BMC Research Notes

DATA NOTE

Open Access

De novo assembly and annotation of the mangrove cricket genome

Aya Satoh[1,2]* , Miwako Takasu[1], Kentaro Yano[2] and Yohey Terai[1]

Abstract

Objectives: The mangrove cricket, *Apteronemobius asahinai*, shows endogenous activity rhythms that synchronize with the tidal cycle (i.e., a free-running rhythm with a period of ~ 12.4 h [the circatidal rhythm]). Little is known about the molecular mechanisms underlying the circatidal rhythm. We present the draft genome of the mangrove cricket to facilitate future molecular studies of the molecular mechanisms behind this rhythm.

Data description: The draft genome contains 151,060 scaffolds with a total length of 1.68 Gb (N50: 27 kb) and 92% BUSCO completeness. We obtained 28,831 predicted genes, of which 19,896 (69%) were successfully annotated using at least one of two databases (UniProtKB/SwissProt database and Pfam database).

Keywords: *Apteronemobius asahinai*, Circatidal rhythm, Illumina, Nanopore

Objective

Some animals in the intertidal zone, which is influenced by a tidal flooding and ebbing cycle of approximately 12.4 h, show a tidal rhythm in their activity [1–3]. This endogenous rhythm, which persists even under constant conditions, is known as a circatidal rhythm, and it occurs over a range of ~ 11.5 h (predatory mite) [4] to ~ 13.8 h (high-shore limpet) [5]. Although the molecular mechanisms underlying the circadian rhythm (i.e., an endogenous rhythm with a period of ~ 24 h) are well known [6], mechanistic studies of circatidal rhythms are limited [7, 8].

The mangrove cricket (*Apteronemobius asahinai*), an endemic species of mangrove forest floors, is also influenced by tides. This cricket shows a circatidal rhythm in its locomotor activity, with a period of ~ 12.6 h [9, 10]. This endogenous rhythm is not entrained by the light–dark cycle but by periodic inundations [11, 12]. The mangrove cricket is one of only a few model organisms

studied for the purpose of understanding the molecular mechanisms of the circatidal rhythm. Previous work demonstrated that the circatidal rhythm was not disrupted by suppressing the expression of two circadian clock genes, *period* and *Clock* [13, 14]. These findings indicate that the molecular components of the circatidal clock differ from those of the circadian clock in the mangrove cricket. Recently, transcriptome analyses of this species were conducted to reveal circatidal clock-controlled genes [15] or to identify biological processes related to the circatidal rhythm [16]. Here, we provide the draft genome of the mangrove cricket. This information is expected to contribute to future molecular studies by enabling the use of molecular techniques such as GWAS.

Data description

Mangrove crickets were collected from a mangrove forest in Ginoza, Okinawa Prefecture, Japan. To generate highly homozygous individuals, we repeated sibling mating over 7 generations and used two adult males of the eighth generation for DNA extraction (for details, see Data file 1). Genomic DNA from the whole body of a male was extracted using the DNeasy® Blood & Tissue Kit (Qiagen). The NEBNext Ultra II DNA Library

*Correspondence: ayasatohjp@gmail.com
[1] Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Shonan Village, Hayama, Kanagawa 240-0193, Japan
Full list of author information is available at the end of the article

Satoh *et al. BMC Res Notes* (2021) 14:387

Page 2 of 3

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Materials and Methods	PDF	figshare: https://doi.org/10.6084/m9.figshare.16632781 [33]
Data file 2	Illumina raw sequences	FASTQ	DDBJ Sequence Read Archive: https://identifiers.org/insdc.sra:DRX290103 [34]
Data file 3	ONT raw sequences	FASTQ	DDBJ Sequence Read Archive: https://identifiers.org/insdc.sra:DRX290104 [35]
Data file 4	Whole genome sequence data	FASTA	DDBJ: https://identifiers.org/ncbi/insdc:BPSV01000000 [36]
Data file 5	Structural and functional gene annotation	GFF	figshare: https://doi.org/10.6084/m9.figshare.14746056 [37]
Data file 6	predicted protein sequences	FASTA	figshare: https://doi.org/10.6084/m9.figshare.14746056 [38]
Data file 7	predicted transcript sequences	FASTA	figshare: https://doi.org/10.6084/m9.figshare.14746056 [39]

Prep Kit for Illumina (New England BioLabs) was used to construct a library from 500 ng sample DNA. Paired-end (2×150 bp) sequencing was performed on the Illumina HiSeq X platform. For long-read library preparation, genomic DNA from the whole body of another male was extracted using the DNeasy® Blood & Tissue Kit and Genomic-tip 20G Kit (both from Qiagen). Short DNA fragments were removed using Short Read Eliminator Kit (Circulomics). The library was constructed from 415 ng sample DNA using the Rapid Sequencing Kit (SQK-RAD004; Oxford Nanopore Technologies [ONT]). Sequencing was performed twice on the MinION Mk1b with a flow cell R9.4 (FLO-MIN106D; ONT). The Illumina and ONT platforms yielded 217.5 and 14.6 Gb of nucleotide sequence, respectively. The Illumina reads (Data file 2) were assembled and scaffolded using the CLC genomic workbench v20.0.4 [17]. The ONT reads (Data file 3) were trimmed for adapter and low-quality reads using Porechop v0.2.4 [18] and Nanofilt v2.8.0 [19], respectively, and then error-corrected using the Illumina reads by LoRDEC v0.9 [20]. Finally, the error-corrected ONT reads were subjected to gap closing in the scaffolds using TGS-Gapcloser v1.1.1 [21]. The final draft genome (Data file 4) consists of 151,060 scaffolds with a total length of 1,676,217,857 bp, average length of 11,096 bp, and N50 of 27,317 bp. BUSCO analysis using the online interface gVolante [22] identified 983 genes (92.21%) among the 1,066 arthropodal universal orthologs completely, and only 17 genes (1.59%) were missing, indicating high completeness of our draft genome.

RepeatModeler v2.0.1 [23] estimated 2532 repeat sequences, which were utilized by RepeatMasker v4.0.9 [24] to mask the repetitive elements in the genome. The repeat sequences in the assembly comprised 572,734,587 bp (34.17% of the total length). The MAKER v2.31.11 [25] pipeline predicted 28,831 protein-coding genes in the hard-masked genome (Data files 5–7). The average coding sequence length was 997.08 bp, with an average intron length of 1000.45 bp and average number of exons per gene of 4.34. We annotated 16,528 genes (57.3%) via a BLASTP v2.10.1 + [26] search (E-value threshold of 1×10^{-10}) against known proteins in the UniProtKB/SwissProt Database [27]. InterProScan v5.50–84.0 [28] identified 4537 domain families among 17,932 (62.3%) genes via a search of the Pfam database. As a result, 69% of the predicted genes were successfully annotated by at least one of the two methods.

Limitations

The genome size, assessed by the k-mer frequency distribution of the Illumina reads using KmerGenie v1.7051 [29], was estimated to be 1,610,998,267 bp. Based on this estimation, the sequencing depths obtained from the Illumina and ONT platforms were calculated to be $134\times$ and $9\times$, respectively. Since the coverage of ONT reads was low, the usage of them were limited only to the gap closing. The genome size of the mangrove cricket is comparable with the three previously sequenced Gryllidae genomes: *Teleogryllus occipitalis* (1.93 Gb) [30], *Teleogryllus oceanicus* (2.05 Gb) [31], and *Laupala kohalensis* (1.6 Gb) [32].

Satoh *et al. BMC Res Notes*   (2021) 14:387

Page 3 of 3

Availability of data and materials

The data described in this Data note can be freely and openly accessed on DDBJ under BioProject ID: PRJDB11838 and the figshare database. Sequence reads have been deposited at DDBJ Sequence Read Archive under accession number DRX290103 (https://identifiers.org/insdc.sra:DRX290103) [34] and DRX290104 (https://identifiers.org/insdc.sra:DRX290104) [35]. The whole genome sequence data has been deposited at DDBJ under accession number BPSV01000000 (https://identifiers.org/ncbi/insdc:BPSV01000000) [36]. The other data files generated in the current study are available at the figshare database: Data file 1 (https://doi.org/10.6084/m9.figshare.16632781) [33], Data file 5–7 (https://doi.org/10.6084/m9.figshare.14746056) [37–39]. See Table 1 and references [33–39] for details.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Shonan Village, Hayama, Kanagawa 240-0193, Japan. [2]School of Agriculture, Meiji University, Kawasaki, Kanagawa 214-8571, Japan.

References

1. Akiyama T. Circatidal swimming activity rhythm in a subtidal cumacean *Dimorphostylis asiatica* (Crustacea). Mar Biol. 1995;123:251–5.
2. Barnwell FH. Daily and tidal patterns of activity in individual fiddler crab (Genus *Uca*) from the Woods Hole region. Biol Bull. 1966;13:1–17.
3. Satoh A, Momoshita H, Hori M. Circatidal rhythmic behaviour in the coastal tiger beetle *Callytron inspecularis* in Japan. Biol Rhythm Res. 2006;37(2):147–55.
4. Treherne JE, Foster WA, Evns PD, Ruscoe CNE. Free-running activity rhythm in the natural environment. Nature. 1977;269:796–7.
5. Gray DR, Hodgson AN. Endogenous rhythms of locomotor activity in the high-shore limpet, *Helcion pectunculus* (Patellogastropoda). Anim Behav. 1999;57:387–91.
6. Dunlap JC, Loros JJ, DeCoursey PJ. Chronobiology: Biological Timekeeping. Massachusetts: Sinauer; 2004.
7. Bulla M, Oudman T, Bijleveld AI, Piersma T, Kyriacou CP. Marine biorhythms: bridging chronobiology and ecology. Philos Trans R Soc B. 2017;372:20160253.
8. Zhang L, Hastings MH, Green EW, Tauber E, Sladek M, Webster SG, et al. Dissociation of circadian and circatidal timekeeping in the marine crustacean *Eurydice pulchra*. Curr Biol. 2013;23:1863–73.
9. Satoh A. Constant light disrupts the circadian but not the circatidal rhythm in mangrove crickets. Biol Rhythm Res. 2017;48:459–63.
10. Satoh A, Yoshioka E, Numata H. Circatidal activity rhythm in the mangrove cricket *Apteronemobius asahinai*. Biol Lett. 2008;4:233–6.
11. Satoh A, Yoshioka E, Numata H. Entrainment of the cricatidal activity rhythm of the mangrove cricket, *Apteronemobius asahinai*, to periodic inundations. Anim Behav. 2009;78:189–94.
12. Sakura K, Numata H. Contact with water functions as a Zeitgeber for the circatidal rhythm in the mangrove cricket *Apteronemobius asahinai*. Biol Rhythm Res. 2017;48:887–95.
13. Takekata H, Matsuura Y, Goto SG, Satoh A, Numata H. RNAi of the circadian clock gene *period* disrupts the circadian rhythm but not the circatidal rhythm in the mangrove cricket. Biol Lett. 2012;8:488–91.
14. Takekata H, Numata H, Shiga S, Goto SG. Silencing the circadian clock gene *Clock* using RNAi reveals dissociation of the circatidal clock from the circadian clock in the mangrove cricket. J Insect Physiol. 2014;68:16–22.
15. Satoh A, Terai Y. Circatidal gene expression in the mangrove cricket *Apteronemobius asahinai*. Sci Rep. 2019;9:3719.
16. Takekata H, Tachibana S, Motooka D, Nakamura S, Goto SG. Possible biological processes controlled by the circatidal clock in the mangrove cricket inferred from transcriptome analysis. Biol Rhythm Res. 2020. https://doi.org/10.1080/09291016.2020.1838747.
17. CLC Genomic Workbench. https://www.qiagenbioinformatics.com/.
18. Wick R. Porechop. 2018. https://github.com/rrwick/Porechop/.
19. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long read sequencing data. Bioinformatics. 2018;34:2666–9.
20. Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. Bioinformatics. 2014;30:3506–14.
21. Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, et al. TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience. 2020;9:giaa094.
22. Nishimura O, Hara Y, Kuraku S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics. 2017;33:3635–7.
23. Smit AFA, Hubley R. RepeatModeler Open-2.0. http://www.repeatmasker.org.
24. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org.
25. Holt C, Yandell M. MAKER2: an annotation pipeline and genome database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.
26. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and application. BMC Bioinformatics. 2009;10:421.
27. Uniprot. https://www.uniprot.org/. Accessed 19 Nov 2020.
28. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
29. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2013;30:31–7.
30. Kataoka K, Minri R, Ide K, Ogura A, Takeyama H, Takeda M, et al. The draft genome dataset of the Asian cricket *Teleogryllus occipitalis* for molecular research toward entomophagy. Front Genet. 2020;11:470.
31. Pascoal S, Risse JE, Zhang X, Blaxter M, Cezard T, Challis RJ, et al. Field cricket genome reveals the footprint of recent, abrupt adaptation in the wild. Evol Lett. 2019;4:19–33.
32. Blankers T, Oh KP, Bombarely A, Shaw KL. The genomic architecture of a rapid island radiation: recombination rate variation, chromosome structure, and genome assembly of the Hawaiian cricket *Laupala*. Genetics. 2018;209:1329–44.
33. Satoh A, Takasu M, Yano K, Terai Y. Materials and Methods.pdf. figshare. 2021. https://doi.org/10.6084/m9.figshare.16632781.
34. Satoh A, Terai Y. HiSeq X Ten paired end sequencing of SAMD00330124. DDBJ Sequence Read Archive. 2021. https://identifiers.org/insdc.sra:DRX290103.
35. Satoh A, Terai Y. MinION sequencing of SAMD00330124. DDBJ Sequence Read Archive. 2021. https://identifiers.org/insdc.sra:DRX290104.
36. Satoh A, Terai Y. *Apteronemobius asahinai*, whole genome shotgun sequencing project. DDBJ. 2021. https://identifiers.org/ncbi/insdc:BPSV01000000.
37. Satoh A, Takasu M, Yano K, Terai Y. Apteronemobius_asahinai.gff. figshare. 2021. https://doi.org/10.6084/m9.figshare.14746056.
38. Satoh A, Takasu M, Yano K, Terai Y. Apteronemobius_asahinai_proteins.fasta. figshare. 2021. https://doi.org/10.6084/m9.figshare.14746056.
39. Satoh A, Takasu M, Yano K, Terai Y. Apteronemobius_asahinai_transcripts.fasta. figshare. 2021. https://doi.org/10.6084/m9.figshare.14746056.

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