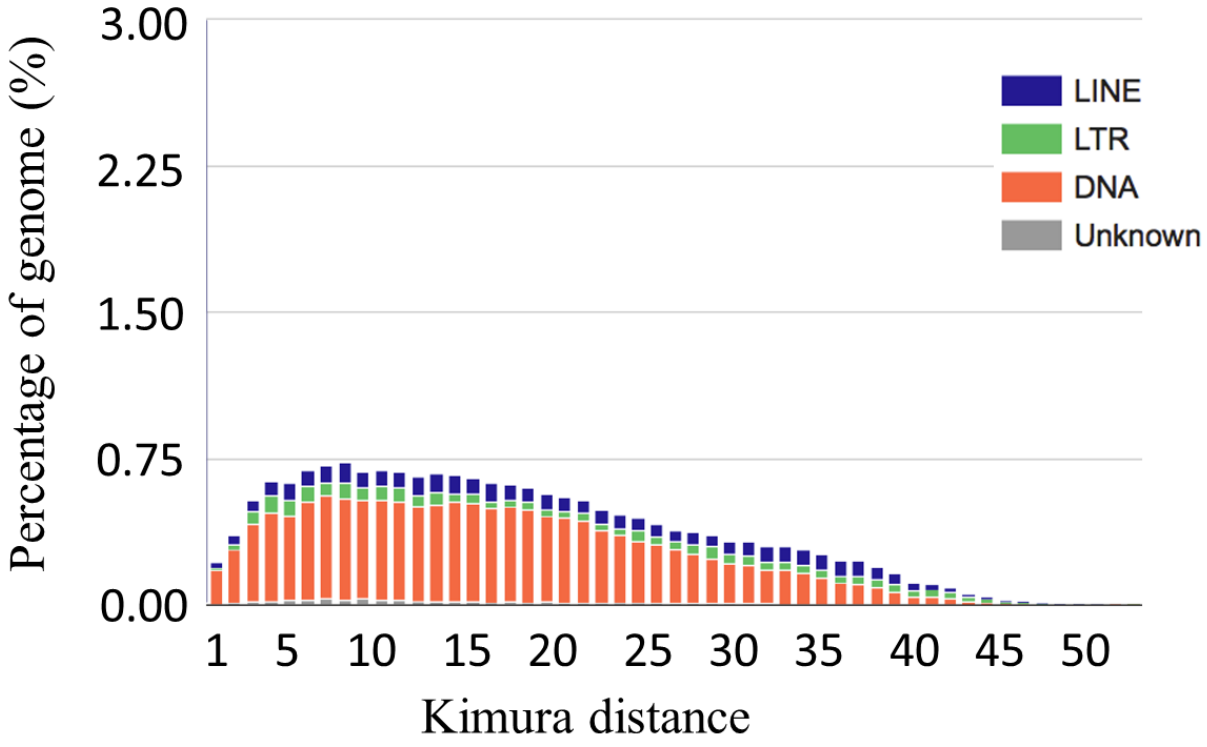


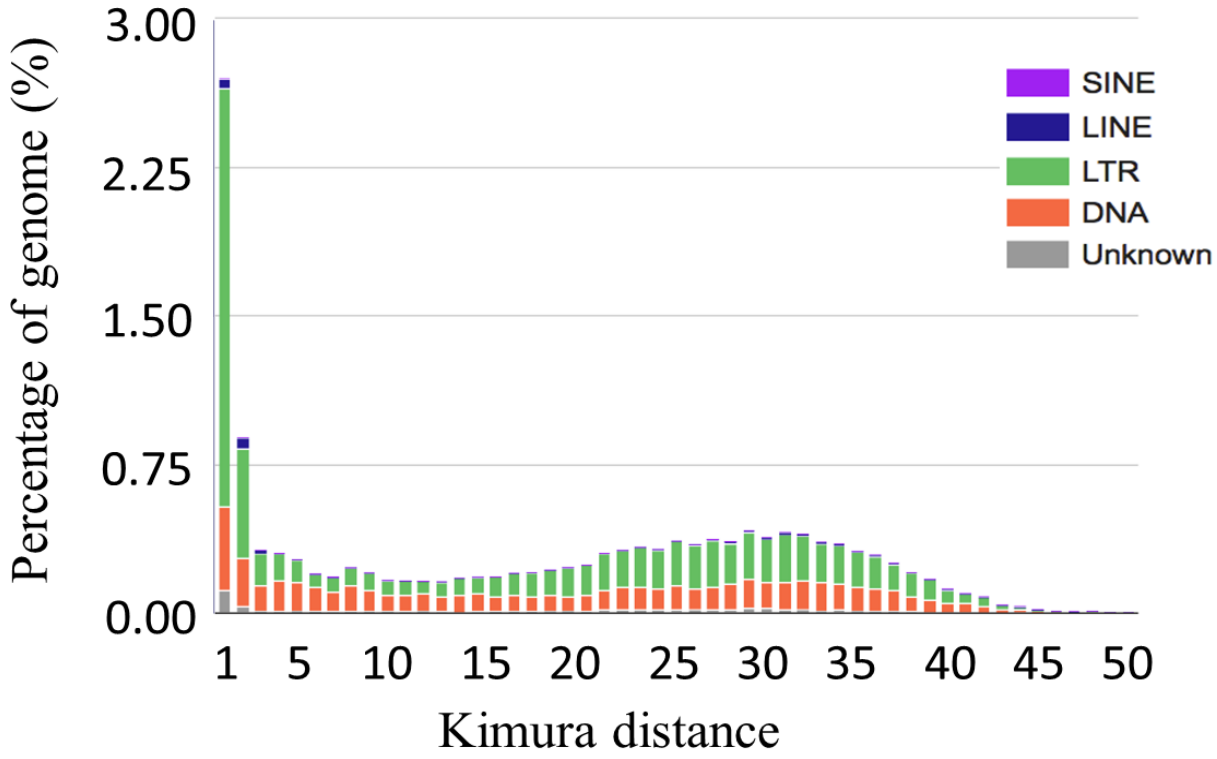
Fig. 2

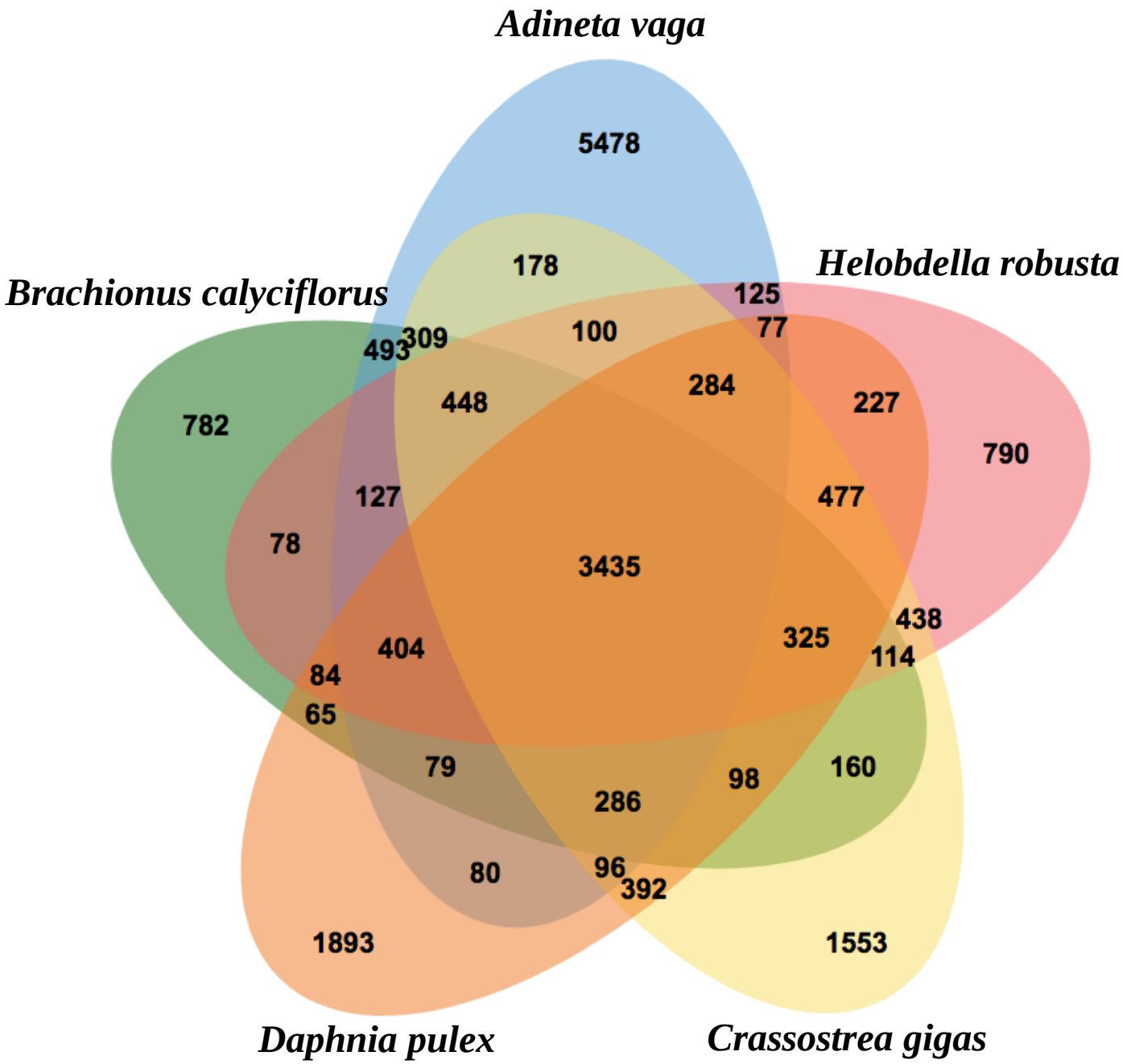

observed
full model
unique sequence
errors
kmer-peaks
Frequency
0
500000
1500000
2500000
0
50
100
150
Coverage

A) *Brachionus calyciflorus*



B) *Adineta vaga*




Adineta vaga
Brachionus calyciflorus
Helobdella robusta
Daphnia pulex
Crassostrea gigas
5478
178
125
77
100
309
493
448
284
227
782
790
127
477
78
3435
438
404
325
114
84
65
79
98
160
286
80
96
392
1893
1553

Suppl. Table S1. Initial gene prediction based on various software

Method		Number	Average coding DNA sequence (CDS) length (bp)
de novo	AUGUSTUS	18,575	1,484
	SNAP	30,445	1,120
	GeneMark	25,482	1,337
Final gene set	Maker	16,114	1,625

Suppl. Table S2. Statistics of RNASeq de novo assembly

Total length	46,718,577
No. of contigs	48,480
GC level	42.90%
N50 of scaffolds (bp)	1,314
Longest scaffolds (bp)	23,361

Suppl. Table S3. Summary of tRNAs

tRNAs decoding Standard 20 AA	1,035
Selenocysteine tRNAs (TCA)	0
Possible suppressor tRNAs (CTA,TTA)	0
tRNAs with undetermined/unknown isotypes	0
Predicted pseudogenes	28
Total tRNAs	1,063

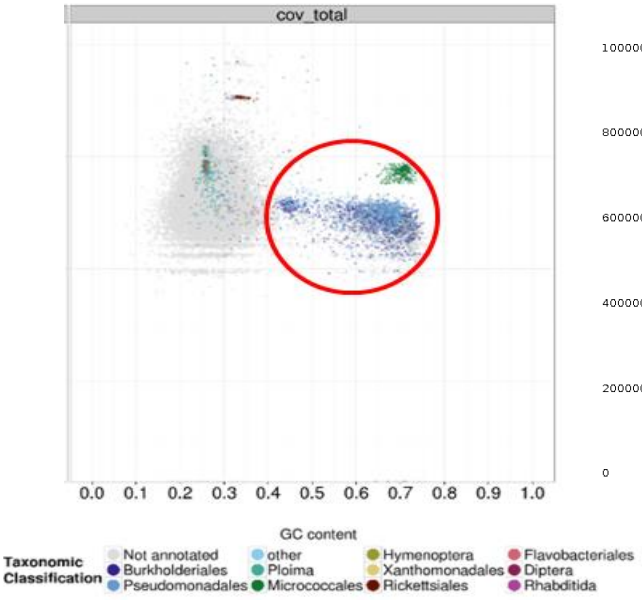
Suppl. Table S4. Comparison of quality assessment for genome assemblies

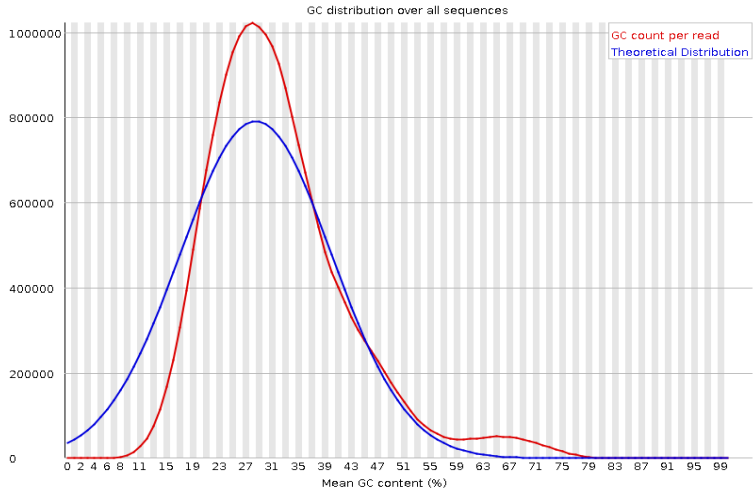
Statistics without reference	Platanus + HM2	Platanus	ALLPATHS-LG
# contigs	1041	5237	49803
# contigs (>= 0 bp)	1041	5237	49803
# contigs (>= 1000 bp)	1041	5237	47714
# contigs (>= 5000 bp)	533	879	3526
# contigs (>= 10000 bp)	432	576	1150
# contigs (>= 25000 bp)	302	439	332
# contigs (>= 50000 bp)	259	372	146
Largest contig	4241357	2603026	4453383
Total length	129636934	134532176	154248041
Total length (>= 0 bp)	129636934	134532176	154248041
Total length (>= 1000 bp)	129636934	134532176	152239261
Total length (>= 5000 bp)	128604942	127054453	69812901
Total length (>= 10000 bp)	127881994	124982599	53981672
Total length (>= 25000 bp)	125882128	122818104	41817174
Total length (>= 50000 bp)	124408248	120320012	35354240
N50	785733	460773	4043
N75	401523	191067	1819
L50	48	81	5162
L75	106	191	20186
GC (%)	25.86	26.49	39.8
Misassemblies			
Unaligned			
Mismatches			
# N's	8252537	14338451	11111636
# N's per 100 kbp	6365.88	10658	7203.75
Genome statistics			
NG50	785733	492152	5879
NG75	401523	213219	2544
LG50	48	76	2638
LG75	106	174	11657
Predicted genes			
# predicted genes (unique)	24358	25257	29637
# predicted genes (>= 0 bp)	93504	96300	74326
# predicted genes (>= 300 bp)	32860	33697	43980
# predicted genes (>= 1500 bp)	3215	2831	3609
# predicted genes (>= 3000 bp)	493	407	481

Suppl. Table S5. TE distribution of a bdelloid rotifer *A. vaga* (Flot et al 2013)

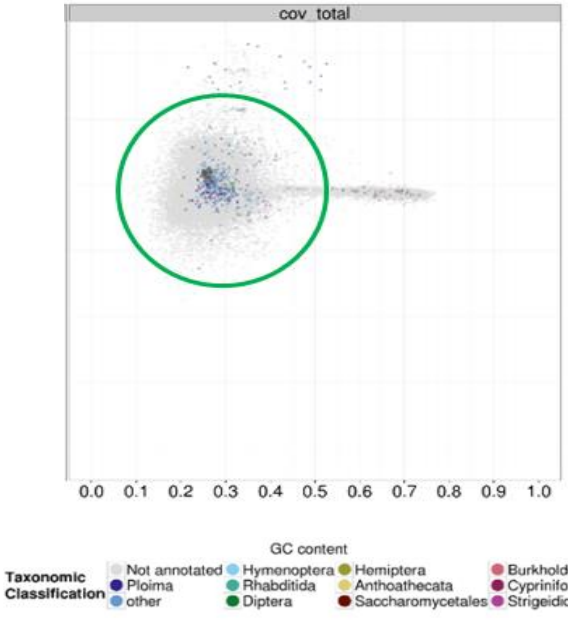
	Class	Length	Percent
DNA	DNA/PiggyBac	34,670	0.02%
	DNA/TcMar-Mariner	65,361	0.03%
	DNA/other	10,388,495	4.76%
LINE	LINE/I-Jockey	198,217	0.09%
	LINE/R2	24,864	0.01%
	LINE/other	470,208	0.22%
LTR	LTR/other	18,139,263	8.32%
SINE	SINE/other	35,175	0.02%
ETC	Retro	156,876	0.07%
	rRNA	78,091	0.04%
	Simple_repeat	4,063,979	1.86%
	Unknown	1,267,684	0.58%
Total		**34,922,883**	**16.01%**

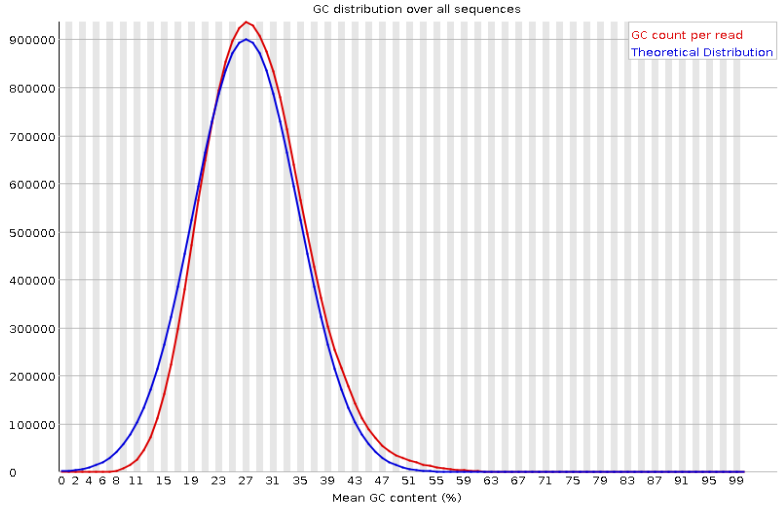
(A)



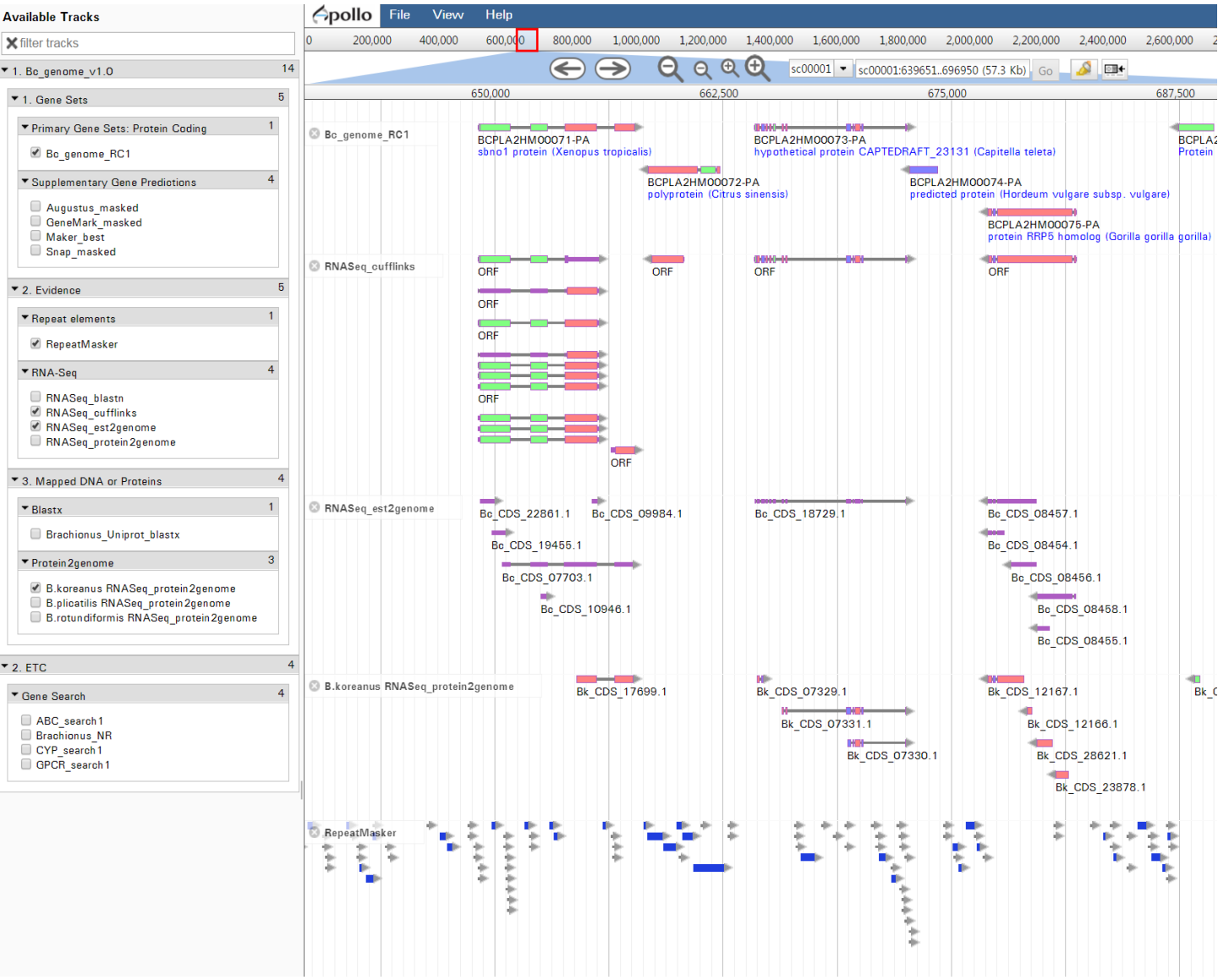


(B)

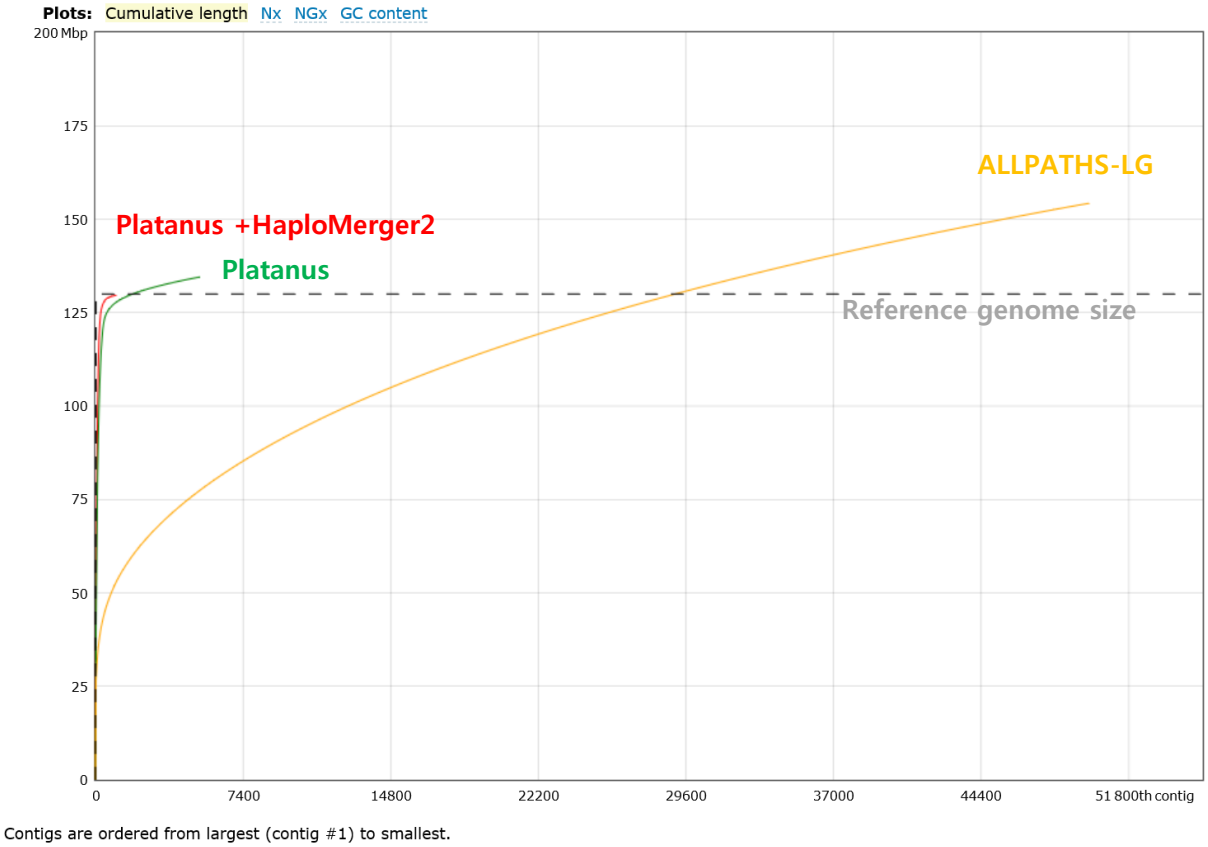




Suppl. Fig. 1. Taxon-annotated GC-coverage plots (blobplots) of (A) before and (B) after contaminant removal in *B. calyciflorus* genome assemblies. Each contig/scaffold in the assembly is represented by a circle and colored dot according to the best match to taxonomically annotated sequence databases (see figure legends) and distributed according to the GC proportion (x-axis) and read coverage (y-axis). The cluster of contigs in a red circle (A) is likely from the presence of contaminants in the sample. The contigs in a green circle (B) shows the removal of contaminant sequences, and the scaffolding of contigs into long contiguous sequences. The graphs on the right side shows the GC distribution over all sequences. After contaminant removal, the graph (red line) is closer to that of the theoretical distribution (blue line).



Suppl. Fig. 2. Screenshot of the genome browser window of *B. calyciflorus* with annotated information.



Suppl. Fig. 3. Cumulative assembly curves showing the relationship between the number of scaffolds (x-axis) and the cumulative span of each assembler (y-axis). Higher-quality assemblies are represented by an almost vertical line, indicating that a relatively small number of scaffolds is required to reach the final genome span; conversely, a long tail indicates that the assembly includes a large number of smaller scaffolds.