DATA NOTE



Two chromosomal reference genome sequences for the

malaria mosquito, Anopheles (Nyssorhynchus) darlingi, Root,

1926 from French Guiana and Peru

[version 1; peer review: 3 approved]

Mathilde Gendrin ¹¹, Katy Heu¹, Marta Moreno ¹², Dionicia Gamboa ¹³, Joseph M Vinetz ¹⁴, Carlos Tong⁵, Jan E Conn^{6,7}, Harriet F Johnson ¹⁸, Haynes Heaton⁹, Martin G Wagah ¹¹⁰, Joanna C Collins¹⁰, Ksenia Krasheninnikova¹⁰, Sarah E Pelan ¹¹⁰, Damon-Lee B Pointon ¹¹⁰, James W Torrance ¹¹⁰, Alan Tracey¹⁰, Marcela Uliano-Silva¹⁰, Jonathan M D Wood ¹¹⁰, Katharina von Wyschetzki¹⁰, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Shane A McCarthy ^{10,11}, Mara K N Lawniczak ¹⁰¹⁰, Daniel E Neafsey^{12,13}, Alex Makunin ¹⁰¹⁰

¹Microbiota of Insect Vectors Group, Institut Pasteur de la Guyane, Cayenne, Cayenne, French Guiana

²Department of Infection Biology, London School Of Hygiene and Tropical Medicine, London, England, UK

³Laboratorio de Malaria: Parásitos y vectores, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias e Ingeniería,

Universidad Peruana Cayetano Heredia, Lima District, Lima Region, Peru

⁴Section of Infectious Disease, Department of Internal Medicine, School of Medicine, Yale University, New Haven, Connecticut, USA ⁵GERESA Loreto Gerencia Regional de Salud Loreto, Unidad de Entomología, Iquitos, Peru

⁶Wadsworth Center, New York State Department of Health, Albany, NY, USA

⁷Department of Biomedical Sciences, College of Integrated Health Sciences, SUNY The State University of New York, Albany, New York, USA

⁸Scientific Operations, Wellcome Sanger Institute, Hinxton, England, UK

⁹CSSE, Auburn University, Auburn, Alabama, USA

¹⁰Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

¹¹University of Cambridge Department of Genetics, Cambridge, England, UK

¹²Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA

¹³Infectious Disease and Microbiome Program, Broad Institute, Cambridge, Massachusetts, USA

v1	First published: 10 Apr 2025, 10:187 https://doi.org/10.12688/wellcomeopenres.23989.1	Open Peer Review			
	Latest published: 10 Apr 2025, 10 :187 https://doi.org/10.12688/wellcomeopenres.23989.1	Approval S	tatus 🗹	~~	
			1	2	3
Abs We	tract present two genome assemblies, each generated from individual	version 1	~	~	~

We present two genome assemblies, each generated from individual female *Anopheles* (*Nyssorhynchus*) *darlingi* (the malaria mosquito; Arthropoda; Insecta; Diptera; Culicidae), from wild populations in French Guiana and Peru. The genome sequences are approximately

view

view

10 Apr 2025

180 megabases in span. The majority of each assembly is scaffolded into three chromosomal pseudomolecules with the X sex chromosome assembled. The complete mitochondrial genomes were also assembled and are both 15.4 kilobases in length. The assemblies differ by two inversions in chromosome arm 2R.

Keywords

Anopheles darlingi, South American malaria vector, genome sequence, chromosomal



This article is included in the Tree of Life gateway.

- 1. **Craig Wilding** D, University College Dublin School of Biology and Environmental Science (Ringgold ID: 123708), Dublin, Ireland
- 2. **Renee Ali**, Johns Hopkins University Bloomberg School of Public Health, Baltimore, USA
- 3. Komal Kumar Raja 🕩, Baylor College of Medicine, Houston, USA

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

Corresponding authors: Mara K N Lawniczak (mara@sanger.ac.uk), Alex Makunin (am60@sanger.ac.uk)

Author roles: Gendrin M: Investigation; Heu K: Investigation; Moreno M: Investigation; Gamboa D: Investigation; Vinetz JM: Investigation; Tong C: Investigation; Conn JE: 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; Torrance JW: Data Curation; Tracey A: Data Curation; Uliano-Silva M: Investigation, Methodology; Wood JMD: Data Curation; von Wyschetzki K: Methodology; McCarthy SA: Investigation, Methodology; Lawniczak MKN: Conceptualization, Funding Acquisition, Investigation, Supervision, Writing – Review & Editing; Neafsey DE: Conceptualization, Funding Acquisition; Makunin A: Formal Analysis, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation

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 (220540), which supports ML. SAM is supported by the Wellcome Trust Grant (WT207492). This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant Number U19AI110818 to the Broad Institute (DEN). The Peruvian Anopheles (Nyssorhynchus) darlingi colonization project was funded by US Public Health Service, National Institutes of Health, National Institute of Allergy and Infectious Diseases Cooperative Agreement U19AI089681 (Amazonian ICEMR) to JMV and NIH R01AI54139-02 to JEC.

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

Copyright: © 2025 Gendrin M *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: Gendrin M, Heu K, Moreno M *et al.* Two chromosomal reference genome sequences for the malaria mosquito, *Anopheles* (*Nyssorhynchus*) *darlingi*, Root, 1926 from French Guiana and Peru [version 1; peer review: 3 approved] Wellcome Open Research 2025, **10**:187 https://doi.org/10.12688/wellcomeopenres.23989.1

First published: 10 Apr 2025, 10:187 https://doi.org/10.12688/wellcomeopenres.23989.1

Species taxonomy

Animalia; Arthropoda; Insecta; Diptera; Culicidae; Anophelinae; Anopheles; *Anopheles darlingi*; Root, 1926 (NCBI txid:43151).

Background

The Anopheles darlingi mosquito is the most important Neotropical malaria vector, spanning an enormous geographic range from southern Mexico to northern Argentina^{1,2}. It belongs to a subgenus (Nyssorhynchus) that exhibits considerable evolutionary divergence from other anopheline subgenera³ harbouring familiar malaria vectors from Africa and Asia, resulting in the proposal to recognize Nyssorhynchus as a genus-level classification⁴. Given the importance of An. darlingi as a neotropical vector, the first short-read genome assembly was published in 2013, based on a collection of male and female mosquitoes from the Brazilian Amazon⁵. The resulting assembly provided a first understanding of this divergent lineage of mosquitoes, but was highly fragmented (over 8000 scaffolds, N50 scaffold length = 81 kb) due to heterozygosity in the sequencing template from many individual mosquitoes and the inherent difficulty of assembling highly contiguous genomes from short reads.

Consequently, improved genome assemblies of An. darlingi were produced as part of the Anopheles Reference Genomes Project (PRJEB51690). Here we present two chromosomally complete genome sequences for Anopheles darlingi, based on single female specimens from Macouria, French Guiana and Cahuide, Loreto Department, Peru. The latter is derived from the first sustained colony of Anopheles (Nyssorhynchus) darlingi established in 2013 from 135 wild-caught female mosquitoes obtained in the northeastern Peruvian Amazon region of Iquitos in the Loreto Department⁶, for which maintenance of heterozygosity was demonstrated. This colony had no significant evidence of a bottleneck, decrease in total alleles, or increase in inbreeding compared with field specimens (founder population). Low-moderate differentiation between field and laboratory populations was detected⁷. The genome assembly for the specimen from French Guiana has been annotated by NCBI RefSeq to support its use as an updated reference for the species.

Genome sequence report for specimen from French Guiana

The genome was sequenced from a single female *Anopheles darlingi* collected from Macouria, French Guiana (5.015, -52.528). A total of 32-fold coverage in Pacific Biosciences single-molecule HiFi long reads (N50 11.387 kb) were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data from a female sibling. Manual assembly curation corrected 13 missing joins or misjoins, reducing the scaffold number by 5.8%.

The final assembly has a total length of 182 Mb in 66 sequence scaffolds with a scaffold N50 of 94.952 Mb (Table 1). 98.89% of the assembly sequence was assigned to three chromosomal-level scaffolds, representing two autosomes

(numbered and oriented against the *An. albimanus* STECLA strain assembly (8; GCA_013758885.1)), and the X sex chromosome (Figure 1–Figure 4; Table 2).

The assembly has a BUSCO 5.3.2⁹ completeness of 97.2% 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.

Genome sequence report for specimen from Peru

The genome was sequenced from a single female *Anopheles darlingi* from an established mosquito colony that originated from Cahuide, Peru (-4.230, -73.460). Unfortunately, this colony no longer exists and is therefore not available for further research. A total of 68-fold coverage in Pacific Biosciences single-molecule HiFi long reads (N50 10.925 kb) and 113-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data from a female sibling. Manual assembly curation corrected 26 missing joins or misjoins and removed 1 duplicated haplotig.

The final assembly has a total length of 182 Mb in 54 sequence scaffolds with a scaffold N50 of 95.091 Mb (Table 3). 98.92% of the assembly sequence was assigned to three chromosomal-level scaffolds, representing two autosomes and the X sex chromosome (Table 4). Synteny between this assembly and that of *An. albimanus* AalbS3 showed identical chromosome arms, with numerous rearrangements within the arms (Figure 5). Compared to the assembly generated from a female from French Guiana, this assembly has two large scale inversions in chromosome arm 2R (Figure 6; Table 5).

The assembly has a BUSCO $5.3.2^{9}$ completeness of 97.0% using the diptera_odb10 reference set.

Chromosome arms delineation

To facilitate the delineation of chromosome arms in both assemblies, we predicted most common tandem repeats using TRASH¹⁰. In autosomes, largest blocks of tandem repeats coincided with expected centromere locations and were mostly composed of repeats with 353 base long monomers (Figure 7, Figure 8). We used outermost copies of this repeat to annotate chromosome arms (Table 6). In the X chromosome, no large tandem repeat blocks were observed on the centromere adjacent side, i.e. at the end of the chromosome sequence. In both assemblies, the majority of chromosome arms ended with tandem repeats with 281 base long monomers, suggesting that these repeats act as telomeres – non-canonical satellite repeats are characteristic for other *Anopheles* species⁸.

Methods

Sample acquisition and nucleic acid extraction

Anopheles darlingi offspring were reared from a wild caught gravid female collected from Macouria, French Guiana (5.015, -52.528) by Katy Heu and Romuald Carinci in April 2019. Individuals from the resulting brood were shipped

Project accession data	
Assembly identifier	idAnoDarlMG_H_01
Species	Anopheles darlingi
Specimen	idAnoDarlMG-H_01
NCBI taxonomy ID	43151
BioProject	PRJEB53253
BioSample ID	ERS10527351
Isolate information	female, whole organism
Raw data accessions	
PacificBiosciences SEQUEL II	ERR9439499
Hi-C Illumina	ERR9356785, ERR9356786
PolyA RNA-Seq Illumina	ERR9356791, ERR9356792, ERR9356793
Genome assembly	
Assembly accession	GCA_943734745
Accession of alternate haplotype	GCA_943734775
Span (Mb)	181.653
Number of contigs	96
Contig N50 length (Mb)	19.203
Number of scaffolds	66
Scaffold N50 length (Mb)	94.952
Longest scaffold (Mb)	94.952
BUSCO* genome score	C:97.2%[S:96.7%,D:0.5%],F:0.6%,M:2.2%,n:3285

Table 1. Genome data for	Anopheles	darlingi from	French	Guiana,
idAnoDarlMG_H_01.	-	-		

* 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/ idAnoDarlMG_H_01/dataset/CALSDZ01/busco.

on dry ice to the UK. From this brood, a single female idAnoDarlMG-H_01 was used for Pacific BioSciences and its sibling female idAnoDarlMG-H_08 was used for Arima Hi-C.

Anopheles darlingi offspring from Peru were reared from a colony in April 2019. The original collection to create this colony used wild females collected from Cahuide, Loreto, Peru (-4.230, -73.460) by the Amazonian International Center of Excellence in Malaria Research (Amazonian ICEMR) team in a long-term project led by Dionicia Gamboa, Marta Moreno, Jan Conn and Joseph Vinetz⁶. A single mated female from

this colony was used to rear a brood to be used for sequencing. From this brood, female sibling idAnoDarlJC-H15_27 was used for Pacific BioSciences and 10x linked reads, and female sibling idAnoDarlJC-H15_22 was used for Arima Hi-C.

For both reference genomes, high molecular weight (HMW) DNA extraction used one whole insect disrupted by manual grinding with a blue plastic pestle in Qiagen MagAttract lysis buffer and then extracted using the Qiagen MagAttract HMW DNA extraction kit with two minor modifications¹¹. The quality of the DNA was evaluated using an Agilent FemtoPulse to



Figure 1. Snail plot summary of assembly statistics for *Anopheles darlingi* **assembly idAnoDarlMG_H_01.** The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 181,652,731 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 (94,951,917 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 sequence lengths (94,951,917 and 71,270,736 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/idAnoDarlMG_H_01/dataset/CALSDZ01/snail.

ensure that most DNA molecules were larger than 30 kb, and preferably > 100 kb. In general, single mosquito extractions range 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 a 0.8X AMpure XP purification. For the specimen from Peru, a small aliquot (less than ~5% of the

total volume) of HMW DNA was set aside for 10X Linked Read sequencing. For PacBio sequencing, 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



Figure 2. Blob plot of base coverage in idAnoDarlMG_H_01 PacBio HiFi reads against GC proportion for *An. darlingi* assembly idAnoDarlMG_H_01. 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/idAnoDarlMG_H_01/dataset/CALSDZ01/blob.

DNA sample. The concentration and quality of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer 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, separate sibling mosquito specimens were used as input material for the Arima V2 Kit according to the manufacturer's instructions for animal tissue. This approach



Figure 3. Cumulative chromosome length for *An. darlingi* **assembly idAnoDarlMG_H_01.** 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/idAnoDarlMG_H_01/dataset/CALSDZ01/ cumulative.



Figure 4. Genome assembly of *An. darlingi*, idAnoDarlMG_H_01: Hi-C contact map. Visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at https://genome-note-higlass.tol.sanger.ac.uk/ l/?d=Pap61sCkTXaWekXO6tp-Wg.

Table 2. Chromosomal pseudomolecules in the genomeassembly of An. darlingi from French Guiana, idAnoDarlMG_H_01.

INSDC accession	Chromosome	Size (Mb)	Count	Gaps
OX030911.1	2RL	94.951	1	10
OX030912.1	3RL	71.271	1	5
OX030913.1	Х	13.401	1	15
OX030914.1	MT	0.015	1	0
	2 Unlocalised	0.030	1	0
	X Unlocalised	1.957	60	0
	Unplaced	0.026	1	0

Table 3. Genome data for Anopheles darlingi from Peru,idAnoDarlJC-H15_27.

Project accession data	
Assembly identifier	ASM3899477v1
Species	Anopheles darlingi
Specimen	idAnoDarlJC-H15_27
NCBI taxonomy ID	43151
BioProject	PRJNA830873
BioSample ID	SAMN33940061
Isolate information	female, whole organism
Raw data accessions	
PacificBiosciences SEQUEL II	SRX26935726
10X Genomics Illumina	SRX26935727, SRX26935728, SRX26935729, SRX26935730
Hi-C Illumina	SRX26803935
Genome assembly	
Assembly accession	GCA_038994775
Span (Mb)	182.384
Number of contigs	86
Contig N50 length (Mb)	13.497
Number of scaffolds	54
Scaffold N50 length (Mb)	95.091
Longest scaffold (Mb)	95.091
BUSCO* genome score	C:97.0%[S:96.3%,D:0.7%], F:0.7%,M:2.3%,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.

Table 4. Chromosomal pseudomolecules in the genome assembly of *An. darlingi* from Peru, idAnoDarlJC-H15_27.

INSDC accession	Chromosome	Size (Mb)	Count
CM077074.1	2RL	95.091	1
CM077075.1	3RL	71.801	1
CM077073.1	Х	13.508	1
CM077077.1	MT	0.015	1
	X Unlocalised	1.974	52

of using a sibling was taken 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 for assembly of specimen from French Guiana, generated by NCBI RefSeq12 and available through VectorBase13 under accession AdarGF1, RNA was extracted from separate whole sibling and unrelated insect specimens (idAnoDarlMG-G_01, idAnoDarlMG-G_02, and idAnoDarlMG-H_11) 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. A 10X Genomics Chromium read cloud sequencing library was also constructed according to the manufacturer's instructions (this product is no longer available). Only 0.5 ng of DNA was used and only 25-50% of the gel emulsion was put forward for library prep due to the small genome size. For Hi-C data generation, following the Arima HiC 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



Figure 5. Synteny between genome assemblies of *An. albimanus* AalbS3 and *An. darlingi*, idAnoDarlMG_H_01. Chromosome arms arrangements are identical, while within arms there are numerous rearrangements.



Figure 6. Synteny between genome assemblies of *An. darlingi* from French Guiana, idAnoDarlMG_H_01, and from Peru, idAnoDarlJC-H15_27. Coordinates of two inversions are provided in Table 5.

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).

Genome assembly

Assemblies were carried out with Hifiasm¹⁴; haplotypic duplications were identified and removed with purge_dups¹⁵. For

 Table 5. Inversions between genome assemblies of

 An. darlingi from French Guiana and Peru.

INSDC accession	Chromosome	Start	End
OX030911.1	2RL	6,114,428	12,943,365
OX030911.1	3RL	18,970,078	21,986,145

the specimen from Peru, one round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes¹⁶. The assembly was then scaffolded with Hi-C data¹⁷ using SALSA2¹⁸. The assembly was checked for contamination as described previously¹⁹. Manual curation was performed using gEVAL²⁰, HiGlass²¹ and Pretext²². The mitochondrial genome was assembled using MitoHiFi²³, which performs annotation using MitoFinder²⁴. The genome was analysed and BUSCO scores were generated within the BlobToolKit environment²⁵. Synteny analyses were performed with D-GENIES²⁶. Tandem repeats were annotated with TRASH¹⁰. Table 7 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.



Figure 7. Tandem repeats annotation for genome assembly of An. darlingi from French Guiana, idAnoDarlMG_H_01.



Figure 8. Tandem repeats annotation for genome assembly of An. darlingi from Peru, idAnoDarlJC-H15_27.

Table 6.	Chromosome	arms annotation	in genome	assemblies	of An.	darlingi.

Chromosome arm	idAnoDarlMG_H_01 start	idAnoDarlMG_H_01 end	idAnoDarlJC-H15_27 start	idAnoDarlJC-H15_27 end
2R	1	52,837,268	1	52,904,211
2L	53,277,590	94,951,917	53,898,552	95,090,923
3R	1	35,602,462	1	35,719,087
3L	36,556,652	71,270,736	36,907,605	71,801,356

Fable	7.	Software	tools	used.
--------------	----	----------	-------	-------

Software tool	Version	Source
hifiasm	0.14	14
purge_dups	1.2.3	15
longranger align	2.2.2	27
freebayes	1.3.1	16
SALSA2	2.2-4c80ac1	18
MitoHiFi	2	23
gEVAL	N/A	20
HiGlass	1.11.6	21

Software tool	Version	Source
PretextView	0.1.x	22
BlobToolKit	3.4.0	28
BUSCO	5.3.2	9
D-GENIES	1.4	26
TRASH	1.2	10

Data availability

European Nucleotide Archive: PRJEB53253, https://identifiers.org/bioproject/PRJEB53253; PRJNA830873, https://identifiers.org/bioproject/PRJNA830873. The genome sequences are released openly for reuse. The *Anopheles darlingi* genome sequencing initiative is part of the Anopheles Reference Genomes project PRJEB51690. All raw sequence data and the assemblies have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1 and Table 3.

Author information

Members of the Wellcome Sanger Institute Scientific Operations: Sequencing Operations collective are listed here: https://doi. org/10.5281/zenodo.4790456.

Acknowledgments

We acknowledge Romuald Carinci (Institut Pasteur de la Guyane) for assistance with field sample collection, and Nithya Swaminathan (Broad Institute) for support with project management and data submission. We are grateful to the Ministerio de Agricultura, Dirección General Forestal y de Fauna Silvestre and the Dirección de Salúd, Gobierno Regional de Loreto Gobierno Regional de Loreto under the auspices of Resolución Directoral Number 0425-2012-AG-DGFFS-DGEFFS for permission to carry out these studies of Anopheles darlingi in Peru.

References

- Komp WHW: The occurrence of Anopheles darlingi root in central America. Am J Trop Med Hyg. 1941; s1-21(5): 659–670. Publisher Full Text
- Rachou RG: Anofelinos do Brasil: comportamento das espécies vetoras de malária. Rev Bras Malariol Doencas Trop. 1958.
 Reference Source
- Harbach RE, Kitching IJ: The phylogeny of Anophelinae revisited: inferences about the origin and classification of Anopheles (Diptera: Culicidae). Zool Scr. 2016; 45(1): 34–47.
 Publisher Full Text
- Foster PG, de Oliveira TMP, Bergo ES, et al.: Phylogeny of Anophelinae using mitochondrial protein coding genes. R Soc Open Sci. 2017; 4(11): 170758.
 PubMed Abstract | Publisher Full Text | Free Full Text
- Marinotti O, Cerqueira GC, de Almeida LGP, et al.: The genome of Anopheles darlingi, the main neotropical malaria vector. Nucleic Acids Res. 2013; 41(15): 7387–7400.
 PubMed Abstract | Publisher Full Text | Free Full Text
- Moreno M, Tong C, Guzmán M, et al.: Infection of laboratory-colonized Anopheles darlingi mosquitoes by Plasmodium vivax. Am J Trop Med Hyg. 2014; 90(4): 612–616. PubMed Abstract | Publisher Full Text | Free Full Text
- Lainhart W, Bickersmith SA, Moreno M, et al.: Changes in genetic diversity from field to laboratory during colonization of Anopheles darlingi root (Diptera: Culicidae). Am J Trop Med Hyg. 2015; 93(5): 998–1001.
 PubMed Abstract | Publisher Full Text | Free Full Text
- Compton A, Liang J, Chen C, et al.: The beginning of the end: a chromosomal assembly of the new world malaria mosquito ends with a novel telomere. G3 (Bethesda). 2020; 10(10): 3811–3819.
 PubMed Abstract | Publisher Full Text | Free Full Text
- 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
- Wlodzimierz P, Hong M, Henderson IR: TRASH: Tandem Repeat Annotation and Structural Hierarchy. *Bioinformatics*. 2023; 39(5): btad308.
 PubMed Abstract | Publisher Full Text | Free Full Text
- 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 8 Jan 2024]. Publisher Full Text
- Pruitt KD, Brown GR, Hiatt SM, et al.: RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 2014; 42(Database issue): D756–63. PubMed Abstract | Publisher Full Text | Free Full Text
- 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
- 14. 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

- 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
- Garrison E, Marth G: Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN]. 2012. Reference Source
- 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
- 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

- Howe K, Chow W, Collins J, et al.: Significantly improving the quality of genome assemblies through curation. GigaScience. 2021; 10(1): giaa153. PubMed Abstract | Publisher Full Text | Free Full Text
- Chow W, Brugger K, Caccamo M, et al.: gEVAL a web-based browser for evaluating genome assemblies. Bioinformatics. 2016; 32(16): 2508–2510. PubMed Abstract | Publisher Full Text | Free Full Text
- 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
- 22. PretextView: OpenGL powered pretext contact map viewer. Github. Reference Source
- Uliano-Silva M, Nunes JGF, Krasheninnikova K, et al.: marcelauliano/MitoHiFi: mitohifi_v2.0. 2021.
 Publisher Full Text
- 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
- 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
- Cabanettes F, Klopp C: D-GENIES: dot plot large genomes in an interactive, efficient and simple way. Peerl. 2018; 6: e4958.
 PubMed Abstract | Publisher Full Text | Free Full Text
- Long Ranger BASIC and ALIGN Pipelines. Software -Genome & Exome -Official 10x Genomics Support. [cited 16 Dec 2022]. Reference Source
- 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

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 10 May 2025

https://doi.org/10.21956/wellcomeopenres.26465.r122270

© **2025 Raja K.** 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.



Komal Kumar Raja 匝

Baylor College of Medicine, Houston, USA

The manuscript by Gendrin et al presents two high quality genome assembly sequences for the species *Anopheles darlingi* from French Guiana and Peru. The authors have achieved 182 Mb of assemblies with N50 of about 95 Mb and BUSCO completeness of >97% suggesting good quality of the genome sequences. This assembly is a substantial improvement over the fragmented 2013 version of the genome. The authors have sufficiently introduced the biology and evolutionary context of the species including colony origin. The experimental and bioinformatics methodology is robust. The use of TRASH for centromeric and teloomeric repeat annotation is commendable. Overall, this manuscript represents a significant advance in *An. darlingi* genomics and will be a valuable resource for malaria vector research.

Minor comment

Please be consistent between 'idAnoDarlMG-H_01' and 'idAnoDarlMG_H_01'

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others? $\ensuremath{\mathsf{Yes}}$

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Single cell genomics, bioinformatics, genomics and Neurobiology

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 10 May 2025

https://doi.org/10.21956/wellcomeopenres.26465.r122268

© **2025 Ali R.** 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.



Renee Ali

Molecular Microbiology and Immunology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA

A good addition to the genomic data available for this important neotropical vector *Anopheles darlingi*. The methods were well explained and reproducible. It would be nice to produce another assembly or compare to the first version if possible to observe whether there where also inversions on the chromosome arm 2R.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others? $\ensuremath{\mathsf{Yes}}$

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetics, mosquito populations, understudied vectors

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 26 April 2025

https://doi.org/10.21956/wellcomeopenres.26465.r122272

© **2025 Wilding C.** 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.



Craig Wilding 匝

University College Dublin School of Biology and Environmental Science (Ringgold ID: 123708), Dublin, Leinster, Ireland

In this article, the authors describe the generation and annotation of chromosomal-scale genome sequences from two separate *Anopheles darlingi* individuals collected from geographically isolated (French Guiana and Peru) populations. Both genomes are similarly high quality (in terms of scaffold-length and BUSCO completeness) and were generated using a combination of PacBio HiFi and Hi-C scaffolding for genome sequencing, and Illumina for RNASeq (on a specimen from French Guiana).

Having two assemblies has permitted the authors to identify two major chromosomal inversions on chromosome 2R.

The methods utilized are reported clearly although syntenic comparison for the French Guianan sample was reported to be undertaken versus the *"An. albimanus* STECLA strain assembly (GCA_013758885.1)" whilst synteny for the Peruvian sample is reported as generated against *"An. albimanus* AalbS3". Whilst I think these are the same, for clarity, it would be better to be consistent in reference to the *An. albimanus* strain. Note, I believe that it should be the STELCA strain (not STECLA).

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others? $\ensuremath{\mathsf{Yes}}$

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary genetics/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.