



OPEN

DATA DESCRIPTOR

The first high-quality genome assembly and annotation of *Patiria pectinifera*

Jaehoon Jung^{1,2}, So Yun Jhang^{1,3}, Bongsang Kim^{1,2}, Bomin Koh^{1,2}, Chaeyoung Ban¹,
Hyojung Seo¹, Taeseo Park⁴, Won-Jae Chi⁵, Soonok Kim⁵, Heeбал Kim^{1,2,3} & Jaewoong Yu¹✉

The blue bat star, a highly adaptive species in the East Sea of Korea, has displayed remarkable success in adapting to recent climate change. The genetic mechanisms behind this success were not well-understood, prompting our report on the first chromosome-level assembly of the *Patiria* genus. We assembled the genome using Nanopore and Illumina sequences, yielding a total length of 615 Mb and a scaffold N50 of 24,204,423 bp. Hi-C analysis allowed us to anchor the scaffold sequences onto 22 pseudochromosomes. K-mer based analysis revealed 5.16% heterozygosity rate of the genome, higher than any previously reported echinoderm species. Our transposable element analysis exposed a substantial number of genome-wide retrotransposons and DNA transposons. These results offer valuable resources for understanding the evolutionary mechanisms behind *P. pectinifera*'s successful adaptation in fluctuating environments.

Background & Summary

Adapting to rapidly changing environments is a critical issue in evolutionary biology amidst global warming. The East Sea, known for its extreme seasonal fluctuations in temperature (5–15 °C for the coasts of the Korean Peninsula) and salinity (30~34 permil)^{1,2}, is one of the world's ecoregions most heavily impacted by climate change^{3,4}. Among the species inhabiting this region is the blue bat star (*Patiria pectinifera*, Fig. 1a), which thrives in a wide range of environments due to its adaptability and resilience⁵. While recent studies have shown that transposable elements (TEs) can accelerate genetic responses to stress, the underlying mechanisms remain unclear^{6,7}.

Besides its adaptation ability, *P. pectinifera* is commonly used as a model organism in developmental biology due to its prevalence, echinoderm's intermediate phylogenetic position with respect to chordates and protostomes, and ease of maintenance in the laboratory^{8–10}. Furthermore, large and transparent oocytes, which remain viable after removal from the gonads, were used for studies of oocyte maturation, fertilization, and larval development¹¹. A peptide known as "Gonad Stimulating Substance (GSS)", which exhibits properties similar to relaxin, has been identified in *P. pectinifera*¹². This peptide serves as potential evidence of an invertebrate gonadotropin responsible for gamete maturation, exerting its reproductive functions via a G protein-coupled receptor. To date, despite the researchers' attention and biological importance, the genetic background and evolutionary history of *P. pectinifera* have not been fully understood. There have been studies about tissue-specific antimicrobial peptides (AMPs) gene expression in the species using RT-qPCR and RACE PCR¹³, and researchers have found novel genes and peptides from the species^{10,13}. However, its genome has not been constructed thoroughly, needless to say that there was just one whole-genome sequencing dataset for the species in sequence read archive (based on 2023-04-27 release data in SRA, NCBI¹⁴).

This study strives to advance our understanding of the genetic mechanisms facilitating the successful adaptation of *P. pectinifera* in variable environments. We utilized long-read sequencing data and Hi-C techniques to construct a chromosome-level genome assembly for the species, enabling the analysis of its distinctive genomic

¹eGnome, Inc., 26 Beobwon-ro 9-gil, Songpa-gu, Seoul, 05836, Republic of Korea. ²Department of Agricultural and Life Sciences and Research Institute of Population Genomics, Seoul National University, Seoul, Republic of Korea.

³Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, 151-742, Republic of Korea. ⁴Animal Resources Division, National Institute of Biological Resources, Incheon, 22689, Republic of Korea. ⁵Microorganism Resources Division, National Institute of Biological Resources, Incheon, 22689, Republic of Korea. ✉e-mail: jwyu@genome.co.kr

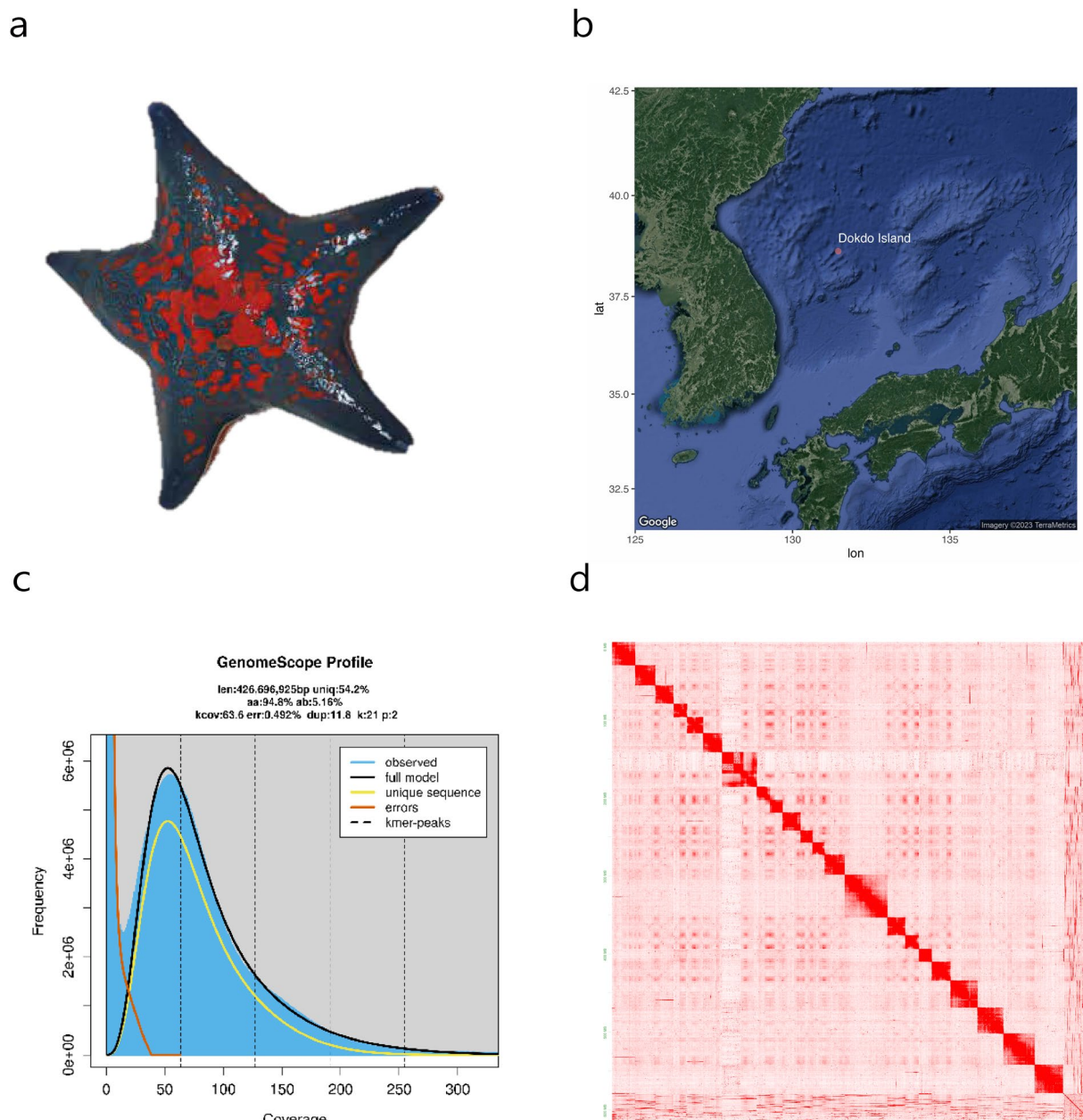


Fig. 1 Chromosome-level assembly of *P. pectinifera* genome. **(a)** Image presentation of *P. pectinifera*. **(b)** Geographical origin of the collected sample. The map has been generated by R package “ggmap”. **(c)** K-mer (21-mer) plot of the species for genome size and heterozygosity estimation. **(d)** Hi-C contact map of representing 22 pseudochromosomes.

feature – high heterozygosity level. We acknowledge that our comparative genome analyses are preliminary and serve as a steppingstone for further research.

Our insights into the genomic features of *P. pectinifera* augment our understanding of the species’ adaptive capabilities and furnish valuable resources for ongoing studies in evolutionary biology. The chromosome-level genome assembly of *P. pectinifera* can act as a baseline for subsequent comparative genomics studies, thereby enhancing our knowledge of the evolution and adaptation of other echinoderm species, even though we acknowledge that in-depth comparative genomic analysis among diverse echinoderm taxa is required for more definitive conclusions.

Methods

Sample information and collection. *P. pectinifera* sample was collected from Dokdo Island (38°14′25.1″N 131°52′13.63″E) in June 2019 for genome assembly analysis (Fig. 1a,b).

DNA library construction and sequencing. The high molecular weight (HMW) genomic DNA was extracted from gonads utilizing PrepGEM universal DNA extraction kit (microGEM, #PUN0100) for the purpose of sequencing with the Oxford Nanopore Technology (ONT) platform. This extraction process, which was adapted from the CTAB method¹⁵ with a 2% polyvinylpyrrolidone (PVP) (comprising 1% of molecular weight [MW] 10,000 and 1% of MW 40,000) (Sigma-Aldrich, Burlington, MA, USA). Subsequently, DNA concentration was ascertained via the Quant-iT PicoGreen[®] assay (Invitrogen, Waltham, MA, USA) while the Synergy HTX Multi-Mode microplate reader (Biotek, Rochester, VT, USA) was used to measure the absorbance at 260 nm and 230 nm (A260/A230). The DNA quality was then verified using gel electrophoresis. Short Read Eliminator Kit (Circulomics, Baltimore, MD, USA) was used to eliminate short genomic fragments under 10 kb. The library was prepared using the ONT 1D ligation Sequencing kit (SQK-LSK109, Oxford Nanopore Technologies, Oxford, UK) with the native barcoding expansion kit (EXP-NBD104) following the manufacturer's protocol. The constructed library was loaded onto a MinION flow cell (FLO-MIN106 R9.4 Flow Cell) (Oxford Nanopore Technologies) and PromethION Flow Cell (FLO-PRO002) (Oxford Nanopore Technologies). Sequencing was performed on a MinION MK1b and PromethION sequencer with MinKNOW software v19.10.1.

As part of the Hi-C library construction procedure, gonad tissue was introduced to 1% formaldehyde to facilitate chromatin fixation. After this, nuclei isolation was conducted following a method¹⁶. The fixed chromatin was then subjected to treatment with HindII-HF (New England BioLabs, Ipswich, MA, USA), following which the 5' overhangs were populated with nucleotides and biotin-14-dCTP (Invitrogen). This allowed for ligation of free blunt ends. Subsequent to ligation, the DNA was purified with unligated DNA ends having biotin effectively removed. Fragmentation and size-based selection were performed on the Hi-C DNA. Finally, the ThruPLEX[®] DNA-seq Kit (Takara Bio USA, Inc, Mountain View, CA, USA) was employed to complete the preparation of the Hi-C library. Fragmentation and size selection were performed to shear the Hi-C DNA. The Hi-C library preparation was performed using the ThruPLEX[®] DNA-seq Kit (Takara Bio USA, Inc, Mountain View, CA, USA).

To sequence with the Illumina platform, genomic DNA was randomly sheared using the Covaris S220 system. The fragmented DNA was utilized for library preparation using the TruSeq DNA PCR-Free kit (Illumina, San Diego, CA, USA), adhering to the manufacturer's instructions. Whole-genome shotgun sequencing was performed using Illumina NovaSeq. 6000 with a 150 bp paired-end. Also, the prepared Hi-C library was sequenced with the same platform. All the obtained reads were quality controlled by trimming adaptor sequences and low-quality reads using Trimmomatic v0.39¹⁷ for Illumina reads and Porechop v0.2.4 (-q 7 option, <https://github.com/rrwick/Porechop>) and NanoFilt¹⁸ (-k 5000 option) for Nanopore reads.

RNA library construction, sequencing and transcriptome construction. Total RNA was extracted from three distinct tissues - the digestive gland, gonad, and stomach of a *P. pectinifera* individual. This extraction was done using TRIzol Reagent (Invitrogen, Waltham, MA, USA) as per the guidelines provided by the manufacturer. The total RNA concentration was then gauged using the Quant-iT[™] RNA Assay Kits (Invitrogen), and the absorbance at 260 nm and 280 nm (A260/A280) was quantified using the Synergy HTX Multi-Mode microplate reader (Biotek, Rochester, VT, USA). The complementary DNA (cDNA) library was then constructed using the cDNA-PCR Sequencing Kit (SQK-PCS109, Oxford Nanopore Technologies), alongside the PCR Barcoding Kit (SQK-PBK004, Oxford Nanopore Technologies), again adhering to the manufacturer's protocol. All three tissue samples underwent sequencing in the same flow cell using the same procedure as the one employed for the previously mentioned DNA sequencing. We also used the Illumina platform to generate high-quality short reads. Using the TruSeq Stranded mRNA Prep kit, we constructed the cDNA library and sequenced it in the Illumina NovaSeq. 6000 with 100-bp paired-end reads.

For transcriptome assembly, filtered long-read RNA sequencing data from 3 tissues were error-corrected with short reads using TALC software¹⁹. The TALC-corrected reads were aligned to the assembled genome using minimap2 splice-aware option²⁰, alignment data were sorted with Samtools v1.9.21, and reference-based transcriptome assembly was performed using StringTie²².

Genome Size Estimation and heterozygosity estimation using short read data. The filtered high-quality short-read sequencing data were used to calculate 21-mer distribution through Jellyfish v2.3.0 program²³. From the jellyfish output, genome size, repeat content, and heterozygosity were estimated using GenomeScope2 Program²⁴ at 426 Mb, 45.8%, and 5.16%, respectively (Fig. 1c).

High-quality genome construction. We utilized NextDenovo v2.4.0 (<https://github.com/Nextomics/NextDenovo>) to assemble the *P. pectinifera* genome, using a total of 25.95 Gb data in Nanopore sequences with N50 21 kb. After the assembly, assembled contigs were polished with Illumina short reads using four times of NextPolish v1.1.0²⁵. Then we employed the Hi-C technology to obtain chromosome-level genome assembly. Firstly, the paired-end Illumina reads were mapped onto the polished assembly using HiC-Pro v3.0.0²⁶ with default parameters to check the quality of the raw Hi-C reads. Then Juicer v1.6²⁷ and 3D-DNA v180419²⁸ were applied to cluster the genomic contig sequences into potential chromosomal groups. Afterward, contig orientations were validated and ambiguous fragments were removed with manual curation using Juicebox v1.13²⁹, whereby consecutive contigs were linked to generate a high-quality genome assembly (Fig. 1d). The final chromosome-level genome assembly was created with the largest 22 pseudo-chromosomes, which correspond to the known karyotype of the species³⁰ (hereafter referred to as "Chromosomes"), and 1,458 small scaffolds with a length of 615 Mb (Table 1).

Repeat annotation. The identification of repetitive regions in the genomes was accomplished through a two-step process involving ab-initio prediction and homology-based repeat search. Firstly, *ab-initio* prediction approach was accomplished by using RepeatModeler v2.0.1³¹ with RECON v1.08³², RepeatScout v1.0.6³³, rmbblast

Statistics	Value
Number of Scaffolds	22 (+1458)
Genome Length	579,098,568 (+36,262,640)
Min bp in pseudochromosomes	15,291,986
Max bp in pseudochromosomes	54,886,374
Average bp of pseudochromosomes	26,322,662.18
N50	24,204,423
QV	31.77 (0.00066 error rate)

Table 1. Genome Assembly statistics of *P. pectinifera*.

	Detection Tool		Repeatmasker	Repeatmasker + RFSB
	Total		281,711,337	
Class I	LTR	Gypsy	6,061,638	36,080,823
		Copia	169,828	6,893,215
		others	3,383,226	3,680,142
	Non-LTR	LINE	10,121,821	10,532,230
		SINE	147,399	419,396
Class II			Tc1-Mariner	420,985
			Zator	29,369
			hAT	437,019
			CMC	347,430
			others	2,767,384
Helitron			975,329	1,513,271
Simple repeat			4,188,228	4,188,228
Low complexity			585,869	585,869
Others (Include unknown)			252,075,812	328,827

Table 2. Repeat annotation results comparison.

v2.10.0 (<http://www.repeatmasker.org/RMBlast.html>) and TRF v409³⁴ to predict interspersed repeats. With ‘-LTRStruct’ option in Repeatmodeler, LTRs were also identified using LtrHarvest³⁵ (on genomertools-1.5.9³⁶), Ltr_retriever v2.9.0³⁷, MAFFT v7.407³⁸, CD-HIT v4.8.1³⁹, and Ninja v1.2.2⁴⁰. This library, in addition to RepBase⁴¹, was then utilized to perform a homology-based repeat search throughout the genome using RepeatMasker⁴².

Given the large amounts of unclassified elements observed in our data (89.36% of all repeats), we conducted an additional transposon classification analysis using RFSB tool⁴³. This was done to further elucidate and identify the unidentified repeat sequences (Table 2).

Gene prediction and functional annotation. A total of four rounds of genome annotation were performed with MAKER3 pipeline⁴⁴. Initial evidence-based annotation was performed using exonerate v2.4.0⁴⁵, with proteome evidences (*A. rubens*: GCF_902459465.1, *A. planci*: GCF_001949145.1, *P. miniata*: GCF_015706575.1, *S. purpuratus*: GCF_000002235.5, and *P. pectinifera*: 26 proteins from UniProt database⁴⁶) and reference-based transcriptome assembly data. The first round was used to train the gene prediction software SNAP⁴⁷ with the alignment above. In the next round, trained SNAP gene models were used for *ab initio* MAKER3 annotation. Then, the SNAP gene model was trained with second annotation data, and the augustus/genemark gene model constructed from braker2⁴⁸ with transcript alignment data and protein data was used as a third-round annotation. The final round was performed with the EVM module weights of 8, 2, and 1, respectively, to the transcript, protein, and *ab initio* evidence. As a result, 40,468 genes were predicted and the gene structure annotation, predicted CDS and protein sequences and annotation of repeats were uploaded in figshare (See Data Records).

Data Records

The final genome assembly has been deposited at GenBank with accession number JASAOE000000000⁴⁹, and raw sequencing data from the Nanopore, Illumina, and Hi-C libraries have been deposited at NCBI with accession numbers SRR24423632, SRR24423633, SRR24423634 under BioProject SRP435816⁵⁰.

The transcriptome data of 3 tissues were deposited with accession numbers SRR24423629, SRR24423630, SRR24423631 under the same Bioproject.

All results from genome annotation are available in figshare⁵¹.

Genome	<i>P. pectinifera</i>
Results	C:98.5% [S:96.4%, D:2.1%], F:0.4%, M:1.1%, n:954
Complete BUSCOs (C)	940
Complete and single-copy BUSCOs (S)	920
Complete and duplicated BUSCOs (D)	20
Fragmented BUSCOs (F)	4
Missing BUSCOs (M)	10
Total BUSCO groups searched (n)	954

Table 3. Genome assessment of complete genome.

Technical Validation

Evaluation of the genome assembly. To evaluate the quality of the genome assembly, we first assessed the completeness of the assembled genome using BUSCO v3.0.2b⁵². The final genome assembly yielded a Complete BUSCO score of 98.5% (Table 3). Second, we measured the Base Accuracy QV value using Merqury⁵³, resulting in a QV value of 31.77 (0.00066 error rate) (Table 1). In addition, we computed the N50 and L50 metrics for the genome, which turned out to be 24,204,423 and 11 respectively (Table 1). These quality assessments suggest a high-quality genome assembly.

Code availability

All procedures and workflows employed in data processing adhered to the guidelines and protocols outlined in the respective bioinformatics software manuals. This study did not involve the development of any specialized code.

Received: 24 May 2023; Accepted: 29 August 2023;

Published online: 20 September 2023

References

- Kim, K., Kim, Y.-G., Cho, Y.-K., Takematsu, M. & Volkov, Y. Basin-to-basin and year-to-year variation of temperature and salinity characteristics in the East Sea (Sea of Japan). *Journal of oceanography* **55**, 103–109 (1999).
- Sin, E., Min, W. G., Kim, Y.-B. & Kim, T. W. Respiration of the sea urchin *Mesocentrotus nudus* in response to large temperature fluctuations. *Marine environmental research* **144**, 178–185 (2019).
- Free, C. M. *et al.* Impacts of historical warming on marine fisheries production. *Science* **363**, 979–983 (2019).
- Fisheries, M. O. O. A. (ed Ministry of Oceans and Fisheries) (Ministry of Oceans and Fisheries, 2014).
- Martinez, A. S., Byrne, M. & Coleman, R. A. What and when to eat? Investigating the feeding habits of an intertidal herbivorous starfish. *Marine biology* **163**, 1–13 (2016).
- Makarevitch, I. *et al.* Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS genetics* **11**, e1004915 (2015).
- Horváth, V., Merenciano, M. & González, J. Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends in Genetics* **33**, 832–841 (2017).
- Ikegami, S., Tamura, S. & Kanatani, H. Starfish gonad: action and chemical identification of spawning inhibitor. *Science* **158**, 1052–1053 (1967).
- Mita, M. *et al.* A relaxin-like peptide purified from radial nerves induces oocyte maturation and ovulation in the starfish, *Asterina pectinifera*. *Proceedings of the National Academy of Sciences* **106**, 9507–9512 (2009).
- Kim, C. H. *et al.* Identification of a novel starfish neuropeptide that acts as a muscle relaxant. *Journal of neurochemistry* **137**, 33–45 (2016).
- Davydov, P., Shubravvi, O. & Vassetzky, S. in *Animal species for developmental studies* 287–311 (Springer, 1990).
- Mita, M. Relaxin-like gonad-stimulating substance in an echinoderm, the starfish: a novel relaxin system in reproduction of invertebrates. *General and comparative endocrinology* **181**, 241–245 (2013).
- Kim, C.-H. *et al.* Identification of a novel antimicrobial peptide from the sea star *Patiria pectinifera*. *Developmental & Comparative Immunology* **86**, 203–213 (2018).
- Katz, K. *et al.* The Sequence Read Archive: a decade more of explosive growth. *Nucleic acids research* **50**, D387–D390 (2022).
- Porebski, S., Bailey, L. G. & Baum, B. R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant molecular biology reporter* **15**, 8–15 (1997).
- Zhang, M. *et al.* Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *nature protocols* **7**, 467–478 (2012).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- De Coster, W., D’hert, S., Schultz, D. T., Cruets, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
- Broseus, L. *et al.* TALC: transcript-level aware long-read correction. *Bioinformatics* **36**, 5000–5006 (2020).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome biology* **20**, 1–13 (2019).
- Marçais, G. & Kingsford, C. Jellyfish: A fast k-mer counter. *Tutorialis e Manuais* **1**, 1–8 (2012).
- Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature communications* **11**, 1–10 (2020).
- Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* (2020).
- Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology* **16**, 1–11 (2015).
- Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems* **3**, 95–98 (2016).
- Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

29. Durand, N. C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3**, 99–101 (2016).
30. Saotome, K. & Komatsu, M. Chromosomes of Japanese starfishes. *Zoological science* **19**, 1095–1103 (2002).
31. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).
32. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome research* **12**, 1269–1276 (2002).
33. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
34. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* **27**, 573–580 (1999).
35. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* **9**, 1–14 (2008).
36. Gremme, G., Steinbiss, S. & Kurtz, S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM transactions on computational biology and bioinformatics* **10**, 645–656 (2013).
37. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant physiology* **176**, 1410–1422 (2018).
38. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772–780 (2013).
39. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
40. Wheeler, T. J. in *International Workshop on Algorithms in Bioinformatics*. 375–389 (Springer).
41. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna* **6**, 1–6 (2015).
42. Chen, N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* **5**, 4.10. 11–14.10. 14 (2004).
43. Riehl, K., Riccio, C., Miska, E. A. & Hemberg, M. TransposonUltimate: software for transposon classification, annotation and detection. *Nucleic Acids Research* **50**, e64–e64 (2022).
44. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Current protocols in bioinformatics* **48**, 4.11. 11–14.11. 39 (2014).
45. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics* **6**, 1–11 (2005).
46. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic acids research* **28**, 45–48 (2000).
47. Korf, I. Gene finding in novel genomes. *BMC bioinformatics* **5**, 1–9 (2004).
48. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* **3**, lqaa108 (2021).
49. NCBI GenBank <https://identifiers.org/ncbi/insdc:JASAOE000000000> (2023).
50. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRP435816> (2023).
51. Jung, J. *et al.* gene structure annotation, predicted CDS and protein sequences and annotation of repeats of patiria pectinifera assembly. [figshare https://doi.org/10.6084/m9.figshare.23118647.v2](https://doi.org/10.6084/m9.figshare.23118647.v2) (2023).
52. Seppy, M., Manni, M. & Zdobnov, E. M. in *Gene prediction 227–245* (Springer, 2019).
53. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome biology* **21**, 1–27 (2020).

Acknowledgements

This work was supported by a grant from the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIBR202131103, NIBR202227203).

Author contributions

Conception and Design: J.J., J.Y., H.K., S.K., Sample Collection: W.C., S.K., T.P. and C.B., Laboratory experiments: B.Koh. Analysis and interpretation of data: J.J., H.S., B.Kim, S.J., Drafting the manuscript: J.J. and J.Y.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023