

RESEARCH ARTICLE

Characterization of the complete mitogenome of *Anopheles aquasalis*, and phylogenetic divergences among *Anopheles* from diverse geographic zones

Luis Martinez-Villegas¹, Juliana Assis-Geraldo², Leonardo B. Koerich³, Travis C. Collier⁴, Yoosook Lee⁵, Bradley J. Main⁶, Nilton B. Rodrigues¹, Alessandra S. Orfano¹, Ana C. A. M. Pires¹, Thais B. Campolina¹, Rafael Nacif-Pimenta¹, Djane C. Baia-da-Silva^{7,8,9}, Ana P. M. Duarte^{7,8,9}, Ana C. Bahia¹⁰, Claudia M. Rios-Velásquez¹¹, Marcus V. G. Lacerda^{7,8,9,11}, Wuelton M. Monteiro^{7,8,9}, Gregory C. Lanzaro⁵, Nagila F. C. Secundino^{1,8,9}, Paulo F. P. Pimenta^{1,7,8,9*}



OPEN ACCESS

Citation: Martinez-Villegas L, Assis-Geraldo J, Koerich LB, Collier TC, Lee Y, Main BJ, et al. (2019) Characterization of the complete mitogenome of *Anopheles aquasalis*, and phylogenetic divergences among *Anopheles* from diverse geographic zones. PLoS ONE 14(9): e0219523. <https://doi.org/10.1371/journal.pone.0219523>

Editor: Igor V. Sharakhov, Virginia Polytechnic Institute and State University, UNITED STATES

Received: April 17, 2019

Accepted: June 25, 2019

Published: September 3, 2019

Copyright: © 2019 Martinez-Villegas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results of this study are available in the NCBI Database as part of the *Anopheles aquasalis* whole genome shotgun sequencing project; accession number: NJHH00000000.

Funding: This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley - USA supported by NIH S10 OD018174 Instrumentation Grant. The following Brazilian agencies partially funded this study: Fundação do

1 Laboratory of Medical Entomology, Institute René Rachou, Oswaldo Cruz Foundation, Minas Gerais, FIOCRUZ, Belo Horizonte, MG, Brazil, **2** Biosystems Informatics and Genomics Group, Institute René Rachou, Oswaldo Cruz Foundation, Minas Gerais, FIOCRUZ, Belo Horizonte, MG, Brazil, **3** Laboratory of Physiology of Haematophagous Insects, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil, **4** Daniel K. Inouye US Pacific Basin Agricultural Research Center (PBARC), United States Department of Agriculture, Agricultural Research Service, Hilo, Hawaii, United States of America, **5** Vector Genetics Laboratory, Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California-Davis, Davis, California, United States of America, **6** Davis Arbovirus Research and Training, School of Veterinary Medicine, University of California-Davis, Davis, California, United States of America, **7** Institute of Clinical Research Borborema, Tropical Medicine Foundation Dr. Heitor Vieira Dourado, Manaus, AM, Brazil, **8** Graduation Program in Tropical Medicine, Amazonas State University, Manaus, AM, Brazil, **9** Foundation of Tropical Medicine Dr. Heitor Vieira Dourado, Manaus, AM, Brazil, **10** Institute of Biophysics Carlos Chagas Filho, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, **11** Institute Leonidas and Maria Deane, Oswaldo Cruz Foundation, FIOCRUZ, Manaus, AM, Brazil

* pfppimenta@gmail.com, paulo.pimenta@fiocruz.br

Abstract

Whole mitogenome sequences (mtDNA) have been exploited for insect ecology studies, using them as molecular markers to reconstruct phylogenies, or to infer phylogeographic relationships and gene flow. Recent *Anopheles* phylogenomic studies have provided information regarding the time of deep lineage divergences within the genus. Here we report the complete 15,393 bp mtDNA sequences of *Anopheles aquasalis*, a Neotropical human malaria vector. When comparing its structure and base composition with other relevant and available anopheline mitogenomes, high similarity and conserved genomic features were observed. Furthermore, 22 mtDNA sequences comprising anopheline and Dipteran sibling species were analyzed to reconstruct phylogenies and estimate dates of divergence between taxa. Phylogenetic analysis using complete mtDNA sequences suggests that *A. aquasalis* diverged from the *Anopheles albitarsis* complex ~28 million years ago (MYA), and ~38 MYA from *Anopheles darlingi*. Bayesian analysis suggests that the most recent ancestor of *Nyssorhynchus* and *Anopheles* + *Cellia* was extant ~83 MYA, corroborating current estimates of ~79–100 MYA. Additional sampling and publication of African, Asian, and

Instituto Oswaldo Cruz (FIOCRUZ – PAPES VI); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Institutos Nacionais de Ciência e Tecnologia (INCTs Entomologia Molecular e Elimina Malária); Fundações de Amparo à Pesquisa dos Estados de Minas Gerais e Amazonas (FAPEMIG and FAPEAM). LMV was a Ph.D. student supported by the Programa de Estudantes - Convênio de Pós-Graduação (PEC-PG - CAPES). GCL was a fellow sponsored by Programme Science without Borders (CAPES); PFPP, NFCS, and WMM are research fellows supported by CNPq.

Competing interests: The authors have declared that no competing interests exist.

North American anopheline mitogenomes would improve the resolution of the *Anopheles* phylogeny and clarify early continental dispersal routes.

Introduction

The mitogenome of most insects is composed of a small double-stranded circular molecule of 14–20 kb in length. It contains 37 genes including 13 protein-coding genes (PCG), 22 transfer RNA genes (tRNA) and two ribosomal RNA genes (small (srRNA) and large (lr-RNA) ribosomal subunits). Additionally, it contains an A+T rich control region that is involved in the initiation of transcription and replication [1, 2]. The order of the genes within the mitogenome is highly conserved and can be traced to the ancestral gene arrangement from the Bilateria [3], which differs slightly from ancestral ecdysozoan and arthropod mitogenomes [4].

A total of 475 formally recognized species of *Anopheles* are currently known [5]. Until recently, knowledge regarding the evolution, divergence time, and phylogenetic relationships among representative species within this genus were scarce. Despite its medical importance worldwide, this gap has been due to the relevance and focus on African anophelines [6–9]. The widespread existence of cryptic species complicates taxonomic and phylogenetic analyses in the genus *Anopheles* [6, 7, 10, 11]. Nevertheless, the recent publication of 16 anopheline genomes [8] may provide, in the near future, new full mitogenomes as parallel assemblies of the genomic data produced. This would enable more accurate phylogenomic reconstructions and enhance estimates of the divergence times among members of this genus.

The current hypothesis regarding *Anopheles* evolution is mostly based on the geographic distribution of extant species [12]. It proposes that major mosquito lineages, including *Anopheles*, originated in Western Gondwana approximately 145 to 100 million years ago (MYA) during the periods recognized as late Jurassic or early Cretaceous [12, 13]. The genus *Anopheles* would have emerged in what now is South America and, following rapid diversification and migration across land bridges, colonized most of the Earth's favorable habitats [12, 14]. Recent publications hypothesize that the *Anopheles* subgenera *Nyssorhynchus* and *Anopheles* + *Cellia* diverged between 79–100 MYA [6, 15, 16], suggesting that their most recent common ancestor might have lived before the geological split of Western Gondwana ~95 MYA [16].

The human malaria cycle probably evolved in Africa, where first interactions between parasites, their anopheline hosts, and hominids occurred [7]. As malaria-infected humans migrated out of Africa, they carried *Plasmodium* parasites with them, leaving their African mosquito vectors behind. This migration gave rise to a journey of adaptation between the *Plasmodium* parasite and no less than 34 different anopheline mosquito species worldwide [17]. Based on available evidence of historical, archaeological and genetic type, it is believed that human malaria was introduced into the Americas by Europeans who transferred both, *Plasmodium falciparum* and *Plasmodium vivax* (the most prevalent malaria parasite species) to the indigenous population [18–20]. Although Neotropical anophelines were already adapted to feed upon primates, including humans, interactions between Neotropical malaria vectors, humans, and malaria parasites, can be considered geologically and evolutionarily recent [7, 16].

Anopheles aquasalis is a relevant Neotropical malaria vector of *P. vivax* on the Atlantic and Pacific coasts from Central America to Southern Brazil. *An. aquasalis* has been reared under laboratory conditions since 1995, being a well-established model for experimental studies involving the interaction of malaria vectors with several *Plasmodium* species [21]. Nonetheless, little is known about its evolutionary relationship with other *Anopheles* sibling species.

Mitochondrial genome sequence data are useful to infer phylogenetic and phylogeographic relationships [22–26]. Here we present a preliminary characterization of the mitogenome of *A. aquasalis* assembled by Next Generation shotgun sequencing (NGS). We compared the mtDNA genome sequence, and some of its features, with other selected anopheline mitogenomes. We applied Bayesian analysis to reconstruct phylogenetic relationships and estimate divergence times amongst other human malaria vectors. The implications of these findings were briefly discussed regarding the evolutionary history of anophelines in general. A more thorough and balanced analysis of anopheline mitogenomes, including representatives from North America, Asia and Africa would provide an in-depth description of dispersal routes throughout evolutionary and geological times.

Methods

Anopheles aquasalis colony

A. aquasalis were obtained from a colony established at the Medical Entomology Laboratory at Instituto de Pesquisas René Rachou-FIOCRUZ (Fiocruz, Minas Gerais). The mosquitoes originally came from a colony established in 1995 in Rio de Janeiro [27, 28], and are currently kept under laboratory conditions as previously described [29].

Single mosquito DNA extraction

Genomic DNA from a single adult female *A. aquasalis* was extracted using the QiagenDNeasy blood and tissue kit (Qiagen, Hilden, Germany) according to the protocol for purification of insect DNA with a minor modification: Qiagen EB buffer rather than AE buffer was used to avoid possible interference of EDTA with the Nextera kit enzymes. The purified *Anopheles* genomic DNA was then quantified with the Qubit HS (Life Technologies, USA) system and used to construct the genomic library.

Whole genome shotgun sequencing

The genomic DNA was processed using the Nextera DNA sample preparation kit (Epicentre Biotechnologies, Madison, WI). Thirty nanograms of sample DNA were fragmented utilizing 5 μ l of Tagment DNA Enzyme with 25 μ l of Tagment Buffer. Tagmentation reactions included in the Nextera kit were performed by incubating the sample for 5 min at 55°C. The tagmented DNA was purified using the QiagenMinElute protocol (QIAGEN, Germany). Purified DNA was eluted from the column with 11 μ l of nuclease-free water. Purified DNA (5 μ l) was used as the template in a 20- μ l volume for limited-cycle PCR (5 cycles) and processed as outlined in the Nextera protocol (Illumina). Amplified DNA was purified using the AMPure Bead cleanup (Beckman Coulter, USA) according to the manufacturer's protocol. The fragment size distribution of the tagmented DNA was analyzed utilizing a 2100 Bioanalyzer with a 7500 DNA assay kit (Agilent Technologies, Santa Clara, CA). Fragments of ~600 bp long were carried out for sequencing. The library was sequenced on one lane of an Illumina HiSeq2000 instrument to generate paired-end reads. Sequencing was performed by The Vincent J. Coates Genomics Sequencing Laboratory (GSL) at the University of California, Berkeley.

Mitochondrial genome assembly

Sequences were assembled *de novo* using Velvet v1.2.10 [30] with a k-mer size of 41, according to the scripts and parameters suggested by the Velvet Manual (<https://www.ebi.ac.uk/~zerbino/velvet/Manual.pdf>) and in-house protocols from the UC Davis Vector Genetics Laboratory. The assembled contigs were aligned to the mtDNA sequence of *An. gambiae* (GenBank

No. L20934.1) and *A. darlingi* (GenBank No. GQ918273.1) using the MUMmerv3.0 software [31] to identify and confirm the *de novo* assembly of the *A. aquasalis* mitogenome.

Sequence analysis: Composition and genomic features of the *A. aquasalis* mitogenome

The assembled mitogenome was manually inspected for repeats at the beginning and end of the assembly to infer circularity. Automatic annotation of the mitogenome fasta file was performed with MITOS [32], followed by a manual curation based on the GenBank file format. Manual inspection of the predicted Protein-Coding Genes (PCG's), ribosomal RNA (rRNA) genes, transfer RNA (tRNA) genes, and the AT rich region was performed with Artemis -release 16- [33]. The nucleotide sequences of PCGs were translated based on the invertebrate mtDNA genetic code.

The manual curation of coding regions and rRNA genes was mainly carried by sequence comparison with published insect mitogenome sequences such as *A. gambiae* (GenBank No. L20934.1) and *A. darlingi* (GenBank No. GQ918273.1) amongst others. Careful attention was given to PCGs comparing the predicted Open reading frames (ORFs) to the UniProt database (<http://www.uniprot.org/>) giving more weight to similarity hits that had experimental validation.

The MITOS annotated tRNAs were verified and their secondary structures predicted with the tRNAscan-SE search server v1.21 [34] with default settings: the invertebrate mitochondrial codon predictors, and a cove score cut off of 5 (software tool available at: <http://lowelab.ucsc.edu/tRNAscan-SE/>). Some tRNA genes could not be detected by tRNAscan-SE. They were identified by direct comparison and sequence similarity to tRNAs of other dipterans or anophelines. These tRNA genes were modeled with RNAstructure [35]. To maintain a uniform format, all of the 22 figures were generated with RNAstructure.

To visualize the annotated mitogenome, circular representation of it was generated with Blast Ring Image Generator (BRIG) [36]. For such purpose, a GenBank formatted file and the fasta file were employed according to the Brig 0.95 Manual. Available at: <http://ufpr.br/sourceforge.net/project/brig/BRIGMANUAL.pdf>.

Nucleotide composition analyses, expressed as AT%, were performed for individual PCGs, full mtDNA, concatenated PCGs, concatenated tRNAs, lrRNA (16S), srRNA (12S) and concatenated rRNAs. For the aforementioned targets, composition bias based on strand asymmetry values were estimated using the following formulae for skews: AT skew = $[A-T]/[A+T]$ and GC skew = $[G-C]/[G+C]$ as proposed by Perna & Kocher (1995) [37] on an Excel spreadsheet. Codon bias was assessed estimating the relative synonymous codon usage (RSCU). All the above compositional analyses (except bias estimation) were performed using MEGA v6.0 [38]. The phylogenetic and comparative analyses performed henceforth relied upon available sequences and published literature regarding anopheline and culicine species detailed in Table 1.

Comparative analyses between anophelines from different geographic regions

Comparative analyses regarding nucleotide composition and strand asymmetry were performed between *A. aquasalis* and four other anophelines representative of different geographic regions: *Anopheles punctulatus* (GenBank No. JX219738.1) from South East Asia, *A. gambiae* (GenBank No. L20934.1) from Africa, *A. darlingi* North (GenBank No. GQ918272.1) from Central America, and *A. darlingi* South (GenBank No. GQ918273.1) from South America. For

Table 1. List of the insect species used in this study with their corresponding GenBank number.

Species	Family	Length (bp)	GenBank No.	Vector of malaria in (continent)	Reference
<i>Anopheles aquasalis</i>	Culicidae	15393	NJHH00000000	South America	This study
<i>Anopheles punctulatus</i> (isolate ITN_PNG-18)	Culicidae	15412	JX219738.1	Oceania	[6]
<i>Anopheles farauti</i> 4 (isolate 7_10–11)	Culicidae	15412	JX219735.1	Oceania	[6]
<i>Anopheles farauti</i> 4 (isolate 8_11–12)	Culicidae	15412	JX219736.1	Oceania	[6]
<i>Anopheles hinesorum</i>	Culicidae	15336	JX219734.1	Oceania	[6]
<i>Anopheles koliensis</i> (isolate ESP001B)	Culicidae	15412	JX219743.1	Oceania	[6]
<i>Anopheles dirus</i> A (isolate A1)	Culicidae	15404	JX219731.1	Southeast Asia	[6]
<i>Anopheles dirus</i> A (isolate A2)	Culicidae	15404	JX219732.1	Southeast Asia	[6]
<i>Anopheles cracens</i> (isolate B1)	Culicidae	15412	JX219733.1	Southeast Asia	[6]
<i>Anopheles albitarsis</i> F	Culicidae	15418	HQ335349.1	South America	[39]
<i>Anopheles albitarsis</i> G	Culicidae	15474	HQ335346.1	South America	[39]
<i>Anopheles deaneorum</i>	Culicidae	15424	HQ335347.1	South America	[39]
<i>Anopheles janconnae</i>	Culicidae	15425	HQ335348.1	South America	[39]
<i>Anopheles oryzalimnetes</i>	Culicidae	15422	HQ335345.1	South America	[39]
<i>Anopheles darlingi</i> North	Culicidae	15386	GQ918272.1	Central America	[40]
<i>Anopheles darlingi</i> South	Culicidae	15385	GQ918273.1	South America	[40]
<i>Anopheles quadrimaculatus</i> (A strain Orlando)	Culicidae	15455	L04272.1	North America	[41]
<i>Anopheles gambiae</i>	Culicidae	15363	L20934.1	Africa	[42]
<i>Culex pipiens</i>	Culicidae	14856	HQ724614.1	NA	[43]
<i>Aedes aegypti</i>	Culicidae	16655	EU352212.1	NA	Unpublished
<i>Aedes albopictus</i>	Culicidae	16665	NC_006817.1	NA	Unpublished
<i>Drosophila melanogaster</i>	Drosophilidae	19517	U37541.1	NA	[44]

For anopheline species, the continent in which they exert their malaria vectorial activity is shown. The sequence length reflects the number of base pairs assembled not considering Ns.

<https://doi.org/10.1371/journal.pone.0219523.t001>

each mtDNA genome, base composition (expressed as AT%) and strand asymmetry (AT and GC skew) were calculated as explained above.

Additionally, a CDS nucleotide similarity comparison between *A. aquasalis*, the above-cited anophelines, plus *Anopheles albitarsis* (GenBank No. HQ335349.1)—another South American brackish-water tolerant species—was performed using the Blast2sequence online tool [45] available at <http://blast.ncbi.nlm.nih.gov/Blast.cgi#>.

Phylogenetic analysis and molecular dating

To further our knowledge on the *Anopheles* genus phylogeny, as well as to estimate the divergence time or split between *A. aquasalis* and other Neotropical anophelines, we reproduced the phylogeny and molecular dating analyses performed by Logue *et al.*, (2013) [6] using the tools and parameters suggested by the authors adding the assembled *A. aquasalis* mitogenome.

Briefly, 21 insect mitogenomes were selected from the ones used by Logue *et al.*, (2013) [6] and their sequences retrieved from the NCBI databank (Table 1). Then, for each of the 13 PCGs, the following actions were performed: the DNA sequences were translated into amino acid sequences using the invertebrate mt genetic code, then they were aligned to each other with the MAFFT alignment engine, and the aligned amino acid sequences were reverse-translated back into nucleotide sequences. All the above steps were performed with the online tool Translator X [46] using default parameters (available at: <http://translatorx.co.uk/>). Afterward, the aligned sequences from all 13 mt genes were concatenated using FASconCAT [47]. The

concatenated PCG sequences from the 22 mitogenomes were analyzed with jModeltest v0.1.1 [48] to determine the best nucleotide substitution model according to the Akaike Information Criterion.

Bayesian phylogenies were reconstructed, and node ages inferred, using BEAST v1.7.5 [49]. The following program parameters were used: an uncorrelated lognormal relaxed clock model allowing for rate heterogeneity among species; the GTR + G substitution model; the SRD06 model of partitioning, which enables estimation of nucleotide substitution parameters separately for the 1st + 2nd and 3rd codon positions (this apparently provides a better fit for protein-coding nucleotide data), and a Yule model for tree reconstruction. With the above parameters, three independent runs of 20 million generations were performed, saving the generated trees every 1,000 generations. All runs were then combined after a burn-in of 10% using LogCombiner v1.7.2 and afterwards, Tracer v1.5 was used to verify the mixing of the Markov chains (both tools are part of the BEAST pipeline). The maximum credibility tree was determined using TreeAnnotator v1.7.2 and visualized with FigTree v1.4.3 available at <http://tree.bio.ed.ac.uk/software/figtree/>. Divergence times were estimated using BEAST v1.7.5 following the instructions provided by the developers at <http://beast.bio.ed.ac.uk/>. In addition to the aforementioned parameters, the *Drosophila-Anopheles* divergence time was set as the calibration point using a prior distribution normally distributed around a mean of 260 million years ago (MYA) ranging from 243 to 276 MYA as suggested by Gaunt & Miles (2002) [50].

Results

Composition and genomic features

The complete mitogenome of *A. aquasalis* was assembled into a single contig of 15,393bp. The expected 37 genes in animal mtDNA, comprising 13 protein-coding genes, two rRNA genes (12S and 16S), 22 tRNA genes and a control region were identified (shown in Figs 1 and 2). *Aaquisalis_mt.gb*, shows the annotated features in GenBank format. [NCBI as part of the *Anopheles aquasalis* whole genome shotgun sequencing project; accession number NJHH00000000]. A short region of 229 bp, located within the A-T rich region, failed to be assembled due to low coverage. That is consistent with the report showing lower coverage of the AT-rich region using Illumina Nextera kit than some other enzymatic sheering protocols [51].

The annotated genes are encoded on both the heavy (22 genes) and light (15 genes) strands with some ORFs overlapping adjacent genes. In total, there are 38 overlapping nucleotides between 14 neighboring genes with the junctions spanning from 1 to 7 bp. Excluding the control region, we found 44 intergenic nucleotides (IGNs) at 7 locations with their lengths ranging from 1 to 17 bp. As in other dipterans and metazoans [2, 25, 52], the most common start codon was ATG (6 PCGs). Incomplete or truncated termination codons were annotated in the following PCGs: *COI*, *COII*, *COIII*, *ND5*, and *ND4* (Fig 1). Other Anopheline mosquitoes also shared the same incomplete termination codons.

The mtDNA of *A. aquasalis*, as in other anophelines or insects [15, 53] includes 22 tRNA genes with anticodons representing 20 different amino acids, with a length ranging from 64–72 bp, and a total length of 1477 bp when concatenated. The lengths of the 12S and 16SrRNA genes are 793 and 1324 bp respectively, both being encoded on the light strand (L). As suggested for metazoans, the ends of both rRNA genes were assumed to stretch up to the boundaries of flanking genes [54]. Like reported for other dipterans species, the 16SrRNA gene is flanked by the *tRNA^{Leu}* and *tRNA^{Val}* genes while the 12SrRNA gene is placed between *tRNA^{Val}* and the control region. The A-T content for both was 82.5% and 79.9% respectively, resembling the composition of other dipterans as compared by Zhao *et al.*, (2013). [25]. The 22

Gene	Strand	Location	Length (bp)	IGN	Codon		AT %
					Start / Anti	Stop	
<i>tRNA^{Ile}</i>	H	1-68	68	-3	GAT		
<i>tRNA^{Gln}</i>	L	66-134	69	-1	TTG		
<i>tRNA^{Met}</i>	H	134-202	69	0	CAT		
<i>ND2</i>	H	203-1228	1026	-2	ATT	TAA	81.97
<i>tRNA^{Trp}</i>	H	1127-1295	69	-1	TCA		
<i>tRNA^{Cys}</i>	L	1295-1358	64	0	GCA		
<i>tRNA^{Tyr}</i>	L	1359-1423	65	-2	GTA		
<i>COI</i>	H	1422-2958	1537	0	TCG	T ³	69.14
<i>tRNA^{Leu}</i>	H	2959-3024	66	2	TAA		
<i>COII</i>	H	3027-3711	685	0	ATG	T ³	73.98
<i>tRNA^{Lys}</i>	H	3712-3782	71	11	CTT		
<i>tRNA^{Asp}</i>	H	3794-3861	68	0	GTC		
<i>ATP8</i>	H	3862-4023	162	-7	ATT	TAA	82.10
<i>ATPase6</i>	H	4017-4697	681	-1	ATG	TAA	74.01
<i>COIII</i>	H	4697-5483	787	0	ATG	T ³	70.74
<i>tRNA^{Gly}</i>	H	5484-5550	67	0	TCC		
<i>ND3</i>	H	5551-5904	354	-2	ATA	TAA	76.55
<i>tRNA^{Arg}</i>	H	5903-5966	64	0	TCG		
<i>tRNA^{Ala}</i>	H	5967-6032	66	0	TGC		
<i>tRNA^{Asn}</i>	H	6033-6100	68	-1	GTT		
<i>tRNA^{Ser}</i>	L	6100-6166	67	1	GCT		
<i>tRNA^{Glu}</i>	H	6168-6233	66	-2	TTC		
<i>tRNA^{Phe}</i>	L	6232-6298	67	0	GAA		
<i>ND5</i>	L	6299-8039	1741	0	GTG	T ³	78.22
<i>tRNA^{His}</i>	L	8040-8104	65	0	GTG		
<i>ND4</i>	L	8105-9446	1342	-7	ATG	T ³	76.81
<i>ND4L</i>	L	9440-9739	300	10	ATG	TAA	83.00
<i>tRNA^{Thr}</i>	H	9750-9817	68	0	TGT		
<i>tRNA^{Pro}</i>	L	9818-9883	66	2	TGG		
<i>ND6</i>	H	9886-10410	525	-1	ATT	TAA	84.76
<i>CYTB</i>	H	10410-11546	1137	-2	ATG	TAA	72.56
<i>tRNA^{Ser}</i>	H	11545-11610	66	17	TGA		
<i>ND1</i>	L	11628-12584	957	-6	ATA	TAA	76.70
<i>tRNA^{Leu}</i>	L	12579-12644	66	0	TAG		
<i>16S rRNA</i>	L	12645-13968	1324	2			
<i>tRNA^{Val}</i>	L	13971-14042	72	0	TAC		
<i>12S rRNA</i>	L	14043-14835	793	0			
Control region		14836-15393	558	0			

³ Termination codons completed via polyadenylation
IGN: Intergenic nucleotide; minus indicates overlapping between genes.

Fig 1. Organization and gene features of the *A. aquasalis* mitochondrial genome.

<https://doi.org/10.1371/journal.pone.0219523.g001>

predicted secondary structures of individual tRNAs are shown in S1 Fig. All folded into the classic cloverleaf secondary structure except *tRNA-Ser* (GCT) in which the dihydrouridine stem “simply” formed a loop.

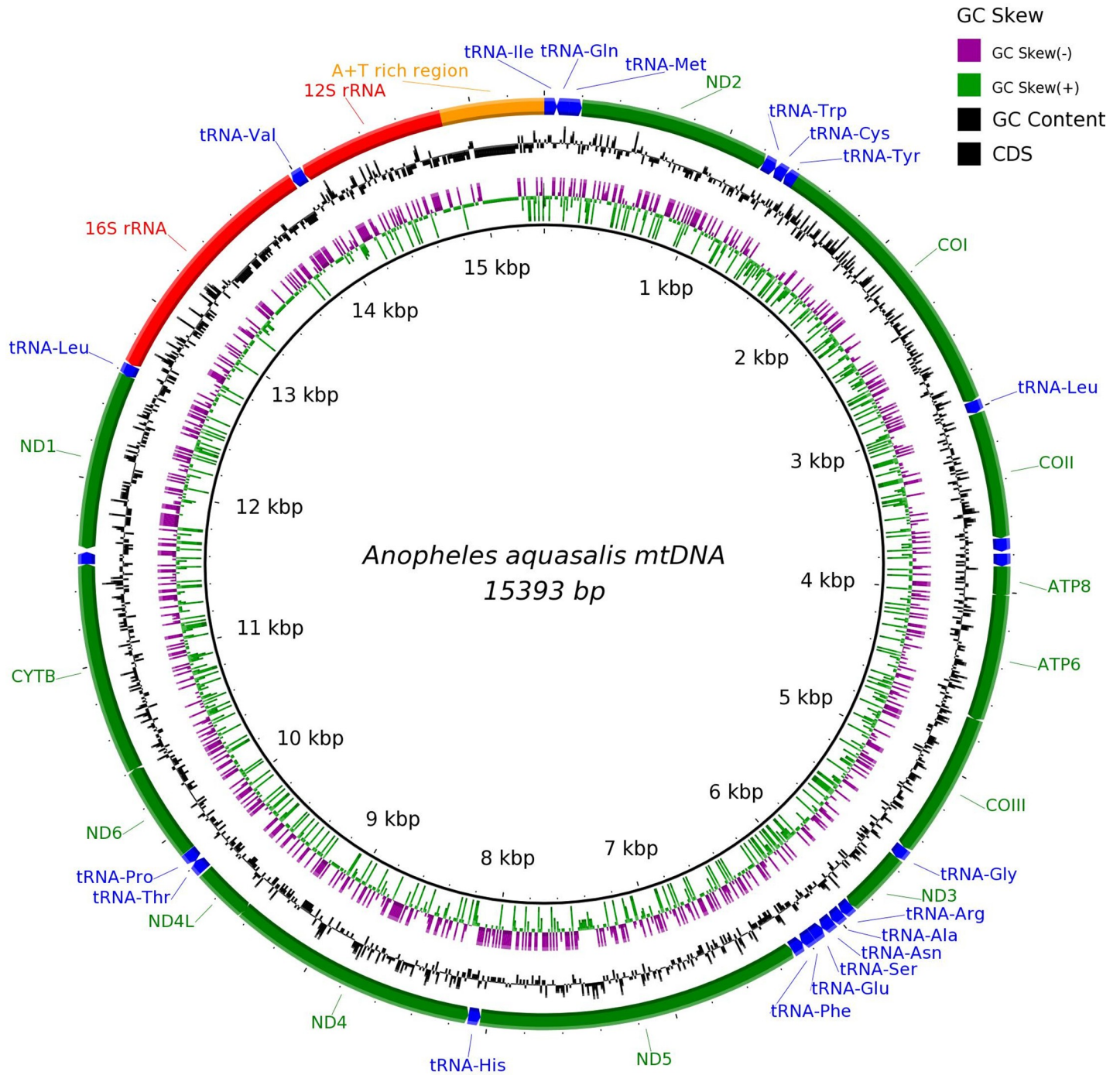


Fig 2. The complete mitogenome of *A. aquasalis*. BRIG visualization showing the protein coding genes, rRNAs and tRNAs in the mtDNA genome of *A. aquasalis*. The black inner ring shows the GC content on the outer surface, whereas AT content is shown on the inner surface. Strand asymmetry of the mitogenome is shown by the GC (+) and (-) skews according to the color key shown in the legend.

<https://doi.org/10.1371/journal.pone.0219523.g002>

The nucleotide composition of the *A. aquasalis* mtDNA is biased towards a high A+T content. The overall AT% content in the whole mitochondrion sequence was 77.16%, and concordantly it had a 22.84% GC. The AT% for individual PCG's, long and short rRNA genes, and concatenated rRNAs and tRNAs are summarized in [S1 Table](#).

Base composition, measured by strand asymmetry analyses (AT and GC skews), were also computed and are shown in [S1 Table](#). Regarding strand compositional biases, whole anopheline mitogenomes have AT- and GC-skews that are similar to the reported for other metazoans [16, 55]. The complete *A. aquasalis* mitogenome has a positive AT skew and a negative GC skew for the majority strand (also known as light -L-), which means that this strand is richer in A and C nucleotides.

The *A. aquasalis* mitogenome contains 3743 codons whose usage is shown in [S2 Table](#). Almost all the codons were present except AGG (S). There are 29 codons (out of 32) of the NNU and NNA types with a relative codon usage bias (RSCU) greater than 1. A strong bias towards A+T rich codons was observed, being the five most prevalent codons (in descending order): UUA (L), UUU (F), AUU (I), AUA (M), and AAU (N) as shown in [S3 Table](#). The most used codon is UUA (L) and the less used codon in the genome is CGC (R). The mitogenome is rich in leucine while cysteine is the rarest amino acid.

Comparative composition and identity analyses between selected anophelines

Compositional patterns based upon %AT and strand asymmetries were sought by comparing the mtDNA sequences from *A. aquasalis* and other human malaria vectors selected for their representative geographic distribution. The computed data and comparative approach rendered the plots shown in [S2 Fig](#). Overall, the plots represent the similar trends of AT% and skew patterns between the compared features, amongst the evaluated anophelines. Matching profiles between invertebrates (mollusks) had been reported before by Plazzi *et al.*, (2013) [52] via these types of plots. The observed trend had a few exceptions like the estimated GC skews from *COI* (from *A. darlingi* North and *A. darlingi* South) and *COII* (from *A. darlingi* South).

When comparing *A. aquasalis* coding nucleotide sequences and rRNA genes with their corresponding counterparts from the selected anopheline species, higher identity values were obtained with the American anophelines than with those from Africa and Asia. This was the general trend across all the compared PCGs and rRNA genes ([S4 Table](#) and [S3 Fig](#)).

Phylogeny and molecular dating analyses

The PCG sequences of 18 *Anopheles* mtDNA genomes were analyzed including the following outgroups: *D. melanogaster*, *C. pipiens*, *Ae. albopictus*, and *Ae. aegypti* (GenBank numbers are shown in [Table 1](#)). The aligned and concatenated sequences from the 22 mitogenomes resulted in a block of 11,514 nucleotides. According to the Akaike Information Criterion, the best nucleotide substitution model for this data set was the General Time Reversible with gamma distribution (GTR + G) model.

A phylogenetic tree was reconstructed using Bayesian analysis with BEAST v1.7.5 [49]. All the phylogenetic relationships were supported with robust posterior probabilities greater than 90%, with the exception of the position of *A. gambiae* (42%) and the internal nodes amongst the South East Asia and Oceania anophelines included (ranging from 85 to 88%). The reconstructed *Anopheles* phylogenetic tree is shown in [Fig 3](#). From examining the tree topology, and considering the current continental distribution, a deep divergence between two *Anopheles* geographical lineages was observed. One clear monophyletic branch grouping Central and South American anophelines, and a second lineage grouping North American, Asian/Oceania, and African anophelines. As expected, this radiation pattern resembles the one published by Logue *et al.*, (2013) [6]. The monophyletic clades corresponding to the *A. punctulatus*, *A. dirus* and *A. albitarsis* groups are also evident.

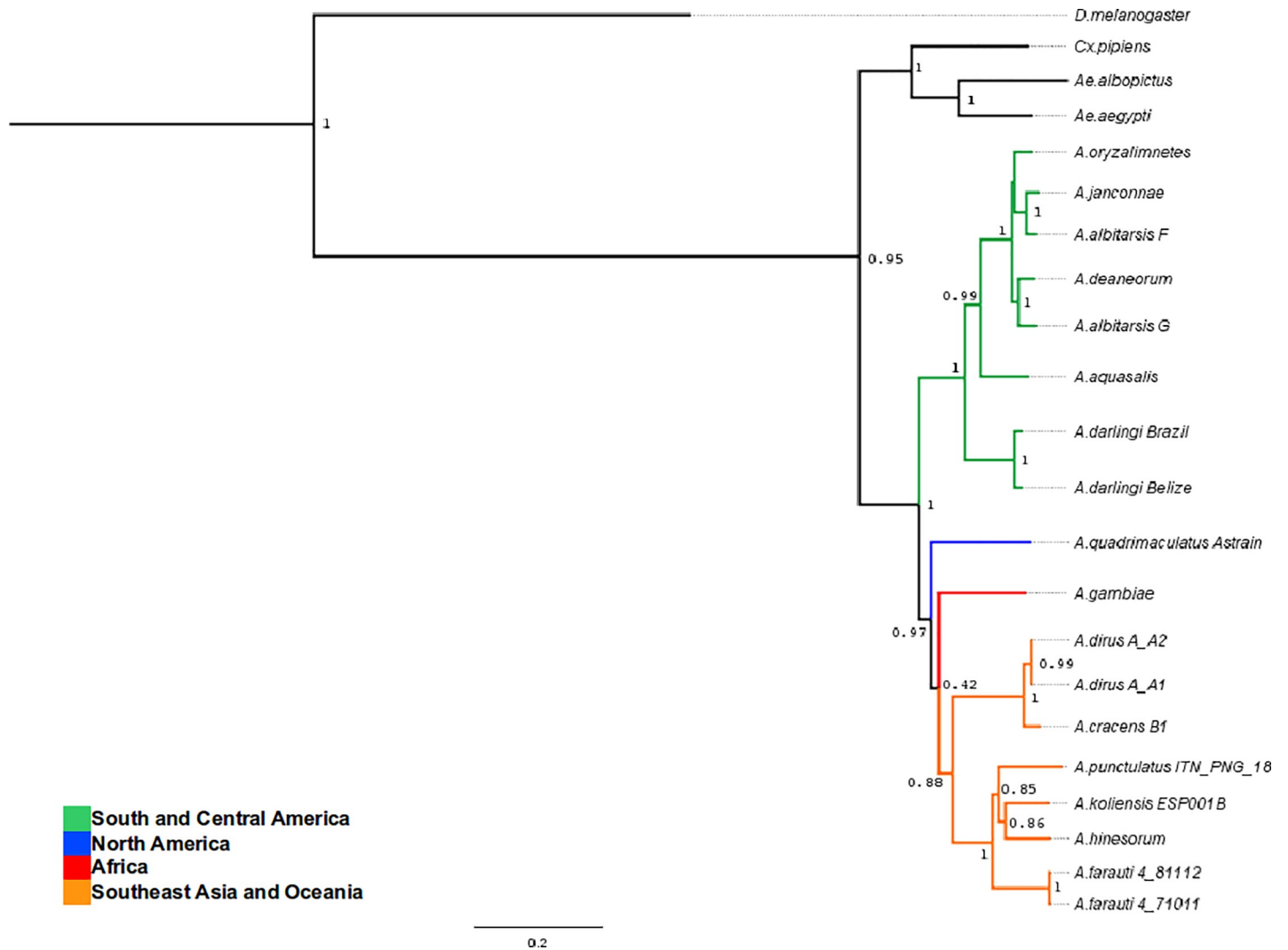


Fig 3. By-geographic region phylogeny of representative *Anopheles*, using the concatenated DNA sequences of all the mitochondrial protein coding genes. The values on the nodes correspond to the posterior probabilities supporting the tree topology. The phylogenetic tree was reconstructed using the concatenated PCGs and the Bayesian Markov Chain Monte Carlo approach (MCMC) analysis implemented in BEAST v1.7.5.

<https://doi.org/10.1371/journal.pone.0219523.g003>

Table 2. Mean divergence times in MYA (million of years ago), and 95% credibility intervals for selected nodes.

<i>Drosophila</i> / <i>Anopheles</i> (Calibration point ~260mya)	258.85	[239.70–278.86]
<i>Anophelinae</i> / <i>Culicinae</i>	136.63	[90.13–189.15]
<i>Anopheles</i> genus	83.23	[54.33–115.88]
<i>Anopheles gambiae</i> / SE Asia—Oceania anophelines	65.50	[43.54–89.80]
<i>Anopheles darlingi</i> / <i>Anopheles albitarsis</i> complex	38.98	[23.54–58.32]
<i>Anopheles aquasalis</i> / <i>Anopheles albitarsis</i> complex	28.56	[17.10–42.12]

Output of the analysis performed with BEAST v1.7.5 as explained in the Methods section. The measures in the second and third column correspond to the mean ages, and 95% credibility intervals determined for each of the selected nodes respectively.

<https://doi.org/10.1371/journal.pone.0219523.t002>

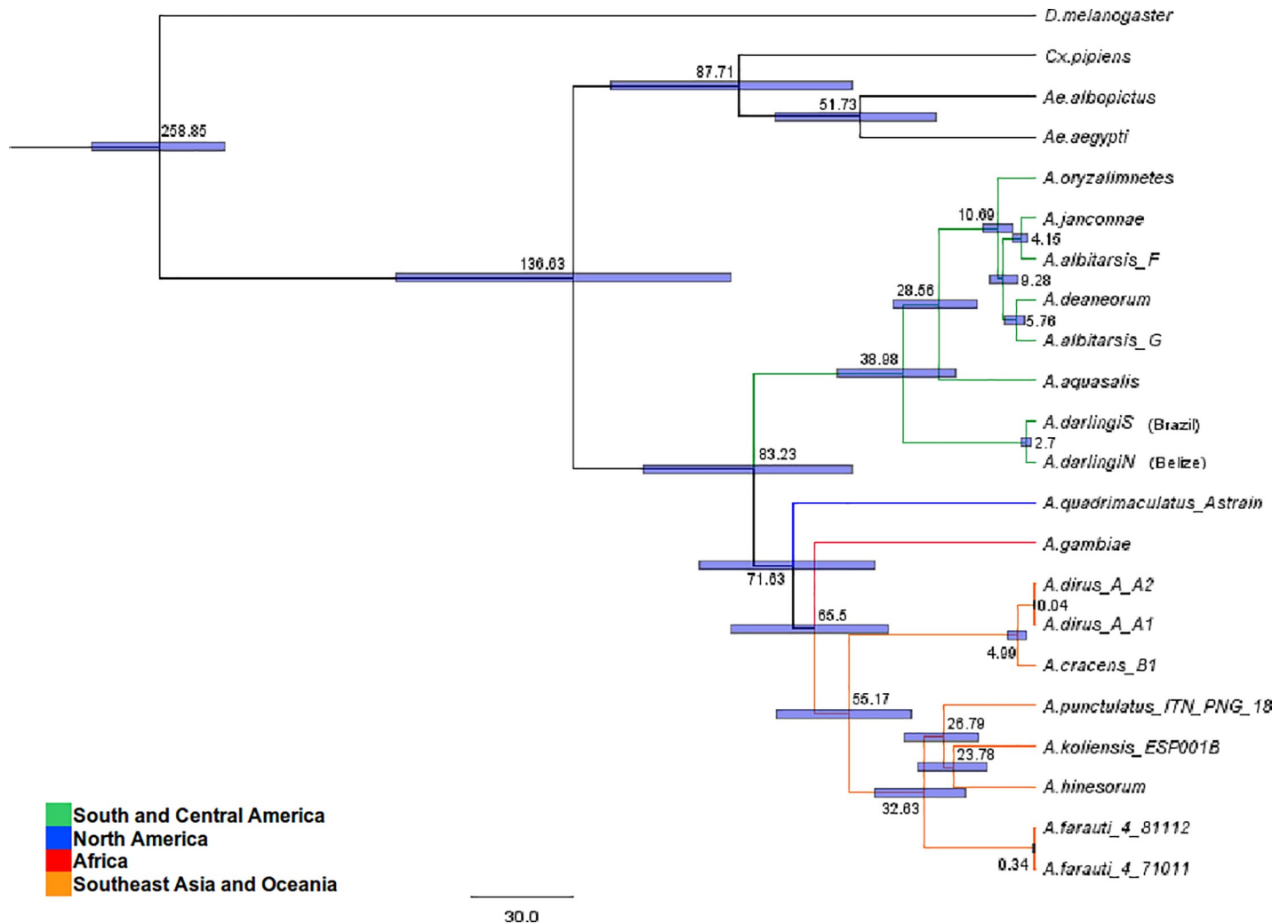


Fig 4. Phylogenetic tree of selected *Anopheles* using the concatenated DNA sequences of all the mitochondrial protein coding genes. The values on the tree nodes correspond to the mean divergence time (MYA) estimated for each event. The bars illustrate the 95% credibility intervals of the divergence times. (Table 2 presents the selected key divergence events).

<https://doi.org/10.1371/journal.pone.0219523.g004>

The currently available fossil record for mosquitoes is poor [7, 56]. Thus, we estimated the divergence times amongst anopheline species using the *Drosophila-Anopheles* divergence time (set at 260 MYA) as the only calibration point time as previously suggested [6, 50]. We dated the most recent common ancestor (MRCA) of all *Anopheles* to 83.23 MYA with a 95% credibility interval ranging from 54.33 to 115.88 MYA as shown in Table 2. Amongst Neotropical anophelines, the MRCA within the *A. albopictus* complex and *A. aquasalis* dates to 28.56 MYA with a 95% credibility interval ranging from 17.10 to 42.12 MYA. This MRCA is younger than the one shared between *A. darlingi* and the *A. albopictus* complex, which dates back to 38.98 MYA as it can be observed in Table 2 and Fig 4.

Discussion

Greater integration of nuclear and mitochondrial genomic studies is necessary to further our understanding of the *Anopheles* genomic evolution, the phylogenetic relationships between vectorial competence traits, and the co-evolutionary history of this genus with the human

malaria parasites they may harbor. Thus, in the current “omics” era, extending the wealth of representative anopheline mitogenomes available is a necessary and feasible task.

Like in most metazoans, the assembled mitochondrial genome of *A. aquasalis* is a small, double-stranded, circular DNA molecule. Encompassed in 15,393 base pairs, we annotated a densely packaged set of 13 protein-coding genes, 22 tRNA genes, and two rRNA genes as shown in Fig 1. Genome structures with highly conserved features, composition, and organization have been reported among insects [51] such as anophelines [15].

When regarding genome architecture, the occurrence of overlapping open reading frames implies a genome with a compact structure. This feature can be observed by mtDNA sequences of culicines, anophelines, and other insect mitogenomes [47, 52, 53]. The reason for such characteristic can be (structurally) attributed to the small intergenic distances between consecutive genes (intergenic distance < x bp) as the following show: < 23 bp in *Ae. aegypti*; < 24 bp in *C. quinquefasciatus*; < 21 bp in *Ae. albopictus*; < 16 bp in *A. gambiae*; < 18 bp in *Anopheles quadrimaculatus*; < 30 bp in *Drosophila yakuba*, and < 30 bp in *D. melanogaster* (as recounted by [53]), and in the case of *A. aquasalis* < 17 bp. Once a free-living prokaryote, mitochondria underwent genome reduction as it transitioned into an obligate endosymbiont harboring tightly packed remnants from the eubacterial genome of its ancestor [54].

The intergenic distances do not account for the control region, which in the *A. aquasalis* mitogenome spans 558bp. This major non-coding region is also known as the A+T-rich region, and it plays a role in the initiation of transcription and replication [55]. The length of this region is highly variable among different insects due to the high rate of nucleotide substitutions, insertions/deletions, and a variable number of tandem repeats it can have [56, 57].

Almost all the tRNA sequences annotated could be folded into typical cloverleaf secondary structures (S1 Fig) exposing the adequate anticodon triplet, except for the DHU-arm of tRNA^{Ser}, which is absent as it has been observed in other insects Li *et al.*, 2013; Zhao *et al.*, 2013 [25, 47] and references therein. Deficiencies regarding secondary structures of tRNA are often observed in protozoans, fungi, algae, plants and low metazoans [58]. The causes of such deficiencies range from aberrant loops and non-Watson-Crick matches to short arms, and all may induce aberrant tRNAs to lose their function, but a “corrective” post-transcriptional RNA editing mechanism has been proposed [59, 60]. The tRNA genes are embedded in variable regions within the mitogenome [51]. Throughout evolution, these regions underwent rearrangements more often than protein-coding regions. Consequently, tRNA order is nowadays explored as an additional tool for enhanced comparative phylogenetic analysis as between species [61].

The nucleotide composition of the complete *A. aquasalis* mtDNA sequence is clearly biased towards a high A+T content (77.16%) as it can be seen in S1 Table. This result was similar to the base composition described for other anophelines. The majority strand (L) of the *A. aquasalis* mtDNA has a compositional bias (positive AT skew and a negative GC skew). This means that this strand is richer in A and C. Though this is the tendency in insects, exceptions exist in arthropod mitogenomes in which the trend is reversed towards a composition with less A & C than T & G on the majority strand [62]. The underlying mechanism that leads to the strand bias, though unknown, has been linked to asymmetric replication and transcription processes. During both, one strand is transiently in a single-stranded state, and thus, it is left more prone to DNA damage. This phenomenon has been widely considered to account for the biased occurrence of mutations between the mtDNA strands [63].

Most of the start codons of the annotated PCGs followed the ATN rule, as described previously for other anophelines [15, 64]. Nonetheless, this was not the case for *COI* and *ND5*; genes that have TCG and GTG start respectively (Fig 1). The latter start codon characteristic has also been reported in other anopheline mitogenomes [15, 64, 65]. As observed in *A. darling* [15], eight genes use the complete stop codon TAA except for *COI*, *COII*, *COIII*, *ND5*, and

ND4, which terminate with a single T (Fig 1). The incomplete stop codon is a reported phenomenon for insect mitogenomes [15, 25, 49, 66]. It has been proposed that the completion of the TAA termination codon is due to a post-transcriptional polyadenylation synthesis [67].

The total number of non-stop codons within the predicted PCGs (3743) was similar to that reported for *A. darlingi* and *A. gambiae* (3733), differing slightly to that of *A. quadrimaculatus* (3715). The nucleotide bias was also reflected in the codon usage within the annotated PCGs. A strong bias towards A+T rich codons like TTA (L), TTT (F), ATT (I), ATA (M), and AAT (N) is a compositional tendency that follows what has been reported as the canonical pattern for Culicidae species [15, 66]. When compared to other insects, such as aphids, the same five codons were the most prevalent, differing only in the exchanged F and I positions [49].

Among the 32 most frequently used codons (RSCU > 1) 29 were of the NNU and NNA type. This indicates that the third position of the codons mostly are either U or A. This differs from the expected trend for dipterans in which, frequently used codons present G or C in the third position [53].

The mitochondrial genome of insects displays unique characteristics such as high codon bias, low GC₃ content, and a highly conserved gene arrangement [51]. These characteristics are associated with the conserved patterns of Shine-Dalgarno sequences found in the transcripts of functional proteins. In addition, the aforementioned characteristics promote the high expression level of the mitochondrial genes [68, 69]. As a consequence of the above, mitochondrial activity of insects and nematodes would render them more tolerant to increased concentrations of reactive oxygen species (ROS) [70]. In the case of anophelines, it is interesting to consider linking this enhanced cellular resistance hypothesis with the detoxification mechanisms [71] towards the oxidative stress triggered as a response to the immune challenge posed by the *Plasmodium* infection [72, 73].

The sequence identity comparison and compositional patterns analyses we performed showed how the nucleotide sequence of protein-coding genes are a conserved feature of anopheline mitogenomes, in which most differences have been reported allocated within the control region [15, 74]. In terms of sequence identity (all values > 86%), a pattern was observed (S3 Fig). In general, the *A. aquasalis* PCGs are more similar to those of the other Neotropical anophelines tested, when compared to those from *A. gambiae* and *A. punctulatus* (African and Asian malaria vectors). Higher sequence similarity could be obeying the phylogeographic and divergence history of anophelines. Considering that, on average, the highest similarity values were obtained with *A. albitarsis*, another brackish-water Neotropical anopheline, from which *A. aquasalis* diverged ~28 MYA (Fig 4). The identity between rRNA sequences ranged from 93% to 98% (S4 Table). Similar identity values described in aphids prompted Wang *et al.*, (2013) [49] to suggest that the predicted and potentially conserved rRNA secondary structures could improve phylogenetic analyses between insect species.

Human malaria vectors from different species complexes are not closely related. This implies that some of the competence traits potentially arose (or were lost) independently, multiple times within the different extant anopheline lineages. Therefore, the evolutionary history of genes associated with vector competence might have been driven by rapid processes. This would imply that such genes are not highly conserved sequences with a single ancestor as origin [75]. The publication of 16 anopheles genomes [8] might have generated, concomitantly, mitogenomes that will broaden our perspective on the phylogenetic and divergence relationships between the members of the *Anopheles* genus. This would be particularly important since both our study, and Logue *et al.*, (2013) [6], encountered the paradoxical situation regarding the scarcity of mitogenomes from dominant malaria vectors pertaining to Africa (until recently) [76, 77] and the East Mediterranean regions, areas still burdened by human malaria [78]. For instance, at the time of our study, we opted to leave out the mitogenome of

Anopheles funestus due the amount of gaps in the mtDNA PCGs sequences (GenBank accession number NC008070). It was not until recently that mitochondrial lineages of this important vector were thoroughly studied [77].

The reconstructed *Anopheles* phylogeny showed a deep divergence between two main *Anopheles* lineages (Fig 3) seemingly driven by their phylogeographic relations and the earth's geologic eras. The tree topology obtained is consistent with the current hypothesis regarding the origin of *Anopheles* mosquitoes in the Gondwana supercontinent during the Cretaceous period [12, 14]. *Anopheles* mosquitoes would have radiated from what is recognized as South America, then arrived in Africa and from there then colonized Europe and North America (with the aid of land bridges), also migrating through Asia, and into the Pacific [6].

However, as previously mentioned, paucity regarding mitogenomes from African, Asian and European anophelines, hindered the chance of determining accurately if African anophelines are the ancestors of those in Europe and North America, or if North American anophelines radiated from South America. For instance, both our results and those of Logue *et al.*, (2013) [6], represent a phylogeny in which the position of *Anopheles quadrimaculatus* seems unresolved.

Nonetheless, Freitas *et al.*, (2015) [7] provides a thorough revision of phylogenetic relations based on COI, COII and 5.8S rRNA genes which allowed for more species to be tested. When comparing with it, we observed how the mitogenome based relations here reported are the same at the subgenus level as Neafsey *et al.*, (2015) [8] and Krzywinski *et al.*, (2001) [10], being (*Nyssorhynchus*, (*Anopheles*, *Cellia*)). Whereas their study was not able to significantly resolve the evolutionary relationship within this taxonomic level, other studies have reported different associations too, namely, *Cellia*, (*Nyssorhynchus*, *Anopheles*) [79]. A recent comparative evolutionary mitochondriomic study strongly supported the sister relationship of *Nyssorhynchus* + *Kerteszia* and *Cellia* + *Anopheles* based on 50 complete mitogenomes [80]. Thus, this evolutionary relationship seems to be conclusively elucidated at this taxonomic level.

It has been stated that interpreting the current distributions of anophelines from an evolutionary context, may be quite problematic [7]. This author reviewed existing alternative hypotheses, some suggesting a different scenario for the evolution of the extant groups of the *Anophelinae* subfamily. The scenario would reflect the ideas presented by Christophers (1933) [81], in which the ancestral lineage of *Anopheles* existed prior to the split of Pangaea, diversifying into the modern species after the breakup of the continents. Nonetheless, it is not the scope of this study to dig deeper into this matter.

Regarding the estimated divergence times, we dated the *Anopheles* MRCA to ~ 83 MYA. The node age (Fig 4) differs from the ~ 79 MYA reported by Moreno *et al.*, (2010) [16], the ~ 93 MYA estimated by Logue *et al.*, (2013) [6] and 110 MYA by Freitas *et al.*, (2015) [78]. These molecular dating estimates, though different, are in a degree of agreement with the breakup of Western Gondwana and the loss of land connections between South America and Africa (90–95 MYA). This geologic event might have prompted the divergence between both *Anopheles* + *Cellia* from *Nyssorhynchus*, and further on, the divergence between *Cellia* and *Anopheles* ~ 71 MYA or ~ 81 MYA [6]. The geographic sorting of lineages thus would coincide with the loss of land bridges connecting Africa and Europe, as well as between Europe and North America. This, in turn, could account for the absence of *Cellia* in the New World and *Nyssorhynchus* in the Afro-Eurasian continents [10, 15]. In agreement with the above Freitas *et al.*, (2015) [78] determined, by ancestral area reconstruction, that a monophyletic clade composed of the Neotropical subgenera *Stethomyia*, *Kerteszia* and *Nyssorhynchus* had a common ancestor distributed along the Americas, and whose early radiation in the continent began around the Late Cretaceous (~ 90 MYA).

The data suggest that *A. aquasalis* diverged from the *A. albitarsis* complex ~ 28 MYA. The node age implies that their MRCA would be younger than the one shared between *A. darlingi* and the *A. albitarsis* complex which dates back to ~ 38 MYA (Fig 4). If Neotropical anophelines bionomics are considered, then *A. aquasalis* would have diverged from the *A. albitarsis* complex, adapting to its narrow coastal ecological niche, outcompeted by inland (*A. darlingi*) and brackish-water adapted (*A. albitarsis*) sibling species [82].

As brought forward by Freitas *et al.*, (2015) [78], when comparing their results to Fontaine *et al.*, (2015) [9], phylogenetic reconstructions for *Anopheles* require robust data from multiple molecular markers, and individuals per species. It is highlighted by the authors that unclear phylogenies regarding African anophelines, are due to complex speciation processes that have been permeated by recurring events of introgressive hybridization. Nonetheless, it drew our attention that, based on X chromosome phylogenies (which reflect species branching order), Fontaine *et al.*, (2015) [9], determined that brackish water adapted *Anopheles merus* and *Anopheles melas* [83], diverged from bionomical different inland sibling species ~1.85–1.47 MYA, whereas *A. aquasalis* might have diverged from inland sibling *A. darlingi* ~38.98 MYA. As evidence for phyllosymbiosis emerges, including examples of non-brackish water adapted or Neotropical anophelines [84], it becomes relevant to assess if co-evolution of *A. aquasalis* and its aquatic niche acquired microbiota, follow this trend. As a growing field of interest [85–87], exploring the ecological adaptations of anophelines to particular breeding sites, and its connection to their microbiota signature profiles, should generate interesting insights when comparing similar bionomical species for example.

The observed differences regarding radiation and divergence node ages have been reported by other authors [7, 12, 15]. Notably, the choices regarding data inclusion have a direct effect on phylogenomic studies. Therefore, the use of full mitogenomes, individual or concatenated genes (PCGs, rRNAs, tRNAs) or nuclear genes, can be accounted as the reason for the differences between molecular evolutionary histories reported to date [16, 51]. Additionally, observed discrepancies may also be due to: depth of species sampling while seeking for representatives of the geographical regions and /or the subgenera studied; the choice (and use) of mutation rates; and the amount of calibration points employed [6, 12, 15]. For example, Moreno *et al.*, (2010) [15] used as an additional calibration point, the Anophelinae and Culicinae divergence age estimated to be 120 MYA [88]. Nevertheless, the most rigorously calculated date available involving the split between a mosquito lineage and a sister taxon is the divergence time of 259.9 MYA between *Drosophila* and *Anopheles* estimated by Gaunt & Miles, (2002), as highlighted by Krzywinski *et al.*, (2006) [12, 44]. As there are many pitfalls inherent to the use of molecular clocks to date divergences, the outcome of age estimation should be interpreted with caution [12].

Molecular dating of the divergence events reported in this study occurred long before humans arrived in America 15–20 millennia ago [89]. Evidence suggests that *Plasmodium* parasites were introduced into the Americas via African slave trade routes during the European invasion [20]. In order to the transmission cycle to establish, the parasite had to adapt to the populations of local mosquitoes. Then, in accordance with the present geographic distribution of anophelines [17], and the known slave trade disembarking ports, some of the first indigenous vectors that *P. falciparum*-infected humans encountered would have been: *A. albimanus*, *A. aquasalis*, *A. darlingi*, *A. albitarsis* and *A. quadrimaculatus* [90].

Conclusion

Considering the evolutionary distance between Old and New World anophelines, it is not surprising that there are marked differences at genetic, ecological and behavioral levels between

them [17]. This probably resulted in a more stringent selection upon *Plasmodium* as it adapted to these new vectors [90]. Multiple Neotropical anophelines became human malaria vectors independently from each other, including the anophelines from South East Asian and Oceania [6]. This suggests that co-occurrences of anopheline traits related to malaria transmission would be the outcome of convergent evolution [91].

Readily available anopheline phylogenies are key to perform co-evolutionary studies between parasites and vectors to gain insights into their interspecific relationships [7]. Recently, Molina-Cruz & Barillas-Mury, (2014) [90], proposed a hypothesis linking the *Plasmodium* protein PFS47, and a still unknown interacting protein from *Anopheles*, acting as a critical determinant of mosquito-parasite compatibility and thus the “emergence” of vector competence in this genus [90, 92, 93]. Considering that the *Pfs47* gene is polymorphic, and presents a robust clonal distribution, we support the need of a better and more accurate *Anopheles* phylogeny. This, in turn, could enable the reconstruction of the history of malaria transmission in the New World.

Supporting information

S1 Fig. *A. aquasalis* mtDNA predicted tRNA structures. 22 tRNAs were identified in the mitogenome of *A. aquasalis* and their cloverleaf secondary structures predicted with RNAStructure. Using Dayhoff's single letter amino acid code.

(TIF)

S2 Fig. Compositional patterns of *A. aquasalis* and other anopheline mitochondrial genomes. A-T content expressed as AT%, AT skew, and GC skew were estimated and plotted for each single PCG and for other genomic regions according to the legend below the chart. Refer to the S5 and S6 Tables to access the tabular data regarding each parameter.

(TIF)

S3 Fig. Radar depiction of the nucleotide identity comparison between mitochondrial PCGs and rRNA genes from *A. aquasalis* and other anophelines. Nucleotide identity (%) between *A. aquasalis* and selected *Anopheles* species ranging from 80 to 100%. The scale is shown along the main vertical axis of the radar plot.

(TIF)

S1 Table. Base composition analysis of the *A. aquasalis* whole mitochondrion DNA sequence, and other annotated features such as individual PCGs, and concatenated rRNA & tRNA genes.

(XLSX)

S2 Table. Codon usage in the *A. aquasalis* mtDNA. Codon frequency and Relative synonymous Codon Usage (RSCU) are shown for each codon.

(XLSX)

S3 Table. Top 10 codons (classified by count) identified in the protein coding genes of the *A. aquasalis* mitogenome.

(XLSX)

S4 Table. Protein coding genes nucleotide sequence identity comparison between *A. aquasalis* and other selected anophelines.

(XLSX)

S5 Table. Comparative strand asymmetry analysis based on AT and GC skews between the *A. aquasalis* mtDNA and selected anopheline species.

(XLSX)

S6 Table. Comparative compositional analysis represented in AT% between the *A. aquasalis* mtDNA and selected anopheline species.

(XLSX)

Acknowledgments

The mitogenome characterization was performed at the Computational Biology facility at FIO-CRUZ-MG. Authors express their gratitude to Dr. Guilherme Oliveira, who lead this group at the time.

Author Contributions

Conceptualization: Paulo F. P. Pimenta.

Formal analysis: Luis Martinez-Villegas, Juliana Assis-Geraldo, Leonardo B. Koerich, Travis C. Collier, Yoosook Lee, Bradley J. Main, Nilton B. Rodrigues, Alessandra S. Orfano, Ana C. A. M. Pires, Thais B. Campolina, Rafael Nacif-Pimenta, Djane C. Baia-da-Silva, Ana P. M. Duarte, Ana C. Bahia, Claudia M. Rios-Velásquez, Marcus V. G. Lacerda, Wuelton M. Monteiro, Gregory C. Lanzaro, Nagila F. C. Secundino, Paulo F. P. Pimenta.

Project administration: Paulo F. P. Pimenta.

Resources: Paulo F. P. Pimenta.

Supervision: Paulo F. P. Pimenta.

Writing – review & editing: Paulo F. P. Pimenta.

References

1. Wolstenholme DR. Genetic novelties in mitochondrial genomes of multicellular animals. *Curr Opin Genet Dev.*1992; 2(6): 918–925. PMID: [1282405](https://pubmed.ncbi.nlm.nih.gov/1282405/)
2. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res.*1999; 27(8): 1767–1780. <https://doi.org/10.1093/nar/27.8.1767> PMID: [10101183](https://pubmed.ncbi.nlm.nih.gov/10101183/)
3. Mwinyi A, Meyer A, Bleidorn C, Lieb B, Bartolomaeus T, and Podsiadlowski L. Mitochondrial genome sequence and gene order of *Sipunculus nudus* give additional support for an inclusion of *Sipuncula* into *Annelida*. *BMC Genomics.*2009; 10: 27. <https://doi.org/10.1186/1471-2164-10-27> PMID: [19149868](https://pubmed.ncbi.nlm.nih.gov/19149868/)
4. Braband A, Cameron SL, Podsiadlowski L, Daniels SR, and Mayer G. The mitochondrial genome of the onychophoran *Opisthopatus cinctipes* (Peripatopsidae) reflects the ancestral mitochondrial gene arrangement of Panarthropoda and Ecdysozoa. *Mol Phylogenet Evol.*2010; 57(1): 285–292. <https://doi.org/10.1016/j.ympev.2010.05.011> PMID: [20493270](https://pubmed.ncbi.nlm.nih.gov/20493270/)
5. Harbach RE. Genus *Anopheles* Meigen, 1818. Mosquito Taxonomic Inventory Book Genus *Anopheles* Meigen, 1818 Mosquito Taxonomic Inventory City.2011.
6. Logue K, Chan ER, Phipps T, Small ST, Reimer L, Henry-Halldin C et al. Mitochondrial genome sequences reveal deep divergences among *Anopheles punctulatus* sibling species in Papua New Guinea. *Malar J.*2013; 12: 64. <https://doi.org/10.1186/1475-2875-12-64> PMID: [23405960](https://pubmed.ncbi.nlm.nih.gov/23405960/)
7. Harbach RE. The phylogeny and classification of *Anopheles*. InTech;2013.
8. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science.*2015; 347(6217): 1258522. <https://doi.org/10.1126/science.1258522> PMID: [25554792](https://pubmed.ncbi.nlm.nih.gov/25554792/)
9. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV et al. Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science.*2015; 347(6217): 1258524. <https://doi.org/10.1126/science.1258524> PMID: [25431491](https://pubmed.ncbi.nlm.nih.gov/25431491/)

10. Krzywinski J, Wilkerson RC, and Besansky NJ. Evolution of mitochondrial and ribosomal gene sequences in anophelinae (Diptera: Culicidae): implications for phylogeny reconstruction. *Mol Phylogenet Evol.*2001; 18(3): 479–487. <https://doi.org/10.1006/mpev.2000.0894> PMID: 11277639
11. Ad Torre, Fanello C, Akogbeto M, Dossou-Yovo J, Favia G, Petrarca V et al. Molecular evidence of incipient speciation within *Anopheles gambiae* ss in West Africa. *Insect molecular biology*. Wiley Online Library;2001; 10(1): 9–18. PMID: 11240632
12. Krzywinski J, Grushko OG, and Besansky NJ. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol.*2006; 39(2): 417–423. <https://doi.org/10.1016/j.ympev.2006.01.006> PMID: 16473530
13. Reidenbach KR, Cook S, Bertone MA, Harbach RE, Wiegmann BM, and Besansky NJ. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. *BMC Evol Biol.*2009; 9: 298. <https://doi.org/10.1186/1471-2148-9-298> PMID: 20028549
14. Krzywinski J and Besansky NJ. Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annu Rev Entomol.*2003; 48: 111–139. <https://doi.org/10.1146/annurev.ento.48.091801.112647> PMID: 12208816
15. Moreno M, Marinotti O, Krzywinski J, Tadei WP, James AA, Achee NL et al. Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malar J.*2010; 9: 127. <https://doi.org/10.1186/1475-2875-9-127> PMID: 20470395
16. Marinotti O, Cerqueira GC, de Almeida LG, Ferro MI, Loreto EL, Zaha A et al. The genome of *Anopheles darlingi*, the main neotropical malaria vector. *Nucleic Acids Res.*2013; 41(15): 7387–7400. <https://doi.org/10.1093/nar/gkt484> PMID: 23761445
17. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M et al. A global map of dominant malaria vectors. *Parasit Vectors.*2012; 5: 69. <https://doi.org/10.1186/1756-3305-5-69> PMID: 22475528
18. De Castro MC and Singer BH. Was malaria present in the Amazon before the European conquest? Available evidence and future research agenda. *Journal of Archaeological Science*. Elsevier;2005; 32(3): 337–340.
19. Culleton R, Coban C, Zeyrek FY, Cravo P, Kaneko A, Randrianarivelosia M et al. The origins of African *Plasmodium vivax*; insights from mitochondrial genome sequencing. *PLoS One.*2011; 6(12): e29137. <https://doi.org/10.1371/journal.pone.0029137> PMID: 22195007
20. Yalcindag E, Elguero E, Arnathau C, Durand P, Akiana J, Anderson TJ et al. Multiple independent introductions of *Plasmodium falciparum* in South America. *Proc Natl Acad Sci USA.*2012; 109(2): 511–516. <https://doi.org/10.1073/pnas.1119058109> PMID: 22203975
21. Pimenta PF, Orfano AS, Bahia AC, Duarte AP, Rios-Velasquez CM, Melo FF et al. An overview of malaria transmission from the perspective of Amazon Anopheles vectors. *Mem Inst Oswaldo Cruz.*2015; 110(1): 23–47. <https://doi.org/10.1590/0074-02760140266> PMID: 25742262
22. Masta SE, Longhorn SJ, and Boore JL. Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. *Mol Phylogenet Evol.*2009; 50(1): 117–128. <https://doi.org/10.1016/j.ympev.2008.10.010> PMID: 18992830
23. Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB et al. Phylogeography's past, present, and future: 10 years after *Avis*, 2000. *Mol Phylogenet Evol.*2010; 54(1): 291–301. <https://doi.org/10.1016/j.ympev.2009.09.016> PMID: 19755165
24. Santini F, Kong X, Sorenson L, Carnevale G, Mehta RS, and Alfaro ME. A multi-locus molecular timescale for the origin and diversification of eels (Order: Anguilliformes). *Mol Phylogenet Evol.*2013; 69(3): 884–894. <https://doi.org/10.1016/j.ympev.2013.06.016> PMID: 23831455
25. Zhao Z, Su TJ, Chesters D, Wang SD, Ho SY, Zhu CD et al. The mitochondrial genome of *Elodia flavipalpis* Aldrich (Diptera: Tachinidae) and the evolutionary timescale of *Tachinid* flies. *PLoS One.*2013; 8(4): e61814. <https://doi.org/10.1371/journal.pone.0061814> PMID: 23626734
26. Perseke M, Golombek A, Schlegel M, and Struck TH. The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. *Mol Phylogenet Evol.*2013; 66(3): 898–905. <https://doi.org/10.1016/j.ympev.2012.11.019> PMID: 23228546
27. de Carvalho SC, Martins Junior AJ, Lima JB, and Valle D. Temperature influence on embryonic development of *Anopheles albittarsis* and *Anopheles aquasalis*. *Mem Inst Oswaldo Cruz.*2002; 97(8): 1117–1120. <https://doi.org/10.1590/s0074-02762002000800009> PMID: 12563476
28. da Silva AN, Dos Santos CC, Lacerda RN, Santa Rosa EP, De Souza RT, Galiza D et al. Laboratory colonization of *Anopheles aquasalis* (Diptera: Culicidae) in Belem, Para, Brazil. *J Med Entomol.*2006; 43(1): 107–109. PMID: 16506455

29. Rios-Velasquez CM, Martins-Campos KM, Simoes RC, Izzo T, dos Santos EV, Pessoa FA et al. Experimental *Plasmodium vivax* infection of key *Anopheles* species from the Brazilian Amazon. *Malar J.*2013; 12: 460. <https://doi.org/10.1186/1475-2875-12-460> PMID: 24359307
30. Zerbino DR and Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.*2008; 18(5): 821–829. <https://doi.org/10.1101/gr.074492.107> PMID: 18349386
31. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C et al. Versatile and open software for comparing large genomes. *Genome Biol.*2004; 5(2): R12. <https://doi.org/10.1186/gb-2004-5-2-r12> PMID: 14759262
32. Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsche G et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.*2013; 69(2): 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023> PMID: 22982435
33. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA et al. Artemis: sequence visualization and annotation. *Bioinformatics.*2000; 16(10): 944–945. <https://doi.org/10.1093/bioinformatics/16.10.944> PMID: 11120685
34. Lowe TM and Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*1997; 25(5): 955–964. <https://doi.org/10.1093/nar/25.5.955> PMID: 9023104
35. Bellaousov S, Reuter JS, Seetin MG, and Mathews DH. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*2013; 41(Web Server issue): W471–W474. <https://doi.org/10.1093/nar/gkt290> PMID: 23620284
36. Alikhan NF, Petty NK, Ben Zakour NL, and Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.*2011; 12: 402. <https://doi.org/10.1186/1471-2164-12-402> PMID: 21824423
37. Perna NT and Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol.*1995; 41(3): 353–358. PMID: 7563121
38. Tamura K, Stecher G, Peterson D, Filipowski A, and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.*2013; 30(12): 2725–2729. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
39. Zhang Z, Schwartz S, Wagner L, and Miller W. A greedy algorithm for aligning DNA sequences. *Journal of Computational biology.* Mary Ann Liebert, Inc.;2000; 7(1–2): 203–214. <https://doi.org/10.1089/10665270050081478> PMID: 10890397
40. Abascal F, Zardoya R, and Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.*2010; 38(Web Server issue): W7–13. <https://doi.org/10.1093/nar/gkq291> PMID: 20435676
41. Kuck P and Meusemann K. FASconCAT: Convenient handling of data matrices. *Mol Phylogenet Evol.*2010; 56(3): 1115–1118. <https://doi.org/10.1016/j.ympev.2010.04.024> PMID: 20416383
42. Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol.*2008; 25(7): 1253–1256. <https://doi.org/10.1093/molbev/msn083> PMID: 18397919
43. Drummond AJ and Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology.* BioMed Central;2007; 7(1): 214.
44. Gaunt MW and Miles MA. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Biology and Evolution.*2002; 19(5): 748–761. <https://doi.org/10.1093/oxfordjournals.molbev.a004133> PMID: 11961108
45. Yamasaki YK, Nieman CC, Chang A, Collier TC, Main BJ, and Lee Y. Improved tools for genomic DNA library construction of small insects [version 1; not peer reviewed]. *F1000Res.*2016; (5): 211.
46. Plazzi F, Ribani A, and Passamonti M. The complete mitochondrial genome of *Solemya velum* (Mollusca: Bivalvia) and its relationships with conchifera. *BMC Genomics.*2013; 14: 409. <https://doi.org/10.1186/1471-2164-14-409> PMID: 23777315
47. Li T, Gao C, Cui Y, Xie Q, and Bu W. The complete mitochondrial genome of the stalk-eyed bug *Chauliops fallax* Scott, and the monophyly of Malcidae (Hemiptera: Heteroptera). *PLoS One.*2013; 8(2): e55381. <https://doi.org/10.1371/journal.pone.0055381> PMID: 23390534
48. Boore JL, Macey JR, and Medina M. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol.*2005; 395: 311–348. [https://doi.org/10.1016/S0076-6879\(05\)95019-2](https://doi.org/10.1016/S0076-6879(05)95019-2) PMID: 15865975
49. Wang Y, Huang XL, and Qiao GX. Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. *PLoS One.*2013; 8(10): e77511. <https://doi.org/10.1371/journal.pone.0077511> PMID: 24147014

50. Poinar GO, Zavortinik TJ, Pike T, and Johnston PA. *Paleoculicis minutus* (Diptera: Culicidae) n. Gen., n. Sp., from Cretaceous Canadian amber, with a summary of described fossil mosquitoes. *Acta Geologica Hispanica*.2000; 35(1): 119–130.
51. Cameron SL. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu Rev Entomol*.2014; 59: 95–117. <https://doi.org/10.1146/annurev-ento-011613-162007> PMID: 24160435
52. Stewart JB and Beckenbach AT. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). *Genome*. NRC Research Press;2005; 48(1): 46–54. <https://doi.org/10.1139/g04-090> PMID: 15729396
53. Behura SK and Severson DW. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One*.2012; 7(8): e43111. <https://doi.org/10.1371/journal.pone.0043111> PMID: 22912801
54. Timmis JN, Ayliffe MA, Huang CY, and Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*. Nature Publishing Group;2004; 5(2): 123. <https://doi.org/10.1038/nrg1271> PMID: 14735123
55. Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol*.1992; 141: 173–216. PMID: 1452431
56. Fauron CM and Wolstenholme DR. Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules. *Nucleic Acids Res*.1980; 8(11): 2439–2452. <https://doi.org/10.1093/nar/8.11.2439> PMID: 6255406
57. Inohira K, Hara T, and Matsuura ET. Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. *Mol Biol Evol*.1997; 14(8): 814–822. <https://doi.org/10.1093/oxfordjournals.molbev.a025822> PMID: 9254919
58. Schneider A and Marechal-Drouard L. Mitochondrial tRNA import: are there distinct mechanisms? *Trends in cell biology*. Elsevier;2000; 10(12): 509–513. PMID: 11121736
59. Tomita K and Weiner AM. Collaboration between CC-and A-adding enzymes to build and repair the 3'-terminal CCA of tRNA in *Aquifex aeolicus*. *Science*.2001; 294(5545): 1334–1336. <https://doi.org/10.1126/science.1063816> PMID: 11701927
60. Li R and Guan MX. Human mitochondrial leucyl-tRNA synthetase corrects mitochondrial dysfunctions due to the tRNA^{Leu}(UUR) A3243G mutation, associated with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like symptoms and diabetes. *Mol Cell Biol*.2010; 30(9): 2147–2154. <https://doi.org/10.1128/MCB.01614-09> PMID: 20194621
61. Silverio MS, Rodovalho VdR, Bonetti AM, de Oliveira GC, Cuadros-Orellana S, Ueira-Vieira C et al. Preliminary characterization of mitochondrial genome of *Melipona scutellaris*, a Brazilian stingless bee. *BioMed research international*. Hindawi;2014; 2014: 1–6.
62. Wei SJ, Shi M, Chen XX, Sharkey MJ, van AC, Ye GY et al. New views on strand asymmetry in insect mitochondrial genomes. *PLoS One*.2010; 5(9): e12708. <https://doi.org/10.1371/journal.pone.0012708> PMID: 20856815
63. Hassanin A, Leger N, and Deutsch J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. *Syst Biol*.2005; 54(2): 277–298. <https://doi.org/10.1080/10635150590947843> PMID: 16021696
64. Beard CB, Hamm DM, and Collins FH. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol Biol*.1993; 2(2): 103–124. PMID: 9087549
65. Mitchell SE, Cockburn AF, and Seawright JA. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome*.1993; 36(6): 1058–1073. PMID: 8112570
66. Behura SK, Lobo NF, Haas B, deBruyn B, Lovin DD, Shumway MF et al. Complete sequences of mitochondrial genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. *Insect Biochem Mol Biol*.2011; 41(10): 770–777. <https://doi.org/10.1016/j.ibmb.2011.05.006> PMID: 21640823
67. Wei Sj, Shi M, He Jh, Sharkey M, and Chen Xx. The complete mitochondrial genome of *Diadegma semiclausum* (Hymenoptera: Ichneumonidae) indicates extensive independent evolutionary events. *Genome*. NRC Research Press;2009; 52(4): 308–319. <https://doi.org/10.1139/g09-008> PMID: 19370087
68. Duret L and Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila* and *Arabidopsis*. *Proc Natl Acad Sci USA*.1999; 96(8): 4482. <https://doi.org/10.1073/pnas.96.8.4482> PMID: 10200288

69. Hiraoka Y, Kawamata K, Haraguchi T, and Chikashige Y. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. *Genes Cells*.2009; 14(4): 499–509. <https://doi.org/10.1111/j.1365-2443.2009.01284.x> PMID: 19335619
70. Pandey A, Suman S, and Chandna S. Predictive role of mitochondrial genome in the stress resistance of insects and nematodes. *Bioinformatics*.2010; 5(1): 21–27. <https://doi.org/10.6026/97320630005021> PMID: 21346874
71. Jaramillo-Gutierrez G, Molina-Cruz A, Kumar S, and Barillas-Mury C. The *Anopheles gambiae* oxidation resistance 1 (OXR1) gene regulates expression of enzymes that detoxify reactive oxygen species. *PLoS One*.2010; 5(6): e11168. <https://doi.org/10.1371/journal.pone.0011168> PMID: 20567517
72. Bahia AC, Oliveira JH, Kubota MS, Araujo HR, Lima JB, Rios-Velasquez CM et al. The role of reactive oxygen species in *Anopheles aquasalis* response to *Plasmodium vivax* infection. *PLoS One*.2013; 8(2): e57014. <https://doi.org/10.1371/journal.pone.0057014> PMID: 23441231
73. Shrinet J, Nandal UK, Adak T, Bhatnagar RK, and Sunil S. Inference of the oxidative stress network in *Anopheles stephensi* upon plasmodium infection. *PLoS One. Public Library of Science*;2014; 9(12): e114461. <https://doi.org/10.1371/journal.pone.0114461> PMID: 25474020
74. Oliveira MT, Azeredo-Espin AM, and Lessinger AC. The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context. *J Mol Evol*.2007; 64(5): 519–527. <https://doi.org/10.1007/s00239-006-0099-6> PMID: 17460806
75. Besansky NJ. Genome analysis of vectorial capacity in major *Anopheles* vectors of malaria parasites. Document hosted by VectorBase.2014.
76. Hanemaaijer MJ, Houston PD, Collier TC, Norris LC, Fofana A, Lanzaro GC et al. Mitochondrial genomes of *Anopheles arabiensis*, *An. gambiae* and *An. coluzzii* show no clear species division. *F1000Res*.2018; 7: 347. <https://doi.org/10.12688/f1000research.13807.2> PMID: 31069048
77. Jones CM, Lee Y, Kitchen A, Collier T, Pringle JC, Muleba M et al. Complete *Anopheles funestus* mitogenomes reveal an ancient history of mitochondrial lineages and their distribution in southern and central Africa. *Scientific Reports*.2018; 8(1): 9054. <https://doi.org/10.1038/s41598-018-27092-y> PMID: 29899497
78. Freitas LA, Russo CA, Voloch CM, Mutaquiha OC, Marques LP, and Schrago CG. Diversification of the Genus *Anopheles* and a Neotropical Clade from the Late Cretaceous. *PLoS One*.2015; 10(8): e0134462. <https://doi.org/10.1371/journal.pone.0134462> PMID: 26244561
79. Sallum MAM, Schultz TR, Foster PG, Aronstein K, Wirtz RA, and Wilkerson RC. Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. *Systematic Entomology. Wiley Online Library*;2002; 27(3): 361–382.
80. Hao YJ, Zou YL, Ding YR, Xu WY, Yan ZT, Li XD et al. Complete mitochondrial genomes of *Anopheles stephensi* and *An. dirus* and comparative evolutionary mitochondrialomics of 50 mosquitoes. *Sci Rep*.2017; 7(1): 7666. <https://doi.org/10.1038/s41598-017-07977-0> PMID: 28794438
81. CHRISTOPHERS SR. The Fauna of British India, including Ceylon and Burma. Diptera. Vol. IV. Family Culicidae. Tribe Anophelini. London Taylor and Francis;1933; 1–371.
82. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW et al. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic precis. *Parasit Vectors*.2010; 3: 72. <https://doi.org/10.1186/1756-3305-3-72> PMID: 20712879
83. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J et al. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis. *Parasit Vectors*.2010; 3: 117. <https://doi.org/10.1186/1756-3305-3-117> PMID: 21129198
84. Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, and Bordenstein SR. Correction: Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biol*.2017; 15(1): e1002587. <https://doi.org/10.1371/journal.pbio.1002587> PMID: 28068336
85. Ngo CT, Romano-Bertrand S, Manguin S, and Jumas-Bilak E. Diversity of the Bacterial Microbiota of *Anopheles* Mosquitoes from Binh Phuoc Province, Vietnam. *Front Microbiol*.2016; 7: 2095. <https://doi.org/10.3389/fmicb.2016.02095> PMID: 28066401
86. Buck M, Nilsson LK, Brunius C, Dabire RK, Hopkins R, and Terenius O. Bacterial associations reveal spatial population dynamics in *Anopheles gambiae* mosquitoes. *Sci Rep*.2016; 6: 22806. <https://doi.org/10.1038/srep22806> PMID: 26960555
87. Akorli J, Gendrin M, Pels NA, Yeboah-Manu D, Christophides GK, and Wilson MD. Seasonality and Locality Affect the Diversity of *Anopheles gambiae* and *Anopheles coluzzii* Midgut Microbiota from Ghana. *PLoS One*.2016; 11(6): e0157529. <https://doi.org/10.1371/journal.pone.0157529> PMID: 27322614

88. Rai KS and Black WC. Mosquito genomes: structure, organization, and evolution. *Adv Genet.* 1999; 41: 1–33. PMID: [10494615](https://pubmed.ncbi.nlm.nih.gov/10494615/)
89. Stringer C. *The origin of our species.* Penguin UK; 2011; 1.
90. Molina-Cruz A and Barillas-Mury C. The remarkable journey of adaptation of the *Plasmodium falciparum* malaria parasite to New World anopheline mosquitoes. *Mem Inst Oswaldo Cruz.* 2014; 109(5): 662–667. <https://doi.org/10.1590/0074-0276130553> PMID: [25185006](https://pubmed.ncbi.nlm.nih.gov/25185006/)
91. White BJ, Collins FH, and Besansky NJ. Evolution of *Anopheles gambiae* in relation to humans and malaria. *Annual review of ecology, evolution, and systematics.* Annual Reviews. 2011; 42: 111–132.
92. Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, Winikor J et al. The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. *Science.* 2013; 340(6135): 984–987. <https://doi.org/10.1126/science.1235264> PMID: [23661646](https://pubmed.ncbi.nlm.nih.gov/23661646/)
93. Canepa GE, Molina-Cruz A, and Barillas-Mury C. Molecular Analysis of Pfs47-Mediated *Plasmodium* Evasion of Mosquito Immunity. *PLoS One.* 2016; 11(12): e0168279. <https://doi.org/10.1371/journal.pone.0168279> PMID: [27992481](https://pubmed.ncbi.nlm.nih.gov/27992481/)