AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

EUKARYOTES

Draft Genome Sequence of *Plasmodium gonderi*, a Malaria Parasite of African Old World Monkeys

Hajime Honma,[a] Satoru Kawai,[b] Daisuke Motooka,[c] Shota Nakamura,[c] Takahiro Tougan,[d] Toshihiro Horii,[d] Nobuko Arisue[d]

Department of International Affairs and Tropical Medicine, Tokyo Women's Medical University, Tokyo, Japan[a]; Laboratory of Tropical Medicine and Parasitology, Dokkyo University School of Medicine, Tochigi, Japan[b]; Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan[c]; Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan[d]

ABSTRACT *Plasmodium gonderi* is a primate parasite whose natural host is the African Old World monkeys. Here, we report the draft genome sequence for *P. gonderi*. The data are useful not only for understanding the evolution of malaria but also for allowing the comparative genomics of malaria parasites.

Received 12 May 2017 **Accepted** 18 May 2017 **Published** 13 July 2017

Citation Honma H, Kawai S, Motooka D, Nakamura S, Tougan T, Horii T, Arisue N. 2017. Draft genome sequence of *Plasmodium gonderi*, a malaria parasite of African Old World monkeys. Genome Announc 5:e00612-17. https://doi.org/10.1128/genomeA.00612-17.

Address correspondence to Nobuko Arisue, arisue@biken.osaka-u.ac.jp.

The natural hosts of *Plasmodium gonderi* are African guenon such as mangabeys (e.g., *Cercocebus atys*, *C. galeritus*, *C. aterrimus*, and *Cercopithecus* spp.) and drills (e.g., *Mandrillus leucophaeus*) (1, 2). *P. gonderi* is usually used as the outgroup species for phylogenetic analysis of the Asian macaque malaria parasite clade (2, 3), which includes *P. vivax*, the most widely distributed human malaria parasite. However, to date, a thorough understanding of the evolutionary process of this malaria parasite has been hampered by the limited genomic resources available on *P. gonderi*. Here, we report the first draft genome sequence for *P. gonderi*, based on a combination of short-read (MiSeq) and long-read (PacBio) sequencing technology.

An infected blood sample of *P. gonderi* (ATCC 30045) was obtained from an experimentally infected Japanese macaque. The investigators adhered to the Guidelines for the Use of Experimental Animals authorized by the Japanese Association for Laboratory Animal Science. The protocol was approved by the Committee on the Ethics of Animal Experiments of the Dokkyo University of School of Medicine (permit no. 0536). Genomic DNA of *P. gonderi* was extracted from parasitized red blood cells using the saponin method (4). Whole-genome sequencing was performed using the MiSeq (Illumina) and PacBio RS II (Pacific Biosciences) platforms. For the MiSeq sequencing, 500 ng of genomic DNA was sheared to about 600 bp; the library was prepared using KAPA library preparation kits (KAPA Biosystems), and then paired-end sequencing (2 × 251 bp) was performed. For the PacBio RS II sequencing, 2 μg of genomic DNA was sheared to about 15 kb; the library was prepared using a DNA template prep kit version 1.0 (Pacific Biosciences), and sequencing was performed. *De novo* assembly of the MiSeq reads was performed with Celera Assembler version 8.1 (5). Scaffolding of the MiSeq contigs with PacBio subreads was performed using SSPACE-LONGREAD version 1.1 (6). GapFiller version 1.10 (7) was used to close gaps. Scaffolds corresponding to 14 chromosomes, 1 apicoplast genome, and 1 mitochondrial genome were constructed using the *P. vivax* genome (8) as a reference. MiSeq and PacBio reads were remapped to the scaffolds, and unmapped reads were collected and assembled into 727 contigs using CLC Genomics Workbench version 7.5.1 (CLC bio/QIAGEN).

Gene prediction was performed using AUGUSTUS (9) implemented on Geneious version 9.1.7 (10), followed by manual correction by comparison with orthologous gene sequences of closely related *Plasmodium* spp.

The final assembly of the *P. gonderi* genome consisted of 743 scaffolds and contigs comprising 33.0 Mb. The mean coverage was 253×, and the maximum length was 3,573,180 bp with an N_{50} of 1.64 Mb and a G+C content of 26.9%. The nuclear genome covers a predicted 5,885 protein-encoding genes, 13 rRNAs, and 44 tRNAs; for the apicoplast genome, 29 protein-encoding genes, 4 rRNAs, and 34 tRNAs were predicted; and the mitochondrial genome is predicted to contain 3 protein-encoding genes.

Accession number(s). The 14 scaffolds corresponding to the 14 chromosomes, 727 contigs of unknown chromosome location, and the mitochondrial and apicoplast genomes of *P. gonderi* have been deposited at DDBJ/GenBank (BioProject PRJDB5590) under the accession numbers BDQF01000001 to BDQF01000743.

ACKNOWLEDGMENTS

This work was funded by grant JSPS KAKENHI 25460516 and by a Joint Research Project grant of the Research Institute for Microbial Diseases, Osaka University. We express our great thanks to the late Kazuyuki Tanabe for preparation of the parasite material and for his valuable advice for this project. This research is partially supported by the National Bio-Resource Project at the National Institute of Physiological Science through the Japan Agency for Medical Research and Development (AMED). Bioinformatics analyses were in part conducted using the computer system at the Genome Information Research Center of the Research Institute for Microbial Diseases at Osaka University. We thank Nirianne M. Q. Palacpac for help with language editing.

REFERENCES

1. Coatney RG, Collins WE, Warren M, Contacos PG. 1971. The primate malarias. U.S. Government Printing Office, Washington, DC.
2. Escalante AA, Cornejo OE, Freeland DE, Poe AC, Durrego E, Collins WE, Lal AA. 2005. A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. Proc Natl Acad Sci U S A 102:1980–1985. https://doi.org/10.1073/pnas.0409652102.
3. Mitsui H, Arisue N, Sakihama N, Inagaki Y, Horii T, Hasegawa M, Tanabe K, Hashimoto T. 2010. Phylogeny of Asian primate malaria parasites inferred from apicoplast genome-encoded genes with special emphasis on the positions of *Plasmodium vivax* and *P. fragile*. Gene 450:32–38. https://doi.org/10.1016/j.gene.2009.10.001.
4. Honma H, Hirai M, Nakamura S, Hakimi H, Kawazu S, Palacpac NM, Hisaeda H, Matsuoka H, Kawai S, Endo H, Yasunaga T, Ohashi J, Mita T, Horii T, Furusawa M, Tanabe K. 2014. Generation of rodent malaria parasites with a high mutation rate by destructing proofreading activity of DNA polymerase δ. DNA Res 21:439–446. https://doi.org/10.1093/dnares/dsu009.
5. Denisov G, Walenz B, Halpern AL, Miller J, Axelrod N, Levy S, Sutton G. 2008. Consensus generation and variant detection by Celera assembler. Bioinformatics 24:1035–1040. https://doi.org/10.1093/bioinformatics/btn074.
6. Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15:211. https://doi.org/10.1186/1471-2105-15-211.
7. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol 13:R56. https://doi.org/10.1186/gb-2012-13-6-r56.
8. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RM, Crabb BS, Del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TW, Korsinczky M, Meyer EV, Nene V, Paulsen I, White O, Ralph SA, Ren Q, Sargeant TJ, Salzberg SL, Stoeckert CJ, Sullivan SA, Yamamoto MM, Hoffman SL, Wortman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggett CM. 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature 455:757–763. https://doi.org/10.1038/nature07327.
9. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.
10. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.