

Genome sequence of *Asaia bogorensis* strain SC1 isolated from an *Aedes aegypti* mosquito crop

Souvik Chakraborty,¹ Joshua B. Benoit,¹ Annette R. Rowe,¹ Joshua D. Sackett¹

AUTHOR AFFILIATION See affiliation list on p. 2.

ABSTRACT Understanding microbe-host interactions is key to combating disease transmission by mosquitoes. Here, we report the genome sequence of *Asaia bogorensis* strain SC1 isolated from a human-blood-fed *Aedes aegypti* mosquito crop. Metabolic pathway characteristics of aerobic respiration were present in the genome, along with multiple putative antibiotic resistance mechanisms.

KEYWORDS mosquito microbiome, crop, metabolic pathways, antibiotic resistance, genomic characterization

The dipteran crop, a food storage organ found in mosquitoes, harbors a diverse microbiome that can impact various aspects of host physiology and ecology (1–3). However, microbial metabolism in the crop remains poorly characterized, especially regarding its potential role in influencing mosquito-borne disease transmission (4). Here, we report the genome sequence of *Asaia bogorensis* strain SC1, isolated from a human-blood-fed *Aedes aegypti* mosquito crop. Importantly, *Asaia* sp. is dominant in the crop of *Ae. aegypti* (1), suggesting a critical role for these bacteria in the organ.

Asaia bogorensis strain SC1 was isolated from human-blood-fed (University of Cincinnati IRB 2021-0971) *Aedes aegypti* mosquito crop. Briefly, the crop was dissected from a mosquito using ethanol-flamed forceps, homogenized in 1× PBS pH = 7.4, and inoculated on Luria-Bertani (LB) agar at 30°C. To ensure a pure culture, individual colonies were sequentially streaked on fresh LB plates three times. DNA was extracted from overnight LB broth cultures (30°C, 200 RPM) using the Qiagen DNeasy PowerSoil Kit (Qiagen, Germantown, MD) and quantified via Qubit (ThermoFisher Scientific, Waltham, MA). For long-read sequencing, the Native Barcoding Kit 24 V14 Kit (Oxford Nanopore Technologies, Oxford, UK) was used to barcode samples and prepare libraries. DNA was not sheared or size selected prior to library prep. Sequencing was performed with an R10.4.1 flow cell (FLO-MIN114) and minION device under high-accuracy mode (280 bp/s). Basecalling was performed using Guppy 6.4.6, and reads with quality scores <7 were removed. Reads were filtered with FilTlong 0.2.1 (5) to remove reads <2,000 bp and the lowest-quality reads comprising 10% of all sequenced bases. Short-read sequencing libraries were prepared with Illumina DNA Prep Kit (San Diego, CA), barcoded with 10 bp UDI indices, and sequenced on an Illumina NovaSeq using 2 × 150 sequencing chemistry (SeqCenter, Pittsburgh, PA). Short-read sequence quality was visualized with FastQC 0.12.0 (6), and mean sequence quality was >Q30 at all positions, so no further quality filtering was conducted. Nanopore long-read sequences were assembled using Flye 2.9.1 (7). Illumina reads were aligned to the long-read assembly with Burrows-Wheeler Aligner 0.7.17 (8). Pilon 1.23 (9) was used to polish the assembly with parameter --fix all. In total, three rounds of polishing were conducted. Genome statistics and assembly quality were determined with QUAST 4.4 (10) and CheckM 1.0.18 (11). The genome was annotated with NCBI PGAP 6.4 (12). Potential metabolic pathways were identified using KEGG's GhostKOALA tool (13). Taxonomy was assigned with GTDB-Tk 2.2.5 (14). Default

Editor Vincent Michael Bruno, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Joshua D. Sackett, joshua.sackett@uc.edu.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 23 June 2023

Accepted 5 September 2023

Published 16 October 2023

Copyright © 2023 Chakraborty et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Sequencing statistics

Parameter	Value
Raw Illumina reads accession no.	SRR23872088
Raw Nanopore minION reads accession no.	SRR23872085
GenBank accession no. (genome assembly)	GCA_030263835.1
No. of raw paired-end Illumina reads	4,199,454
Illumina estimated genome coverage	168×
No. of raw Nanopore MinION reads (total length, base pairs)	1,215,752,376
Nanopore MinION read length N50 (bp) ^a	5,017
Nanopore MinION estimated genome coverage	303
Assembly size (bp)	3,404,989
G + C content (%)	59.69
Estimated genome completeness (%) ^a	99.75%
Estimated contamination (%) ^a	0
No. of contigs	4
Contig sizes (bp)	3,090,019 229,757 72,572 12,641
No. of protein-coding genes	3,025
No. of tRNA genes	58
No. of rRNA operons	5

^aDetermined with CheckM 1.0.18 (9).

parameters were used unless otherwise specified. Computing resources were supplied and maintained by the Ohio Supercomputer Center (15) and the DOE Systems Biology Knowledgebase (16).

The SC1 genome is 3.4 Mbp with 59.69% GC content (Table 1). GTDB identified the nearest neighbor as *Asaia bogorensis* (96.36% average nucleotide identity). GhostKOALA identified complete pathways for central carbon metabolism, including glycolysis and pyruvate oxidation, TCA cycle, and the pentose phosphate pathway. A complete oxidative phosphorylation pathway was annotated. Interestingly, pyruvate decarboxylase (*pdh*) was absent, but other putative fermentation genes, aldehyde dehydrogenases (*aldB*) and alcohol dehydrogenase (*adhA*), were present. GhostKOALA identified genes potentially conferring resistance to vancomycin, beta-lactams, and cationic antimicrobial peptides. The genome sequence of *Asaia bogorensis* strain SC1 from the mosquito crop provides a valuable resource for further investigating host-microbe interactions of public health interest.

ACKNOWLEDGMENTS

This work was funded by the Wiemann Benedict Grant, Department of Biological Sciences, University of Cincinnati, awarded to S.C. J.D.S. was funded by NSF Postdoctoral Research Fellowship NSF-OCE 2126677.

AUTHOR AFFILIATION

¹Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio, USA

AUTHOR ORCIDs

Joshua B. Benoit  <http://orcid.org/0000-0002-4018-3513>

Annette R. Rowe  <http://orcid.org/0000-0002-8761-6857>

Joshua D. Sackett  <http://orcid.org/0000-0003-1689-8890>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	NSF-OCE PRF 2126677	Joshua D. Sackett
University of Cincinnati (UC)	Wiemann Benedict Grant	Souvik Chakraborty

DATA AVAILABILITY

This study is registered under BioProject [PRJNA945045](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA945045). The raw Illumina and Nanopore sequencing reads have been deposited in the Sequence Read Archive and are available under accession numbers [SRR23872088](https://www.ncbi.nlm.nih.gov/sra/SRR23872088) and [SRR23872085](https://www.ncbi.nlm.nih.gov/sra/SRR23872085), respectively. The complete genome sequence with annotations has been deposited in GenBank under accession number [GCA_030263835.1](https://www.ncbi.nlm.nih.gov/nuccore/GCA_030263835.1).

REFERENCES

- Martinez Villegas LE, Radl J, Dimopoulos G, Short SM. 2023. Bacterial communities of *Aedes aegypti* mosquitoes differ between crop and midgut tissues. *Microbiology*. <https://doi.org/10.1101/2022.08.31.506054>
- Boissière A, Tchioffo MT, Bachar D, Abate L, Marie A, Nsango SE, Shahbazkia HR, Awono-Ambene PH, Levashina EA, Christen R, Morlais I. 2012. Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. *PLoS Pathog* 8:e1002742. <https://doi.org/10.1371/journal.ppat.1002742>
- Wang Y, Gilbreath TM, Kukutla P, Yan G, Xu J, Leulier F. 2011. Dynamic gut Microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. *PLoS ONE* 6:e24767. <https://doi.org/10.1371/journal.pone.0024767>
- Weiss B, Aksoy S. 2011. Microbiome influences on insect host vector competence. *Trends Parasitol* 27:514–522. <https://doi.org/10.1016/j.pt.2011.05.001>
- Wick RR, Menzel P. 2019. Filtlong: quality filtering tool for long reads [Online]. <https://github.com/rrwick/Filtlong>
- Andrews S. 2010. FastQC: a Quality control tool for high throughput sequence data [<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>]
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Tatusova T, DiCuccio M, Badredin A, Chetverin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*. 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
- Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-TK: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
1987. *Ohio Supercomputer Center*. Ohio Supercomputer Center. Available from: <http://osc.edu/ark:/19495/f5s1ph73>
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Bretin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. Kbase: The United States department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>