



HAL
open science

A chromosomal reference genome sequence for the malaria mosquito, *Anopheles moucheti*, Evans, 1925

Sandrine N Nsango, Jean-Pierre Agbor, Diego Ayala, Harriet F Johnson, Haynes Heaton, Martin G Wagah, Joanna C Collins, Ksenia Krasheninnikova, Sarah E Pelan, Damon-Lee B Pointon, et al.

► To cite this version:

Sandrine N Nsango, Jean-Pierre Agbor, Diego Ayala, Harriet F Johnson, Haynes Heaton, et al.. A chromosomal reference genome sequence for the malaria mosquito, *Anopheles moucheti*, Evans, 1925. Wellcome Open Research, 2023, 8, pp.507. 10.12688/wellcomeopenres.20259.1 . hal-04487401

HAL Id: hal-04487401

<https://hal.science/hal-04487401v1>

Submitted on 3 Mar 2024











HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



DATA NOTE

A chromosomal reference genome sequence for the malaria mosquito, *Anopheles moucheti*, Evans, 1925 [version 1; peer review: awaiting peer review]

Sandrine N. Nsango¹, Jean-Pierre Agbor¹, Diego Ayala ^{2,3}, Harriet F. Johnson ⁴, Haynes Heaton⁵, Martin G. Wagah ⁴, Joanna C. Collins⁴, Ksenia Krasheninnikova⁴, Sarah E. Pelan ⁴, Damon-Lee B. Pointon ⁴, Ying Sims⁴, James W. Torrance ⁴, Alan Tracey⁴, Marcela Uliano Da Silva⁴, Jonathan M. D. Wood ⁴, Katharina von Wyszetzki⁴, DNA Pipelines collective, Shane A. McCarthy ^{4,6}, Daniel E. Neafsey^{7,8}, Alex Makunin ⁴, Mara Lawniczak ⁴

¹Faculte de Medecine et des Sciences Pharmaceutiques, Universite de Douala, Douala, Littoral, Cameroon

²MIVEGEC, IRD, Montpellier, France

³ESV-GAB, Centre Interdisciplinaire de Recherches Médicales de Franceville (CIRMF), Franceville, Gabon

⁴Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

⁵CSSE, Auburn University, Auburn, Alabama, USA

⁶Department of Genetics, University of Cambridge, Cambridge, England, UK

⁷Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁸Infectious Disease and Microbiome Program, Broad Institute, Cambridge, MA, USA

V1 First published: 08 Nov 2023, 8:507
<https://doi.org/10.12688/wellcomeopenres.20259.1>

Latest published: 08 Nov 2023, 8:507
<https://doi.org/10.12688/wellcomeopenres.20259.1>

Abstract

We present a genome assembly from an individual male *Anopheles moucheti* (the malaria mosquito; Arthropoda; Insecta; Diptera; Culicidae), from a wild population in Cameroon. The genome sequence is 271 megabases in span. The majority of the assembly is scaffolded into three chromosomal pseudomolecules with the X sex chromosome assembled. The complete mitochondrial genome was also assembled and is 15.5 kilobases in length.

Keywords

Anopheles moucheti, African malaria mosquito, genome sequence, chromosomal

Open Peer Review

Approval Status AWAITING PEER REVIEW

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



This article is included in the [Tree of Life](#) gateway.

Corresponding author: Mara Lawniczak (mara@sanger.ac.uk)

Author roles: **Nsango SN:** Investigation; **Agbor JP:** Investigation; **Ayala D:** Investigation; **Johnson HF:** Investigation, Methodology; **Heaton H:** Investigation, Methodology; **Wagah MG:** Visualization; **Collins JC:** Data Curation; **Krasheninnikova K:** Methodology; **Pelan SE:** Data Curation; **Pointon DLB:** Data Curation; **Sims Y:** Data Curation; **Torrance JW:** Data Curation; **Tracey A:** Data Curation; **Uliano Da Silva M:** Investigation, Methodology; **Wood JMD:** Data Curation; **von Wyszetzki K:** Methodology; **McCarthy SA:** Investigation, Methodology; **Neafsey DE:** Conceptualization, Funding Acquisition; **Makunin A:** Formal Analysis, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation; **Lawniczak M:** Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by a Bill & Melinda Gates Foundation Award (INV-009760) to ML. The Wellcome Sanger Institute is funded by the Wellcome Trust (206194), which supports ML. SAM is supported by the Wellcome Trust Grant (WT207492). Field collection was supported by an ANR grant (ANR-18-CE35-0002-01 –WILDING) to DA.

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

Copyright: © 2023 Nsango SN *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: Nsango SN, Agbor JP, Ayala D *et al.* **A chromosomal reference genome sequence for the malaria mosquito, *Anopheles moucheti*, Evans, 1925 [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, 8:507 <https://doi.org/10.12688/wellcomeopenres.20259.1>

First published: 08 Nov 2023, 8:507 <https://doi.org/10.12688/wellcomeopenres.20259.1>

Species taxonomy

Animalia; Arthropoda; Insecta; Diptera; Culicidae; Anophelelinae; Anopheles; *Anopheles moucheti*; Evans, 1925 (NCBI txid:186751).

Background

Anopheles moucheti moucheti (hereafter *An. moucheti*) has the greatest geographical distribution within the *Moucheti* group, which includes *An. bervoetsi* and *An. nigeriensis*^{1,2}. These species can only be distinguished by slight morphological characters³ and / or using molecular tools^{4,5}. *An. moucheti* is widely distributed across forested areas of West and Central Africa, and the persistence of this species is linked to its ability to lay eggs in lentic streams and rivers where its larvae develop preferentially and predominate over other malaria vectors^{1,6}. *An. moucheti* females usually feed indoors and exhibit a preference for feeding on humans⁷ although the species can also be found in forested areas of Gabon, away from any human presence^{8,9}. *An. moucheti* is among the most important vectors of human malaria in the equatorial rain forest of Africa^{1,10}. It is present in a large number of countries, from Guinea to Kenya¹¹, and it has been found naturally infected with *Plasmodium* parasites across Central African countries, sustaining year-round transmission¹. For instance, it can show infective biting rates up to 300 bites per person per year in villages located along slow moving rivers^{7,10}. Moreover, *An. moucheti* has been recently incriminated for contributing to the origin of malaria in humans, being a potential bridge vector for malaria parasites from non-human primates to humans^{8,9}.

The main genetic work carried out on *An. moucheti* has been focused on differentiation between members of the group. To this end, molecular markers have been developed that are based on the sequence variation in the ribosomal internal transcribed spacer ITS1 region⁵. Microsatellite loci have also been developed showing a low level of genetic differentiation among populations in Cameroon^{5,12}. Several polymorphic inversions have been detected in *An. moucheti*, although their role in local adaptation or speciation is not yet understood¹³. Therefore, genomic characterization of this neglected but major malaria vector, *An. moucheti*, is critical for developing sustainable control strategies for malaria elimination in Central Africa.

The genome of the African malaria mosquito, *Anopheles moucheti*, was sequenced as part of the Anopheles Reference Genomes Project (PRJEB51690). Here we present a chromosomally complete genome sequence for *Anopheles moucheti*, based on a single male specimen from Ebogo, Cameroon.

Genome sequence report

The genome was sequenced from a single male *Anopheles moucheti* collected from Ebogo, Cameroon (3.384, 11.466). A total of 41-fold coverage in Pacific Biosciences single-molecule HiFi long reads (N50 13.064 kb) were generated.

Primary assembly contigs were scaffolded with chromosome conformation Hi-C data from a sibling male mosquito. Manual assembly curation corrected 48 missing joins or misjoins and removed 3 haplotypic duplications, reducing the scaffold number by 7.8% and reducing assembly size by 1.8%.

The final assembly has a total length of 271 Mb in 345 sequence scaffolds with a scaffold N50 of 92.649 Mb (Table 1). 84.7% of the assembly sequence was assigned to three chromosomal-level scaffolds, representing two autosomes (numbered and oriented against the *An. gambiae* PEST strain assembly AgamP3 (14; GCF_000005575.2) with chromosome arms homologies for *Myzomyia* series taken from 15), and the X sex chromosome (Figure 1–Figure 6; Table 2).

Table 1. Genome data for *Anopheles moucheti*, idAnoMoucSN_F20_07.

Project accession data	
Assembly identifier	idAnoMoucSN_F20_07
Species	<i>Anopheles moucheti</i>
Specimen	idAnoMoucSN-F20_07
NCBI taxonomy ID	186751
BioProject	PRJEB53257
BioSample ID	ERS10527373
Isolate information	male, whole organism
Raw data accessions	
PacificBiosciences SEQUEL II	ERR9439507
Hi-C Illumina	ERR9356830
PolyA RNA-Seq Illumina	ERR9356831, ERR9356832
Genome assembly	
Assembly accession	GCA_943734755
Accession of alternate haplotype	GCA_943734785
Span (Mb)	271.317
Number of contigs	406
Contig N50 length (Mb)	4.744
Number of scaffolds	345
Scaffold N50 length (Mb)	92.649
Longest scaffold (Mb)	103.512
BUSCO* genome score	C:97.6%[S:97.3%,D:0.3%], F:0.7%,M:2.0%,n:3285

* BUSCO scores based on the diptera_odb10 BUSCO set using BUSCO 5.3.2. C=complete [S=single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/idAnoMoucSN_F20_07/dataset/CALSEA01/busco.

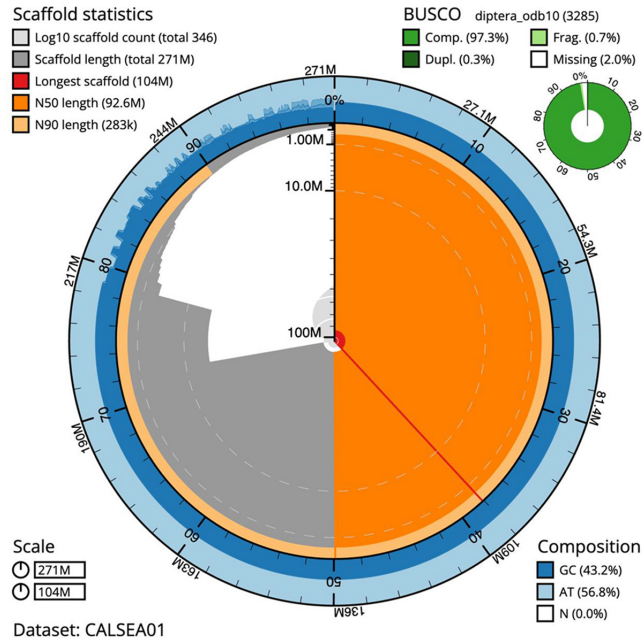


Figure 1. Snail plot summary of assembly statistics for *Anopheles moucheti* assembly idAnoMoucSN_F20_07. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 271,332,225 bp assembly. The distribution of sequence lengths is shown in dark grey with the plot radius scaled to the longest sequence present in the assembly (103,511,506 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 sequence lengths (92,649,126 and 283,162 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the diptera_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/idAnoMoucSN_F20_07/dataset/CALSEA01/snail.

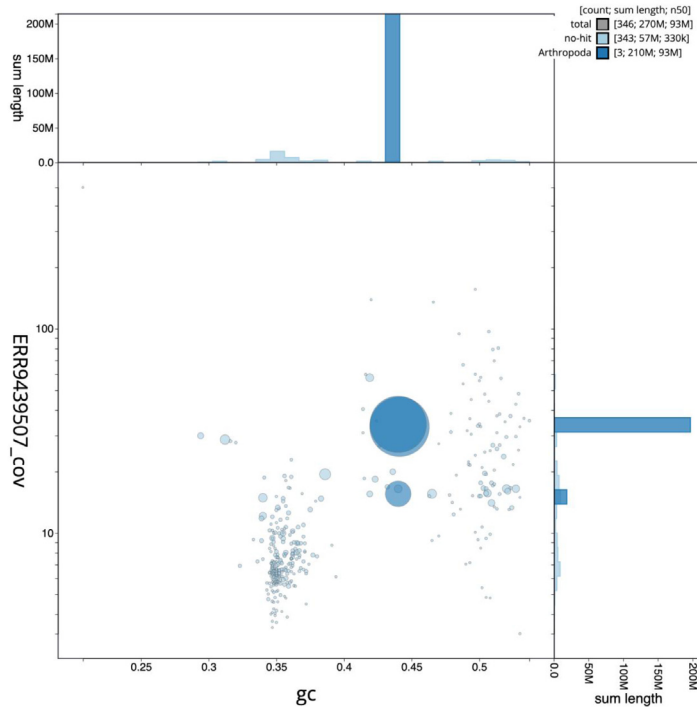


Figure 2. Blob plot of base coverage in idAnoMoucSN-F20_07 PacBio HiFi reads against GC proportion for *An. moucheti* assembly idAnoMoucSN_F20_07. Chromosomes are coloured by phylum. Circles are sized in proportion to chromosome length. Histograms show the distribution of chromosome length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/idAnoMoucSN_F20_07/dataset/CALSEA01/blob.

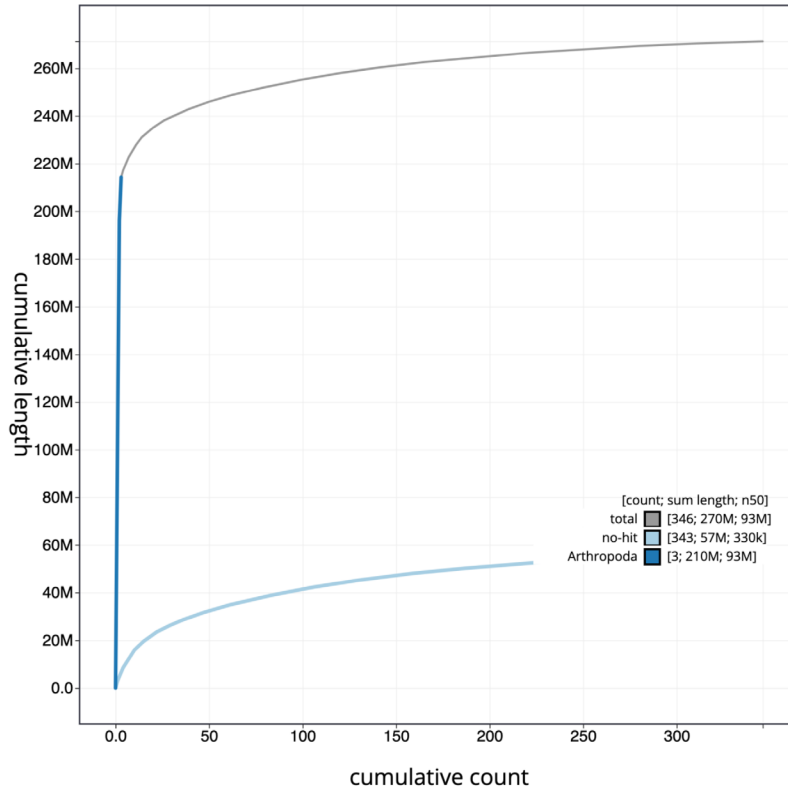


Figure 3. Cumulative chromosome length for *An. moucheti* assembly idAnoMoucSN_F20_07. The grey line shows cumulative length for all chromosomes. Coloured lines show cumulative lengths of chromosomes assigned to each phylum using the buscogenes taxrule. The interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/idAnoMoucSN_F20_07/dataset/CALSEA01/cumulative.

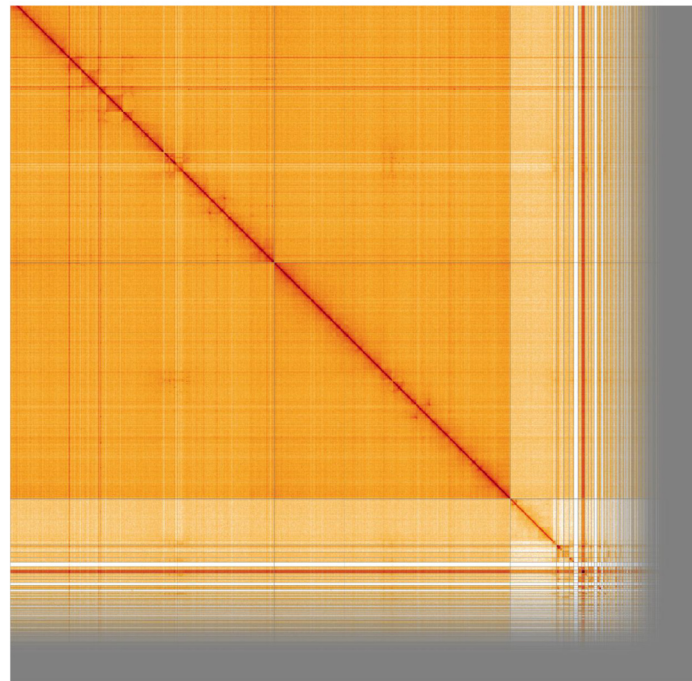


Figure 4. Genome assembly of *An. moucheti*, idAnoMoucSN_F20_07: Hi-C contact map. Visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. X chromosome signal is lower as individuals used for both PacBio and HiC are males. The interactive Hi-C map can be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/?d=GZ6DShODRcSMNgw7a0UIfQ>.

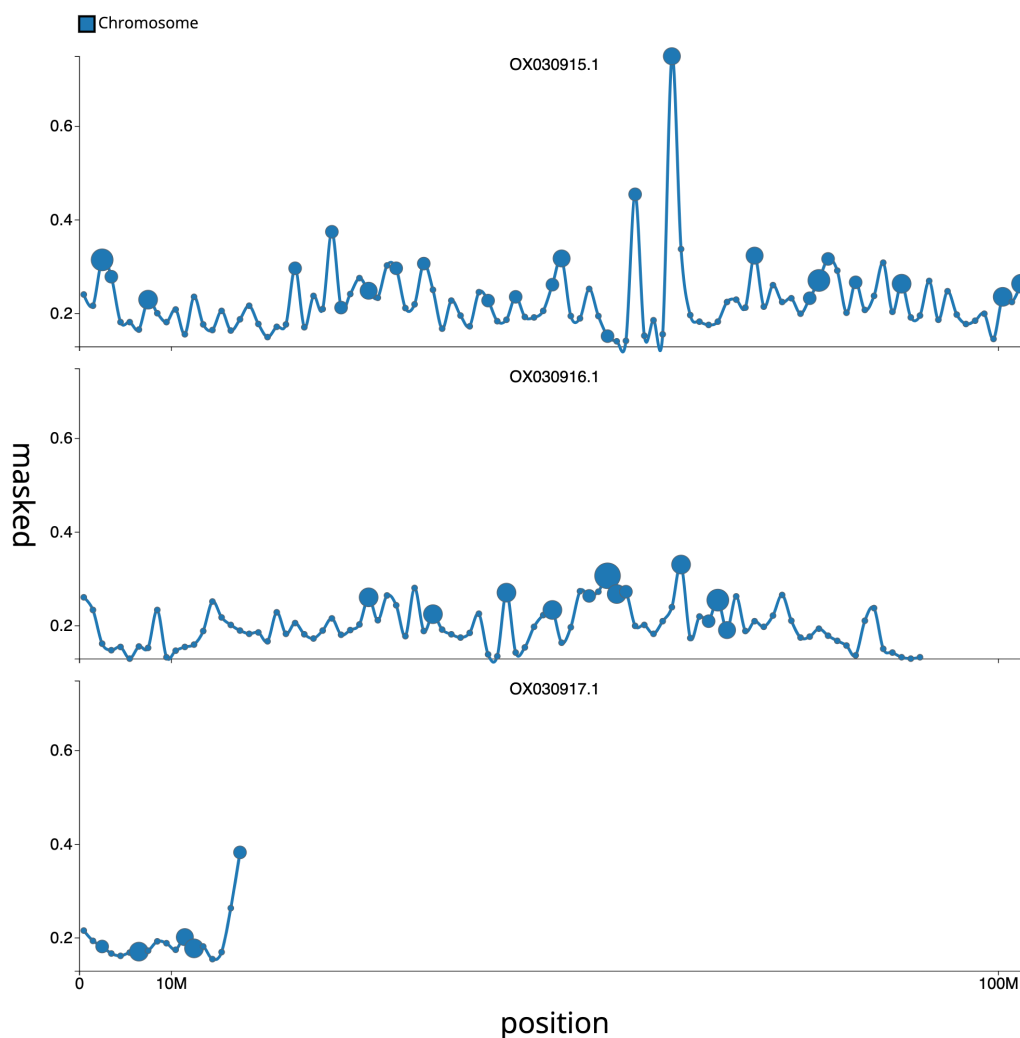


Figure 5. Distribution plot of repeat masked sequence proportion against genomic position for chromosomes of *An. moucheti*, idAnoMoucSN_F20_07. 1Mb windows with assembly gaps highlighted by larger blue circles, size is proportional to total size of gaps in the window (max 7800bp per window).

The assembly has a BUSCO 5.3.2¹⁶ completeness of 97.6% using the diptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition and nucleic acid extraction

Anopheles moucheti offspring were reared from a wild caught gravid female collected from Ebogo, Cameroon (3.384, 11.466) by Sandrine Eveline Nsango. A single male idAnoMoucSN-F20_07 was used for Pacific BioSciences, a single sibling male idAnoMoucSN-F20_09 was used for Arima Hi-C.

For high molecular weight (HMW) DNA extraction one whole insect (idAnoMoucSN-F20_07) was disrupted by manual grinding with a blue plastic pestle in Qiagen MagAttract lysis buffer¹⁷ and then extracted using the Qiagen MagAttract

HMW DNA extraction kit with two minor modifications including halving volumes recommended by the manufacturer due to small sample size (*Anopheles* mosquitoes typically weigh 2–3 mg) and running two elution steps of 100 μ l each to increase DNA yield. The quality of the DNA was evaluated using an Agilent FemtoPulse to ensure that most DNA molecules were larger than 30 kb, and preferably > 100 kb. In general, single mosquito extractions ranged in total estimated DNA yield from 200 ng to 800 ng, with an average yield of 500 ng. Low molecular weight DNA was removed using 0.8X AMPure XP purification. DNA was sheared to an average fragment size of 12–20 kb using a Diagenode Megaruptor 3 at speeds ranging from 27 to 30. Sheared DNA was purified using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration and quality of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer

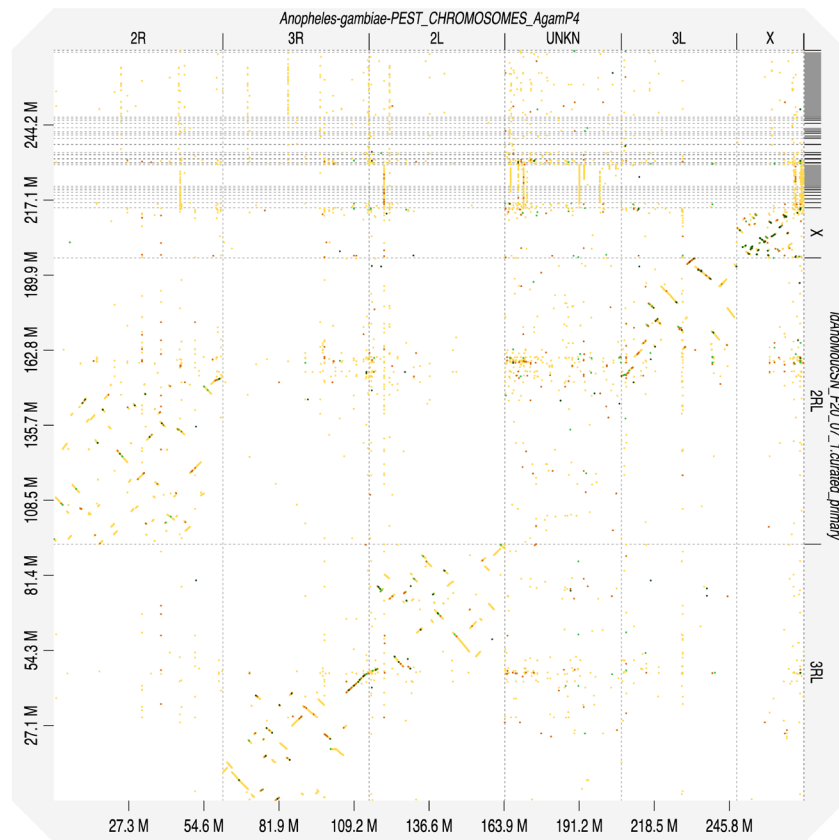


Figure 6. Alignment dotplot between genome assemblies of *An. moucheti* (idAnoMoucSN_F20_07) and *An. gambiae*, AgamP4 (PEST). Visualised in DGenies. Chromosome arms correspondence (moucheti-gambiae): 2R-2R, 2L-3L, 3R-3R, 3L-2L in agreement with 15.

Table 2. Chromosomal pseudomolecules in the genome assembly of *An. moucheti*, idAnoMoucSN_F20_07.

INSDC accession	Chromosome	Size (Mb)	Count	Gaps
OX030915.1	2RL	103.512	1	30
OX030916.1	3RL	92.649	1	15
OX030917.1	X	18.119	1	6
OX030918.1	MT	0.016	1	0
	X Unlocalised	15.547	98	0
	Unplaced	41.489	244	10

and Qubit Fluorometer with the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sheared and cleaned sample on the FemtoPulse system once more. The median DNA fragment size for *Anopheles* mosquitoes was 15 kb and the median yield of sheared DNA was 200 ng, with samples typically losing about 50% of the original estimated DNA quantity through the process of shearing and purification.

For Hi-C data generation, a sibling male mosquito specimen (idAnoMoucSN-F20_09) was used as input material for the Arima V2 Kit according to the manufacturer's instructions for animal tissue. This approach of using a sibling was followed in order to enable all material from a single specimen to contribute to the PacBio data generation given we were not always able to meet the minimum suggested guidance of starting with > 300 ng of HMW DNA from a specimen. Samples proceeded to the Illumina library prep stage even if they were suboptimal (too little tissue) going into the Arima reaction.

To assist with annotation, which will be made available through VectorBase¹⁸ in due course, RNA was extracted from separate whole unrelated lab-reared male and wild-caught female insect specimens (idAnoMoucSN-F4_01, idAnoMoucSN-F7_x) using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 µl RNase-free water, and its concentration was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay. Samples were not always ideally preserved for RNA, so qualities varied but all were sequenced anyway.

Sequencing

We prepared libraries as per the PacBio procedure and checklist for SMRTbell Libraries using Express TPK 2.0 with low DNA input. Every library was barcoded to support multiplexing. Final library yields ranged from 20 ng to 100 ng, representing only about 25% of the input sheared DNA. Libraries from two specimens were typically multiplexed on a single 8M SMRT Cell. Sequencing complexes were made using Sequencing Primer v4 and DNA Polymerase v2.0. Sequencing was carried out on the Sequel II system with 24-hour run time and 2-hour pre-extension. For Hi-C data generation, following the Arima Hi-C V2 reaction, samples were processed through Library Preparation using a NEB Next Ultra II DNA Library Prep Kit and sequenced aiming for 100x depth. RNA libraries were created using the directional NEB Ultra II stranded kit. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi), Illumina NovaSeq 6000 (10X and Hi-C), or Illumina HiSeq 4000 (RNAseq) instruments.

Genome assembly

Assembly was carried out with Hifiasm¹⁹; haplotypic duplications were identified and removed with purge_dups²⁰. The assembly was then scaffolded with Hi-C data²¹ using SALSA2²². The assembly was checked for contamination as described previously²³. Manual curation was performed using gEVAL²⁴, HiGlass²⁵ and PretextView²⁶. The mitochondrial genome was assembled using MitoHiFi²⁷, which performs annotation using MitoFinder²⁸. The genome was analysed and BUSCO scores were generated within the BlobToolKit environment²⁹. Synteny analysis was performed with D-GENIES³⁰ and minimap2³¹. Table 3 contains a list of all software tool versions used, where appropriate.

Ethics/compliance issues

The genetic resources accessed and utilised under this project were done so in accordance with the UK ABS legislation (Nagoya Protocol (Compliance) (Amendment) (EU Exit) Regulations 2018 (SI 2018/1393)) and the national ABS legislation within the country of origin, where applicable.

Table 3. Software tools used.

Software tool	Version	Source
hifiasm	0.14	19
purge_dups	1.2.3	20
SALSA2	2.2-4c80ac1	22
MitoHiFi	2	27
gEVAL	N/A	24
HiGlass	1.11.6	25
PretextView	0.1.x	26
BlobToolKit	3.4.0	32
BUSCO	5.3.2	16
D-GENIES	1.4	30
minimap2	2.24	31

Data availability

European Nucleotide Archive: *Anopheles moucheti* genome assembly, idAnoMoucSN_F20_07. Accession number PRJEB53257; <https://identifiers.org/bioproject/PRJEB53257>.

The genome sequence is released openly for reuse. The *Anopheles moucheti* genome sequencing initiative is part of the Anopheles Reference Genomes project PRJEB51690. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the Sanger Scientific Operations: DNA Pipelines collective are listed here: <https://doi.org/10.5281/zenodo.4790455>.

References

- Antonio-Nkondjio C, Simard F: **Highlights on *Anopheles nili* and *Anopheles moucheti*, Malaria Vectors in Africa.** In: Manguin S editor. *Anopheles Mosquitoes: New Insights into Malaria Vectors*. Rijeka (HR): InTech; 2013. [PubMed Abstract](#)
- Harbach RE, Wilkerson RC: **The insupportable validity of mosquito subspecies (Diptera: Culicidae) and their exclusion from culicid classification.** *Zootaxa*. 2023; **5303**(1): 1–184. [PubMed Abstract](#) | [Publisher Full Text](#)
- Gillies MT, De Meillon B: **The anophelinae of Africa south of the Sahara (Ethiopian zoogeographical region).** South African Institute for Medical Research; 1968. [Reference Source](#)
- Antonio-Nkondjio C, Ndo C, Awono-Ambene P, et al.: **Population genetic structure of the malaria vector *Anopheles moucheti* in south Cameroon forest region.** *Acta Trop*. 2007; **101**(1): 61–68. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kengne P, Antonio-Nkondjio C, Awono-Ambene HP, et al.: **Molecular differentiation of three closely related members of the mosquito species complex, *Anopheles moucheti*, by mitochondrial and ribosomal DNA polymorphism.** *Med Vet Entomol*. 2007; **21**(2): 177–182. [PubMed Abstract](#) | [Publisher Full Text](#)
- Ayala D, Costantini C, Ose K, et al.: **Habitat suitability and ecological niche profile of major malaria vectors in Cameroon.** *Malar J*. 2009; **8**: 307. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Antonio-Nkondjio C, Kerah CH, Simard F, et al.: **Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission.** *J Med Entomol*. 2006; **43**(6): 1215–1221. [PubMed Abstract](#) | [Publisher Full Text](#)

8. Paupy C, Makanga B, Ollomo B, *et al.*: **Anopheles moucheti and Anopheles vinckei are candidate vectors of ape Plasmodium parasites, including Plasmodium praefalciparum in Gabon.** *PLoS One*. 2013; **8**(2): e57294.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Makanga B, Yangari P, Rahola N, *et al.*: **Ape malaria transmission and potential for ape-to-human transfers in Africa.** *Proc Natl Acad Sci U S A*. 2016; **113**(19): 5329–5334.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Mouchet J: **Biodiversité du paludisme dans le monde.** John Libbey Eurotext; 2004.
[Reference Source](#)
11. Irish SR, Kyallo D, Snow RW, *et al.*: **Updated list of Anopheles species (Diptera: Culicidae) by country in the Afrotropical Region and associated islands.** *Zootaxa*. 2020; **4747**(3): zootaxa.4747.3.1.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Annan Z, Durand P, Ayala FJ, *et al.*: **Population genetic structure of Plasmodium falciparum in the two main African vectors, Anopheles gambiae and Anopheles funestus.** *Proc Natl Acad Sci U S A*. 2007; **104**(19): 7987–7992.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Sharakhova MV, Antonio-Nkondjio C, Xia A, *et al.*: **Polymorphic chromosomal inversions in Anopheles moucheti, a major malaria vector in Central Africa.** *Med Vet Entomol*. 2014; **28**(3): 337–340.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Sharakhova MV, Hammond MP, Lobo NF, *et al.*: **Update of the Anopheles gambiae PEST genome assembly.** *Genome Biol*. 2007; **8**(1): R5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Green CA, Hunt RH: **Interpretation of variation in ovarian polytene chromosomes of Anopheles Funestus Giles, A. Pansis Gillies, and A. Aruni?** *Genetica*. 1980; **51**: 187–195.
[Publisher Full Text](#)
16. Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics*. 2015; **31**(19): 3210–3212.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Teltscher F, Johnson H, Lawniczak M: **Manual extraction of High Molecular Weight DNA from single mosquitoes using the Qiagen MagAttract HMW DNA kit.** 2023; [cited 2 Oct 2023].
[Publisher Full Text](#)
18. Giraldo-Calderón GI, Harb OS, Kelly SA, *et al.*: **VectorBase.org updates: bioinformatic resources for invertebrate vectors of human pathogens and related organisms.** *Curr Opin Insect Sci*. 2022; **50**: 100860.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.** *Nat Methods*. 2021; **18**(2): 170–175.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics*. 2020; **36**(9): 2896–2898.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell*. 2014; **159**(7): 1665–1680.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol*. 2019; **15**(8): e1007273.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience*. Oxford University Press, 2021; **10**(1): gaa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a web-based browser for evaluating genome assemblies.** *Bioinformatics*. 2016; **32**(16): 2508–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol*. 2018; **19**(1): 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. **PretextView: OpenGL Powered Pretext Contact Map Viewer.** Github; [Accessed 19 October 2022].
[Reference Source](#)
27. Uliano-Silva M, Nunes JGRN, Krashennikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0. 2021.**
[Publisher Full Text](#)
28. Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour*. 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda)*. 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Cabanettes F, Klopp C: **D-GENIES: dot plot large genomes in an interactive, efficient and simple way.** *PeerJ*. 2018; **6**: e4958.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics*. 2018; **34**(18): 3094–3100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda)*. 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)