



Check for updates

GENOME NOTE

The first complete mitochondrial genome of *Anopheles gibbinsi* using a skimming sequencing approach. [version 1; peer review: 3 approved]

Renee Ali ¹, Mary E. Gebhardt¹, James Sichivula Lupiya², Mbanga Muleba², Douglas E. Norris ¹

¹The W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

²Tropical Diseases Research Centre, Ndola, Zambia

V1 First published: 30 May 2024, 13:553
<https://doi.org/10.12688/f1000research.148473.1>
Latest published: 30 May 2024, 13:553
<https://doi.org/10.12688/f1000research.148473.1>

Abstract

Mosquitoes belonging to the genus *Anopheles* are the only vectors of human malaria. *Anopheles gibbinsi* has been linked to malaria transmission in Kenya, with recent collections in Zambia reporting the mosquito species exhibiting zoophilic and exophilic behavioral patterns with occasional contact with humans. Given the paucity of genetic data, and challenges to identification and molecular taxonomy of the mosquitoes belonging to the *Anopheles* genus; we report the first complete mitochondrial genome of *An. gibbinsi* using a genome skimming approach. An Illumina Novaseq 6000 platform was used for sequencing, the length of the mitochondrial genome was 15401 bp, with 78.5% AT content comprised of 37 genes. Phylogenetic analysis by maximum likelihood using concatenation of the 13 protein coding genes demonstrated that *An. marshallii* was the closest relative based on existing sequence data. This study demonstrates that the skimming approach is an inexpensive and efficient approach for mosquito species identification and concurrent taxonomic rectification, which may be a useful alternative for generating reference sequence data for evolutionary studies among the Culicidae.

Keywords

Anopheles gibbinsi, understudied anopheline, genome skimming, Zambia

Open Peer Review

Approval Status

	1	2	3
version 1			
30 May 2024	view	view	view

1. **Nicola Zadra** , University of Trento, Trento, Italy
NBFC, Palermo, Italy
2. **Luke Ambrose** , The University of Queensland, Saint Lucia, Australia
3. **Alexey Makunin**, Wellcome Sanger Institute, Hinxton, UK

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Renee Ali (rali23@jh.edu)

Author roles: **Ali R:** Conceptualization, Data Curation, Formal Analysis, Investigation, Writing – Original Draft Preparation; **Gebhardt ME:** Data Curation, Investigation, Resources, Writing – Review & Editing; **Lupiya JS:** Investigation, Resources, Writing – Review & Editing; **Muleba M:** Investigation, Resources, Writing – Review & Editing; **Norris DE:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was supported in part by funding from the NSF-Accelerator Project D-688: Computing the Biome (2134862), the NIH Southern and Central Africa International Centers of Excellence for Malaria Research (U19AI089680), Johns Hopkins Malaria Research Institute, Bloomberg Philanthropies, and NIH T32 Training Grant (T32AI138953).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Ali R *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ali R, Gebhardt ME, Lupiya JS *et al.* **The first complete mitochondrial genome of *Anopheles gibbinsi* using a skimming sequencing approach. [version 1; peer review: 3 approved]** F1000Research 2024, **13**:553 <https://doi.org/10.12688/f1000research.148473.1>

First published: 30 May 2024, **13**:553 <https://doi.org/10.12688/f1000research.148473.1>

Introduction

A few species in the genus *Anopheles* are well established primary vectors of human malaria in sub-Saharan Africa, including *Anopheles gambiae* and *An. funestus*. However, studies have also implicated understudied anopheline mosquito species in driving transmission in regions where primary vectors are close to elimination.^{1–3} *Anopheles gibbinsi*, until recently recognized as *An. species 6*, has tested positive for *Plasmodium falciparum* sporozoites in studies from Kenya.^{4,5} Furthermore, *An. gibbinsi* was previously reported from central, eastern and northern Africa and has now been shown to have a geographic range extending into southern Africa. Recent first-time captures for this species in Zambia reported the species largely exhibiting zoophilic and exophilic behavioral patterns; however, a blood meal PCR assay also detected a few specimens positive for human host DNA.⁵ Similar in morphology to other well-established malaria vectors and with a dearth of genetic data available, there is a need for the continued monitoring of *An. gibbinsi* as a potential vector in malaria transmission. It is also important that tools for accurate mosquito identification be developed, due to limitations with the commonly targeted cytochrome oxidase I gene (COI) and the internal transcribed spacer 2 (ITS2) in resolving members of morphologically cryptic species complexes.⁶

The expansion of sequencing strategies has employed the use of mitochondrial genomes (mitogenomes) primarily for species identification and solving discrepancies in the taxonomic classification of metazoan organisms. Mitogenomes have also proved useful in evaluating population structure, chromosomal rearrangements, species introgression, and evolutionary histories.^{7–9} The mitochondrial DNA (mtDNA) is a circular double stranded molecule that encodes 37 genes made up of 13 protein coding genes (PCGs), 22 transfer RNA (tRNA) genes, 2 ribosomal RNA (rRNA) genes and an adenine and thymine (A-T) rich terminal, an essentially non-coding area termed the control region (CR) associated with the replication and transcription of the genome. Maternal inheritance, high copy number, lack of recombination, and absence of introns are characteristics which allow mtDNA to be well suited for accurate molecular identification and rectifying taxonomic classification among species.^{6,10} Here for the first time, we describe a genome skimming approach for recovery of, and characterization of the mitochondrial genome of *An. gibbinsi* and its phylogenetic relationship to other established anopheline vectors of human malaria.

Methods

DNA collection

The *An. gibbinsi* specimens (n = 3) sequenced were collected in Nchelenge, Zambia using a CDC light trap that was placed near an animal. The specimens were stored on silica gel until DNA extraction. Single mosquito specimens were pre-treated¹¹ as described by Chen *et al.* 2021, followed by the extraction protocol as per manufacturer's instructions (Qiagen DNeasy Blood and Tissue Kit, Hilden, Germany). A whole mosquito specimen was placed in a 1.5 mL

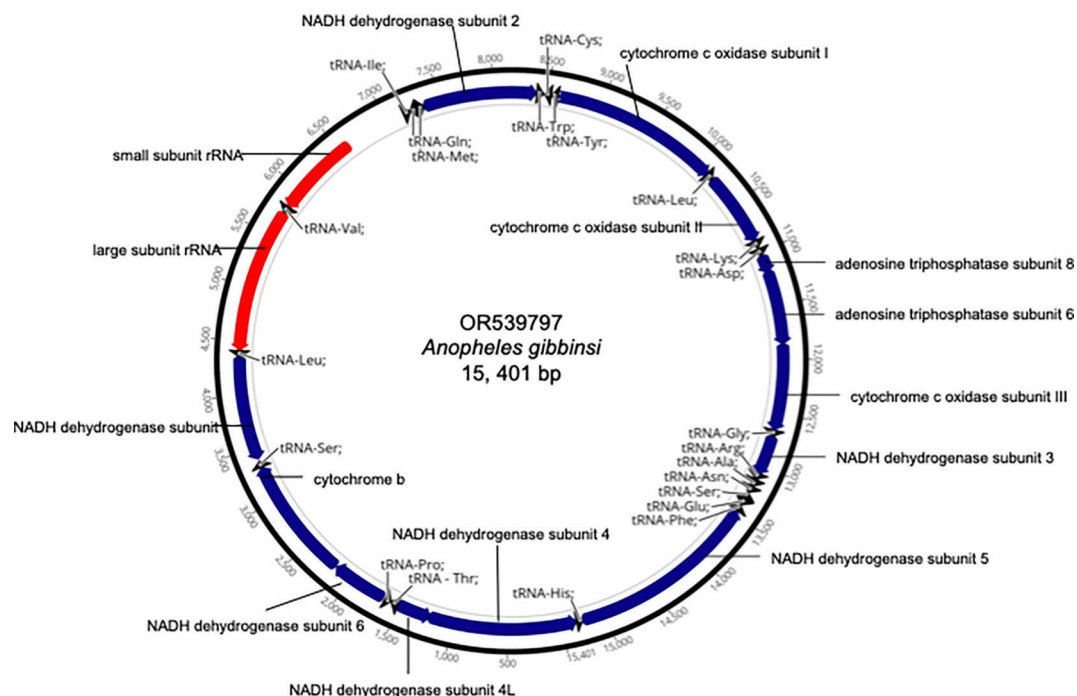


Figure 1. Mitogenome map of *Anopheles gibbinsi* with annotated genes.

Eppendorf tube and homogenized in a cocktail containing 98 μ L of PK buffer (Applied Biosystems, Waltham, Massachusetts, U.S.A.) and 2 μ L of Proteinase K (100 mg/mL), this was followed by an incubation step for 3 hours at 56°C.¹² After incubation, 100 μ L of isopropanol and 100 μ L of Buffer AL (Qiagen DNeasy Blood and Tissue Kit, Hilden, Germany) was added to the lysate and left to incubate at room temperature for 10 minutes. The mixture was pipetted into a DNeasy mini spin column placed in a 2 mL collection tube; extraction protocol was followed⁵ and stored at -20°C prior to sequencing. DNA was shipped to SeqCenter (Pittsburg, U.S.A) for library construction and sequencing. Libraries were sequenced from both ends (150 bp) on Illumina Novaseq 6000 to a depth of 13.3 million reads.

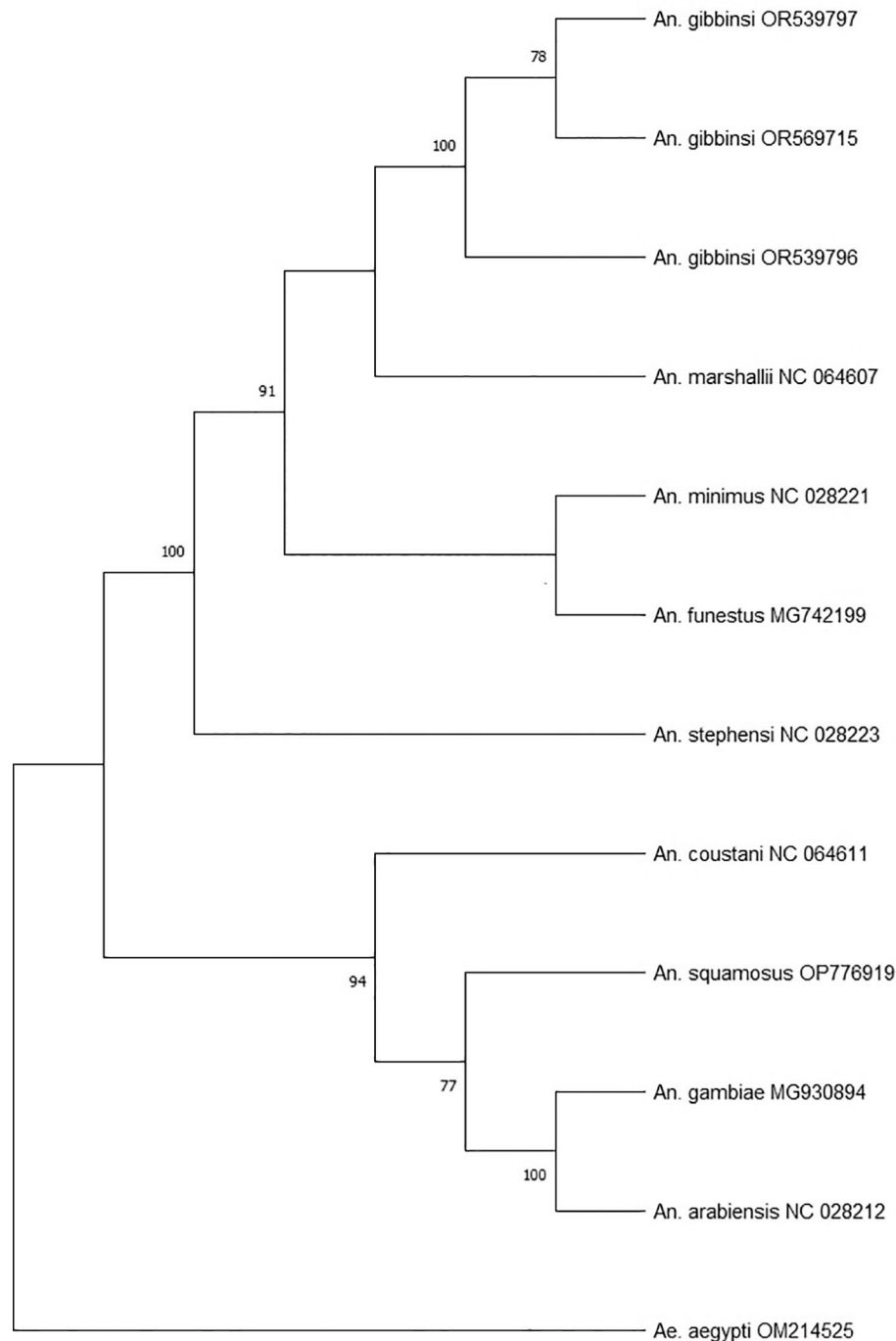


Figure 2. Maximum likelihood tree using the concatenated protein coding genes of *An. gibbinsi* and related *Anopheles*.

The mitochondrial genome contigs were assembled similar to that of the *An. squamosus*¹³ mitogenome using NOVO-Plasty (RRID:SCR_017335) version 4.3.1.¹² Using the invertebrate genetic code under default settings, automatic annotations were conducted using the MITOS website.¹⁴ Adjustments for start and stop codon positions were performed manually in Geneious Prime (RRID:SCR_010519) version 2023.2.1 (Biomatters, Auckland, Australia) to match reference anopheline mitogenomes deposited in NCBI's GenBank database. The mitochondrial genome sequences and their corresponding annotations were submitted to the GenBank database. A representative mitogenome map is provided in [Figure 1](#).

Phylogenetic analysis was performed using the concatenated 13 PCGs of the three *An. gibbinsi* specimens, eight *Anopheles* species and one *Aedes* species as an outgroup. The General Time Reversible (GTR + G + 1) model was identified as the best fit for building the maximum likelihood phylogenetic tree in MEGA (RRID_SCR_023017) version 11¹⁵ using 1000 bootstrap replicates.

Results

The sequencing from the 3 *An. gibbinsi* samples yielded an average of 29,674,206 million reads and of these, approximately 119,364 reads were used to assemble each mitochondrial genome. The contents of the 3 *An. gibbinsi* mitogenomes (GenBank accession numbers OR_539796, OR_539797, OR_569715) included 2 ribosomal RNAs, 22 transfer RNAs and 13 protein coding genes. The representative mitochondrial genome (OR_539797) length was 15,401 bp with an A + T percentage of 78.5% which is comparable to other anopheline mitogenomes deposited in the GenBank database. The cytochrome c oxidase I (COI) fragment spanning 8598–10,133 bp was 94.5% similar to a COI sequence for *An. marshallii* (GenBank YP_010419919).

Phylogenetic analysis ([Figure 2](#)) using the concatenated PCGs revealed *An. marshallii* (NC_064607) as the closest sequenced relative to *An. gibbinsi*, forming a single but weakly supported clade apart from the well-recognized vectors of malaria. Both species belong to the *An. marshallii* complex.

Genome skimming has demonstrated to be a cost-effective approach for generating reference sequence data which can be used for mosquito identification and resolving phylogenies. With the continued monitoring of *An. gibbinsi* as a potential vector for malaria transmission, this study provides a key genomic resource for understanding the phylogenetic relationship of this mosquito species within its complex and with primary vectors of human malaria transmission.

Data availability

GenBank: *Anopheles gibbinsi* mitochondrion, complete genome. Accession numbers [OR539796](#), [OR539797](#), [OR569715](#).¹⁶

Bio Project. Complete mitogenome sequence of *Anopheles gibbinsi* from Nchelenge, Zambia, Accession number [PRJNA1072262](#).¹⁷

SRA. Illumina seq of *Anopheles gibbinsi*. Accession numbers [SRR27842795](#), [SRR27842796](#), [SRR27842794](#).¹⁸

BioSample: *Anopheles gibbinsi* isolates ANGB, AGB1, GN. [SAMN39739077](#), [SAMN39739078](#), [SAMN39739079](#).¹⁹

Acknowledgements

We would like to thank the Zambian field teams and the Tropical Diseases Research Centre for assisting with sample collection. Many thanks to Dr. Yoosook Lee, University of Florida for guidance with bioinformatic analysis.

References

- Mustapha AM, *et al.*: Secondary malaria vectors in western Kenya include novel species with unexpectedly high densities and parasite infection rates. *Parasit. Vectors*. 2021; 14: 252.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Burke A, *et al.*: A new malaria vector mosquito in South Africa. *Sci. Rep.* 2017; 7: 43779.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Stevenson JC, *et al.*: Detection of *Plasmodium falciparum* Infection in *Anopheles squamosus* (Diptera: Culicidae) in an Area Targeted for Malaria Elimination, Southern Zambia. *J. Med. Entomol.* 2016; 53: 1482–1487.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cooke MK, *et al.*: 'A bite before bed': exposure to malaria vectors outside the times of net use in the highlands of western Kenya.

- Malar. J.* 2015; **14**: 259.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Gebhardt ME, *et al.*: **Expanded geographic distribution and host preference of *Anopheles gambiae* (*Anopheles species 6*) in northern Zambia.** *Malar. J.* 2022; **21**: 211.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 6. da Silva AF, Machado LC, de Paula MB, *et al.*: **Culicidae evolutionary history focusing on the Culicinae subfamily based on mitochondrial phylogenomics.** *Sci. Rep.* 2020 Nov 2; **10**(1): 18823.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 7. Smith DR: **The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs?** *Brief. Funct. Genomics.* 2015; **15**: elv027–elv054.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 8. Chen DH, *et al.*: **Mitogenome - based phylogeny of mosquitoes (Diptera: Culicidae).** *Insect Sci.* 2023; **31**: 599–612.
[PubMed Abstract](#) | [Publisher Full Text](#)
 9. Dowling DK, *et al.*: **Evolutionary genetics of the mitochondrial genome: insights from *Drosophila*.** *Genetics.* 2023; **224**: 224.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 10. Hoban ML, *et al.*: **Skimming for barcodes: rapid production of mitochondrial genome and nuclear ribosomal repeat reference markers through shallow shotgun sequencing.** *PeerJ.* 2022; **10**: e13790.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 11. Chen T-Y, *et al.*: **A magnetic-bead-based mosquito DNA extraction protocol for next-generation sequencing.** *JoVE* (*Journal of Visualized Experiments*). 2021; e62354.
[Publisher Full Text](#)
 12. Dierckxsens N, Mardulyn P, Smits G: **NOVOPlasty: de novo assembly of organelle genomes from whole genome data.** *Nucleic Acids Res.* 2017; **45**: e18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 13. Nguyen VT, *et al.*: **The first genome sequence of *Anopheles squamosus* from Macha, Zambia.** *F1000Res.* 2023; **12**: 330.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 14. Bernt M, *et al.*: **MITOS: Improved de novo metazoan mitochondrial genome annotation.** *Mol. Phylogenet. Evol.* 2013; **69**: 313–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
 15. Tamura K, Stecher G, Kumar S: **MEGA11: Molecular Evolutionary Genetics Analysis Version 11.** *Mol. Biol. Evol.* 2021; **38**: 3022–3027.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 16. Ali RLMN, *et al.*: *Anopheles gambiae* mitochondrion, complete genome. [Dataset]. *GenBank.* 2024.
 17. Johns Hopkins Bloomberg School of Public Health: Complete mitogenome sequence of *Anopheles gambiae* from Nchelenge, Zambia. [Dataset]. *BioProject.* 2024.
 18. Johns Hopkins Bloomberg School of Public Health: Illumina seq of *Anopheles gambiae*. [Dataset]. *SRA.* 2024.
 19. Johns Hopkins Bloomberg School of Public Health: Invertebrate sample of *Anopheles gambiae*. [Dataset]. *Biosample.* 2024.

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 22 July 2024

<https://doi.org/10.5256/f1000research.162785.r298679>

© 2024 Makunin A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Alexey Makunin

Wellcome Sanger Institute, Hinxton, England, UK

The manuscript by Ali et al presents three mitogenome assemblies for *Anopheles gibbinsi*, a potential secondary malaria vector mosquito species, generated with low coverage shotgun Illumina sequencing. Overall, the data presentation and text quality are good. I only have several technical comments:

Introduction

"*Anopheles gibbinsi*", until recently recognized as "*An. species 6*", - unclear what this means
"The mitochondrial DNA (mtDNA) is a circular double stranded molecule that encodes 37 genes" - genome structure and number of genes varies across eukaryotes - I'd suggest specifying a taxon here

Methods

Did you do any morphological species identification before extraction?

Did you do any quantification, fragment size and purity estimation for the DNA before library prep?

There is a discrepancy in read counts reported in methods and results: "Libraries were sequenced from both ends (150 bp) on Illumina Novaseq 6000 to a depth of 13.3 million reads." vs "The sequencing from the 3 *An. gibbinsi* samples yielded an average of 29,674,206 million reads"
It might be useful to provide a summary table with sequencing statistics per sample: pre-sequencing metrics (if any available, e.g. total amount of DNA used for library construction or average fragment size), total number of reads, number of reads used in the assembly, assembled mitogenome size before and after correction

"cytochrome c oxidase I (COI) fragment spanning 8598–10,133" - in

<https://www.ncbi.nlm.nih.gov/nuccore/OR539797>, span of COI is 8595..10128

Fig 2 - "forming a single but weakly supported clade" - support value for marshalli-gibbinsi group is not shown on the tree

"Both species belong to the *An. marshallii* complex" - online databases like WRBU and Mosquito Taxonomic Inventory suggest that *An. gibbinsi* is in marshalli group, not complex. Might be useful to add a reference here.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: comparative genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 18 July 2024

<https://doi.org/10.5256/f1000research.162785.r298674>

© 2024 Ambrose L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Luke Ambrose 

The University of Queensland, Saint Lucia, Queensland, Australia

This is a useful report of an additional resource for identifying a cryptic species of *Anopheles*. I have a few suggestions that may improve the study:

1) Mitochondrial DNA alone is not sufficient in most cases to delineate species due to the potential for introgression. This is commonly observed in many diverse organisms. I suggest at minimum a discussion of this. Developing a supporting nuclear based diagnostic, such as an ITS2 RFLP would be preferable and should not be too expensive or difficult. Additionally, it would be useful to identify regions of the mitogenome that may differ between closely related species in the complex and to perform sequencing on samples from across Africa however that could be in a follow up study.

2) The language in the methods is a bit unclear - was the raw data mapped to the *An. squamosus* mitogenome? If so please clarify wording in line with this. I would also like to know average

sequencing depth of the mapped genomes rather than reporting 13.3 million reads (this isn't actually depth I assume but how many read mapped to mitogenomes?)

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genetics of *Anopheles* mosquitoes.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 21 June 2024

<https://doi.org/10.5256/f1000research.162785.r288704>

© 2024 Zadra N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Nicola Zadra 

¹ University of Trento, Trento, Italy

² NBFC, Palermo, Italy

The paper is well-written, and the message delivered is thorough. All the data obtained are available in Genbank, and they are carefully annotated. It discusses the potential role of *Anopheles gibbinsi* in malaria transmission in an area where the primary vector is close to extinction. Currently, there are only a few DNA sequences available for studying this species. Providing a complete mitochondrial genome will significantly aid in understanding its relationships with other *Anopheles* species resolving possible cryptic relations, and quickly identifying and monitoring this potential malaria vector. Data collection and sequencing are well-described and reproducible. The data collection and sequencing methods are well-documented and reproducible. The phylogenetic analysis was conducted using only the 13 protein-coding genes and presented an updated understanding of the relationships among *Anopheles gibbinsi*

and different species in the *Cellia* clade.

I have a few suggestions for improving the clarity and delivery of the paper:

1. Consider enhancing the mitochondrial dataset for phylogenetic reconstruction by including additional mitochondrial sequences from the subgenus *Cellia*. This could provide better insight into the relationship between this species and others within the *Cellia* group, as well as with the primary *Plasmodium* vector.
2. It would be helpful to include the distance to the tree. The current tree structure appears to be an ultrametric tree.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My research area is population genomics and molecular phylogenetics. I have worked on *Aedes* species mitochondrial phylogeny and related flaviviruses.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research