


RESEARCH ARTICLE

Complete mitogenome of *Anopheles sinensis* and mitochondrial insertion segments in the nuclear genomes of 19 mosquito species

Yi-Ran Ding^{1,2} , Bo Li^{1,2}, Yu-Juan Zhang^{1,2}, Qi-Meng Mao^{1,2}, Bin Chen^{1,2} *

1 Chongqing Key Laboratory of Vector Insects, Chongqing Normal University, Chongqing, P.R. China, **2** Institute of Entomology and Molecular Biology, Chongqing Normal University, Chongqing, P.R. China

 These authors contributed equally to this work.

* bin.chen@cqnu.edu.cn



Abstract

Anopheles sinensis is a major malarial vector in China and Southeast Asia. The mitochondria is involved in many important biological functions. Nuclear mitochondrial DNA segments (NUMTs) are common in eukaryotic organisms, but their characteristics are poorly understood. We sequenced and analyzed the complete mitochondrial (mt) genome of *An. sinensis*. The mt genome is 15,418 bp long and contains 13 protein-coding genes (PCGs), two rRNAs, 22 tRNAs and a large non-coding region. Its gene arrangement is similar to previously published mosquito mt genomes. We identified and analyzed the NUMTs of 19 mosquito species with both nuclear genomes and mt genome sequences. The number, total length and density of NUMTs are significantly correlated with genome size. About half of NUMTs are short (< 200 bp), but larger genomes can house longer NUMTs. NUMTs may help explain genome size expansion in mosquitoes. The expansion due to mitochondrial insertion segments is variable in different insect groups. PCGs are transferred to nuclear genomes at a higher frequency in mosquitoes, but NUMT origination is more different than in mammals. Larger-sized nuclear genomes have longer mt genome sequences in both mosquitoes and mammals. The study provides a foundation for the functional research of mitochondrial genes in *An. sinensis* and helps us understand the characteristics and origin of NUMTs and the potential contribution to genome expansion.

OPEN ACCESS

Citation: Ding Y-R, Li B, Zhang Y-J, Mao Q-M, Chen B (2018) Complete mitogenome of *Anopheles sinensis* and mitochondrial insertion segments in the nuclear genomes of 19 mosquito species. PLoS ONE 13(9): e0204667. <https://doi.org/10.1371/journal.pone.0204667>

Editor: Y-h. Taguchi, Chuo University, JAPAN

Received: April 18, 2018

Accepted: September 12, 2018

Published: September 27, 2018

Copyright: © 2018 Ding 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: All relevant data are within the paper and its Supporting Information files.

Funding: This research was supported by the followings: The National Natural Science Foundation of China (31672363, 31372265), National Key Program of Science and Technology Foundation Work of China (2015FY210300) and Science and Technology Major Special Project of Guangxi (GKAA17129002). The funders had no role in study design, data collection and analysis,

Introduction

Mitochondria are eukaryotic cell organelles that are mainly involved in oxidative phosphorylation [1–2]. The conservation, easy alignment, maternal inheritance, and straightforward gene orthology of mitochondrial (mt) genomes have made the mt genome important for studies of phylogeny and evolution [3–6]. Mt genomes are sometimes associated with insecticide resistance. Several transcripts encoding enzymes such as NADH dehydrogenase and ATP synthase, which are involved in the production of energy within the respiratory chain, were over-expressed in *Aedes aegypti* larvae exposed to insecticides [7]. *Anopheles sinensis* is a major

decision to publish, or preparation of the manuscript.

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

malarial vector in Asia [8], and its mt genome was partially sequenced (15,076 bp) and annotated using collections from Shandong, China [9]. *An. sinensis* is being used as a model species to study the molecular mechanism of insecticide resistance [10–11]. To elucidate the function of the mt genome in insecticide resistance, the mt genome need to be completely sequenced and annotated with a laboratory strain originally collected from Jiangsu, China.

Nuclear mitochondrial DNA segments (NUMT) (mitochondrial DNA in the nuclear genome) exist widely in eukaryotes [12]. NUMTs probably arise from nonhomologous recombination of nuclear DNA with mt genome segments from damaged mitochondria [13,14]. The NUMTs found in the nuclear genome are highly similar to mtDNA sequences, so they are easily amplified using universal primers of mtDNA target sequence [15–17]. The existence of NUMTs does not only increases the difficulty of obtaining mtDNA target sequences, but also leads to incorrect conclusions in molecular identification [16–19]. An possible example is the molecular identification of *Dendrocygna arcuate*, which are completely different based on COI-NUMTs and COI-mtDNA, respectively [20]. At the same time, the transferred mtDNA sequences are thought to be molecular fossils in the nuclear genome and thus identification of NUMTs can help in phylogeny and evolution research [21]. Therefore, NUMTs can be useful molecular markers for the phylogenetic studies of *Homo sapiens* [22] and *Arabidopsis thaliana* [23].

Therefore, the genome-wide identification of NUMTs is very important for the molecular identification and phylogenetic study. NUMTs have been identified in several insect species, including *Atta cephalotes* [24], *Podisma pedestris* [25] and *Sitobion miscanthi* [26]. There is no NUMT in the *An. gambiae* genome [21], whereas NUMTs exist in *Culex quinquefasciatus* and *Aedes aegypti* genomes with the NUMT number and total length varying between the two species [27]. The genomes of *An. gambiae* [28], *Cx. quinquefasciatus* [29], *Aedes aegypti* [30] and *Ae. albopictus* [31] have been sequenced and annotated. Recently 16 *Anopheles* genomes have been reported [32]. The nuclear genome of *An. sinensis* has been sequenced at the Chongqing Normal University, China. We asked if the number, total length and density of NUMTs are correlated with genome size. Does the mt genome contribute to nuclear genome expansion? Which kinds of genes are most likely to be transferred to the nuclear genome? Sixteen mosquito species, with known phylogeny and known nuclear genome and mt genome sequences, provided us with an opportunity to answer these questions.

We sequenced the complete mt genome of *An. sinensis* and analyzed its characteristics, including gene organization, base composition, codon usage, and tRNA secondary structure. We also studied the NUMTs of 19 mosquito species for which both mt genome and nuclear genome sequences are available. The NUMT characteristics, including NUMT number, position, and density in each nuclear genome were determined. This study is significant for the annotation of the complete mt genome of *An. sinensis* and for understanding NUMT status and characteristics in mosquitoes.

Materials and methods

Sequencing and analysis of *An. sinensis* mt genome

The *An. sinensis* sample for mt genome sequencing was from an *An. sinensis* pyrethroid-susceptible laboratory colony. This colony was established from individuals collected in Wuxi, Jiangsu Province, China and cultured at Chongqing Normal University, China. Mitochondrial genomic DNA was isolated from a single adult female using the sodium dodecyl sulfate (SDS)/proteinase K digestion method [33].

Mitochondrial DNA fragments, ranging in size from 822 bp to 1371 bp, were amplified by PCR using a primer set specific for mosquitoes [34]. A 25- μ l PCR reaction was prepared with

Takara Ex Taq polymerase (Takara, Japan) and 1.5 μ l of 25 mM MgCl₂, 2.5 μ l of 10 \times PCR Buffer (Mg²⁺ free), 3 μ l of a dNTP mixture (2.5 mM each), 1 μ l of each 10 mM primer [34], 0.2 μ l of 5 U/ μ l Taq polymerase and 1 μ l of the mt genome DNA template. PCR thermal cycling included a 5 min initial denaturation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 48°C–55°C for 45 s, and elongation at 68°C for 1 min, followed by a final elongation for 10 min at 72°C. The PCR products were electrophoresed on a 1.0% agarose gel and then purified using a QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). The purified PCR products were directly sequenced using the primer sets except for the control region. The purified products were loaded into pMD-19T vectors, cloned, and then sequenced. All fragments were sequenced in both directions.

The obtained sequences were edited using DANMAN (<http://www.lynnon.com/>) and were identified in reference to annotated mosquito mt genome sequences through alignment using Clustal X [35]. The sequences of PCGs were translated into amino acids using MEGA version 6.0 [36]. Almost all tRNAs were also recognized by the online tRNAscan-SE Search Server v.1.21 [37], and the tRNAs (*tRNA^{Arg}*, *tRNA^{Ser(AGN)}*) that could not be found by tRNAscan-SE were confirmed by sequence homology comparison. The CR was examined for repeats and special structures with the aid of the Tandem Repeats Finder (<http://tandem.bu.edu/trf/trf.html>) [38]. The nucleotide composition was calculated using DNA Star (<http://www.dnastar.com/>). Codon usage bias was calculated using MEGA version 6.0 [36]. Strand asymmetry was evaluated by AT Skew and GC Skew using the formulae AT skew = [A%–T%]/ [A%+T%] and GC skew = [G%–C%]/ [G%+C%], respectively [39].

Nuclear mitochondrial DNA insertion analysis

Eighteen mosquito species' nuclear genome sequences in FASTA format were downloaded from VectorBase (<https://www.vectorbase.org/>). The genome sequence of *An. sinensis* was obtained from Chongqing Normal University (unpublished data) (S1 Table). The *An. sinensis* genome has an assembly scaffold size of 194.49 Mb, with gene area coverage 98.90%. Of the 19 mosquito species with genomes analyzed for NUMTs, three species belonged to the Culicinae with two species in the genus *Aedes* and one in the genus *Culex*. The remaining 16 species were all in the genus *Anopheles* and belonged to the Anophelinae. The 18 previously determined mt genome sequences were downloaded from NCBI, and the mt genome sequence of *An. sinensis* was obtained in the present study (S1 Table).

The NUMTs were identified using mt genome sequences to search against the nuclear genome sequence using BLASTN for each species. The significance threshold for the BLASTN search was set to $E < 10^{-4}$, with the window slide of 30 bp [40]. The number, total length and density (NUMTs per Mb of nuclear genome sequence) were calculated with Excel 2010 and used to measure the basic characteristics of NUMTs' occurrence in the genome. Excel 2010 was also used to count the number and different lengths of the NUMTs. The statistical analyses were carried out using TIBCO Statistica (<https://www.tibco.com/products/tibco-statistica>). An R package was written to calculate the location and frequency of different areas of the mt genome sequences that had been transferred to the nuclear genome.

Results and discussion

An. sinensis mt genome and its organization

We determined the complete mt genome sequence of *An. sinensis*. It is a typical circular and double-stranded molecule of 15,418 bp (GenBank accession MF322628). The obtained mt genome sequence was 342 bp longer than an earlier report of the *An. sinensis* mt genome [9], and the difference was mainly due to earlier incomplete sequencing of the mt genome control

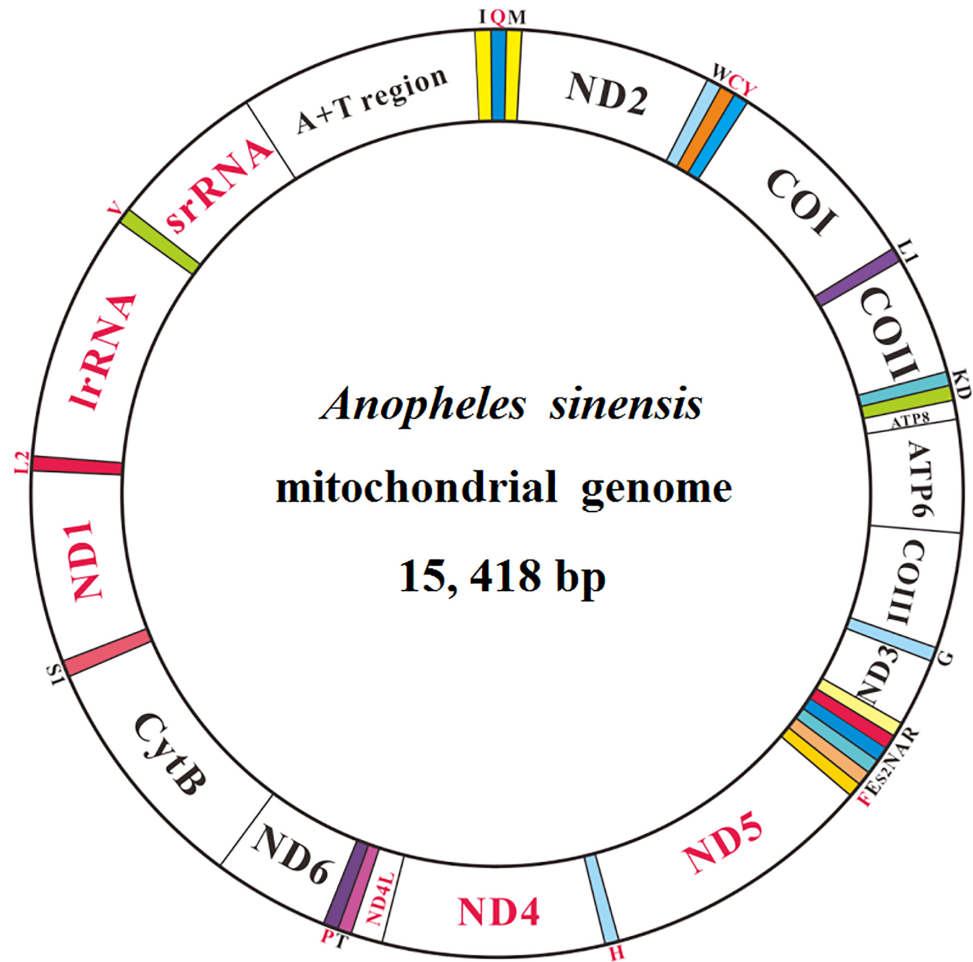


Fig 1. Structure of the *Anopheles sinensis* mt genome. The color-filled blocks indicate tRNAs, and the unfilled white blocks denote protein-coding genes (PCGs), rRNAs and control region (CR). The tRNAs, PCGs, rRNAs and CR located on the majority strand are marked in black, and those on the minority strand are marked in red.

<https://doi.org/10.1371/journal.pone.0204667.g001>

region (CR). Other genes lengths in our assembly of *An. sinensis* mtgenome are same as previous assembly sequence (Genbank accession NC028016) expect for control region. The two mt genome sequences have a 98.6% similarity, and a total of 151 single nucleotide polymorphisms (SNPs) were identified between them, except for CR.

The mt genome sequence of *An. sinensis* contains a conserved set of 37 genes, including 13 protein-coding genes (PCGs), two rRNA genes (IrRNA and srRNA), 22 tRNA genes (tRNAs) and a large non-coding region (CR, also known as the AT-rich region) (Fig 1). Twenty-three genes are located on the majority strand (J-strand), while 14 genes reside on the minority strand (N-strand) (Fig 1 and Table 1). The gene arrangement is the same as in previously published mosquito mt genomes, and the unique difference of the arrangement from other dipteran species mt genomes is that the latter have the order trnA-trnR [41]. This arrangement difference might be associated with different adaptations and evolutionary histories of mosquitoes [42].

Characteristics of the *An. sinensis* mt genome

The A+T content of *An. sinensis* mt genome is 78.34%, and the A+T contents of PCGs, tRNAs, rRNAs and CR are 76.85%, 78.59%, 81.46% and 93.58%, respectively (Table 2).

Table 1. Organization of the *Anopheles sinensis* mt genome.

Gene	Position	Strand	Size (bp)	Anticodon	Start codon	Stop codon	No. of intergenic nucleotides
<i>tRNAIle</i>	1–68	+	68	GAT			
<i>tRNAGln</i>	66–134	-	69	TTG			-3
<i>tRNAMet</i>	134–202	+	69	CAT			-1
<i>ND2</i>	203–1228	+	1026		ATT	TAA	0
<i>tRNATrp</i>	1227–1295	+	69	TCA			-2
<i>tRNACys</i>	1295–1358	-	64	GCA			-1
<i>tRNATyr</i>	1360–1425	-	66	GTA			1
<i>COI</i>	1424–2965	+	1542		TCG	TAA	-2
<i>tRNALeu (UUR)</i>	2961–3026	+	66	TAA			-5
<i>COII</i>	3028–3712	+	685		ATG	T	1
<i>tRNALys</i>	3713–3784	+	72	CTT			0
<i>tRNAAsp</i>	3793–3861	+	69	GTC			8
<i>ATP8</i>	3862–4023	+	162		ATT	TAA	0
<i>ATP6</i>	4017–4697	+	681		ATG	TAA	7
<i>COIII</i>	4697–5483	+	787		ATG	T	-1
<i>tRNAGly</i>	5484–5550	+	67	TCC			0
<i>ND3</i>	5551–5904	+	354		ATA	TAA	0
<i>tRNAArg</i>	5903–5966	+	64	TCG			-2
<i>tRNAAla</i>	5967–6032	+	66	TGC			0
<i>tRNAAsn</i>	6033–6099	+	67	GTT			0
<i>tRNASer (AGN)</i>	6102–6168	-	67	GCT			2
<i>tRNAGlu</i>	6170–6235	+	66	TTC			1
<i>tRNAPhe</i>	6234–6300	-	67	GAA			0
<i>ND5</i>	6301–8043	-	1743		GTG	TAA	0
<i>tRNAHis</i>	8044–8108	-	65	GTG			0
<i>ND4</i>	8106–9450	-	1345		ATG	T	-3
<i>ND4L</i>	9444–9743	-	300		ATG	TAA	-7
<i>tRNAThr</i>	9750–9814	+	65	TGT			6
<i>tRNAPro</i>	9815–9880	-	66	TGG			0
<i>ND6</i>	9883–10407	+	527		ATT	TAA	2
<i>CYTB</i>	10407–11543	+	1137		ATG	TAA	-1
<i>tRNASer (UCN)</i>	11542–11607	+	66	TGA			-2
<i>ND1</i>	11628–12572	-	945		ATT	TAA	20
<i>tRNALeu (CUN)</i>	12579–12644	-	66	TAG			6
<i>lrRNA</i>	12645–13972	-	1328				0
<i>tRNAVal</i>	13973–14044	-	72	TAC			0
<i>srRNA</i>	14045–14841	-	797				0
CR	14842–15418		577				0

<https://doi.org/10.1371/journal.pone.0204667.t001>

This result is similar to the universal feature presumed from earlier reported mosquito mt genomes [43–51] in that the CR has the highest A+T content, followed by rRNAs. For the 13 PCGs in the *An. sinensis* mt genome, the third codon position has a higher A+T content (94.24%), followed by the first codon position (69.46%) and the second codon position (66.84%). The results supports the data from other known mosquito mt genomes [43–51] in that the 3rd codon position has the highest A+T content, followed by the 1st codon position and 2nd codon position.

Table 2. Nucleotide composition of the *An. sinensis* mt genome.

Mitogenome region	T content %T	C content %C	A content %A	G content %G	A+T content %A+T	AT Skew	GC Skew
Whole genome	38.14	12.50	40.20	9.15	78.34	0.026	-0.155
Protein-coding genes	43.94	11.04	32.90	12.12	76.85	-0.144	0.047
First codon position	37.72	10.93	31.74	19.61	69.46	-0.086	0.284
Second codon position	46.32	19.21	20.52	13.96	66.84	-0.386	-0.158
Third codon position	47.79	2.97	46.45	2.79	94.24	-0.014	-0.033
Protein-coding genes-J	41.52	13.05	34.08	11.36	75.60	-0.098	-0.069
First codon position	34.21	13.18	32.85	19.76	67.06	-0.020	0.200
Second codon position	44.68	21.55	20.99	12.78	65.66	-0.361	-0.255
Third codon position	45.68	4.41	48.39	1.53	94.07	0.029	-0.485
Protein-coding genes-N	47.79	7.84	31.04	13.32	78.83	-0.212	0.259
First codon position	43.30	7.36	29.98	19.36	73.28	-0.182	0.449
Second codon position	48.92	15.48	19.78	15.82	68.70	-0.424	0.011
Third codon position	51.15	0.69	43.37	4.79	94.52	-0.082	0.747
tRNA genes	39.02	9.21	39.57	12.20	78.59	0.007	0.140
tRNA genes-J	37.99	10.53	40.73	10.76	78.72	0.035	0.011
tRNA genes-N	40.53	7.31	37.87	14.29	78.40	-0.034	0.323
rRNA	43.15	6.54	38.31	12.00	81.46	-0.059	0.294
Control region	49.91	3.99	43.67	2.43	93.58	-0.067	-0.243

<https://doi.org/10.1371/journal.pone.0204667.t002>

AT-skew and GC-skew have also been widely used to measure the nucleotide compositional behaviors of mt genomes [42]. The AT skew and GC skew of the *An. sinensis* mt genome are 0.026 and -0.155, respectively (Table 2). The AT-skew values are positive, and the GC-skew values are negative for all other mosquito mt genomes [43–51], which indicated overall mt genome preference for the use of A and C. The PCGs of the *An. sinensis* mt genome show an overall negative AT-skew (-0.144) and positive GC-skew (0.047). It is a common phenomenon that the PCGs of mosquito mt genomes prefer to use T and G [43–51].

All 13 PCGs in the *An. sinensis* mt genome use ATN as the start codon, except for *COI*, which uses the special start codon TCG, and *ND5*, which uses GTG as the start codon. Use of GTG as a start codon has been documented for mtDNA-encoded proteins in various organisms, including *Anopheles* species [52]. All 13 PCGs use the complete stop codon TAA, except for *COII*, *COIII* and *ND4*, which use the incomplete T as a stop codon. There is no other mosquito species with a mt genome that uses TAG as a stop codon [43–51]. The usage bias of amino acids for the 13 PCGs was identified in the *An. sinensis* mt genome. Leu has the highest percentage (16.05%), followed by Phe (9.67%), Ser (9.30%) and Ile (9.24%), and Cys has the lowest percentage (1.10%) (S1 Fig). This order is similar to other mosquito mt genomes [43–51]. Leucine has an inferred high usage frequency, and as a hydrophobic amino acid, it can be a component of many transmembrane proteins in the mitochondria.

There are a total of 3733 codons in the *An. sinensis* mt genome, excluding termination codons, which is within the codon number range of other insect mt genomes (3585–3746) [53]. For the relative synonymous codon usage (RSCU), UUA is the most used codon (RSCU value 5.28) in the *An. sinensis* mt genome, followed by CGA (3.24), UCA (2.60), GGA (2.50), UCU (2.50), CCU (2.24), GCU (2.14), GUU (2.10) and ACA (2.10), and ACG is the least used codon (0.02) (Fig 2). The third codon position has a higher usage frequency of A (46.45%) or U (47.79%) in the *An. sinensis* mt genome. This phenomenon is consistent with previously reported mosquito mt genomes [43–51].

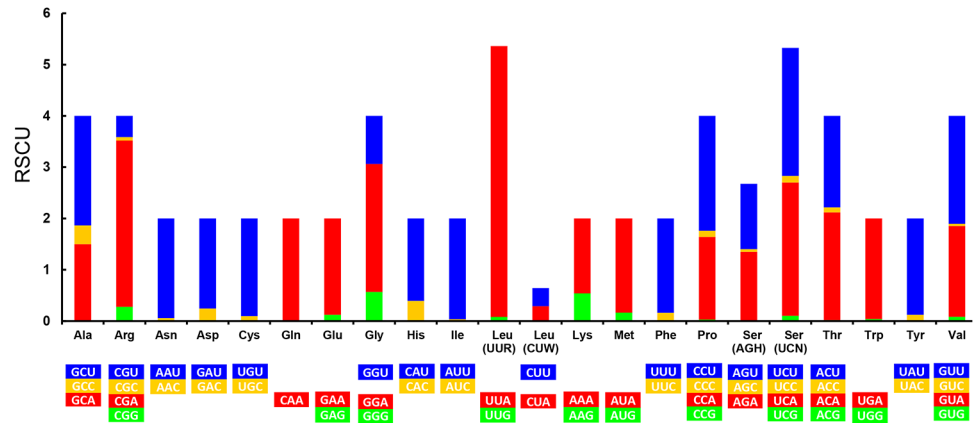


Fig 2. Relative synonymous codon usage (RSCU) in the *An. sinensis* mt genome. The RSCU values are represented on the y-axis, while the codon families for each amino acid are denoted on the x-axis.

<https://doi.org/10.1371/journal.pone.0204667.g002>

All tRNAs of the *An. sinensis* mt genome can be folded into the typical clover-leaf secondary structure, except for *tRNA^{Ser(AGN)}*, which has lost the DHU stem [S2 Fig]. This is a common feature in metazoan mt genomes [42]. Consistent with other known mosquito mt genomes, all tRNA lengths range from 64 to 72 bp. There are 18 mismatches, all as GU base pairs, found in the 12 tRNAs of the *An. sinensis* mt genome (S2 Fig). The large subunit rRNA (*16SrRNA*), 1328 bp long, is located between *ND1* and *tRNA^{Val}*, and the small subunit rRNA (*12SrRNA*), 797 bp long, is between *tRNA^{Val}* and control region, both on the minority strand. The total A+T content of rRNA genes is 81.46%.

The CR plays an important role in the regulation of replication and the transcription of the mt genome [54–55]. The CR region of the *An. sinensis* mt genome is 577 bp long and is located between *12SrRNA* and *tRNA^{Ile}*. The A+T content of this region (93.58%) is higher than other regions of the *An. sinensis* mt genome. There is a poly-T stretch of 18 bp to be identified, which may be a recognition site for the initiation of replication in the mt genome [56]. In addition, there are two 46 bp long tandem repeats found in the CR. The tandem repeat structures in CRs are common, but the length and tandem repeat time vary in the other known mosquito mt genomes [6].

NUMTs and their comparison in mosquitoes

We identified NUMTs of 19 mosquito species with both known mt genomes and nuclear genome sequences. In the subfamily Anophelinae, all 16 species investigated belong to the genus *Anopheles*. Out of the 16, 10 species with genome sizes 141.20–268.44 Mb have no NUMTs, and the remaining 6 species with genome sizes 132.94–217.57 Mb each have 1–3 NUMTs (Table 3). In the subfamily Culicinae, *Cx. quinquefasciatus* (genome size 574.57 Mb), *Ae. aegypti* (1,342.21 Mb) and *Ae. albopictus* (1,868.07 Mb) have 13, 122 and 196 NUMTs, respectively. This suggests that the NUMT number is significantly correlated with genome size ($R = 0.9871, p = 5.71E-15$).

The NUMT lengths in the six Anophelinae species range from 43 bp to 309 bp, while lengths in the three Culicinae species range from 37 bp to 15,580 bp. If lengths are divided into three classes (large-size (>2 kb), medium-size (200 bp to 2 kb) and small-size (< 200 bp), it is seen that the large-sized NUMTs only exist in the three Culicinae species, with the largest NUMT (15,580 bp) existing in *Cx. quinquefasciatus* (S2 Table). This suggests that larger genomes can house longer NUMTs. If we assemble the NUMTs, the numbers of large-sized,

Table 3. Number, total length and density of nuclear mitochondrial segments (NUMTs) in 19 mosquito species of nuclear genomes.

Genus/Subgenus/Series	Species	Sizes of mitogenome (kb)/nuclear genome (Mb)	Number/total length (bp) of NUMTs	NUMT density (bp/Mb nuclear genome)
Anophelinae				
<i>Anopheles/Cellia/Neomyzomyia</i>	<i>An. farauti</i> ^a	15.412/175.52	2/208	1.19
	<i>An. dirus</i>	15.404/209.79	0/0	0.00
<i>Anopheles/Cellia/Neocellia</i>	<i>An. stephensi</i>	15.387/216.26	0/0	0.00
	<i>An. maculatus</i> ^b	14.850/141.20	0/0	0.00
<i>Anopheles/Cellia/Myzomyia</i>	<i>An. minimus</i>	15.395/195.70	1/43	0.22
	<i>An. culicifacies</i>	15.364/198.03	0/0	0.00
<i>Anopheles/Cellia/Pyrethophorus</i>	<i>An. christyi</i> ^b	14.967/169.04	1/309	1.83
	<i>An. epiroticus</i>	15.379/216.83	0/0	0.00
	<i>An. melas</i>	15.366/222.01	0/0	0.00
	<i>An. merus</i>	15.365/244.34	0/0	0.00
	<i>An. coluzzii</i>	15.441/218.22	0/0	0.00
	<i>An. arabiensis</i>	15.369/239.13	0/0	0.00
<i>Anopheles/Anopheles/Myzorhynchus</i>	<i>An. gambiae</i>	15.363/268.44	0/0	0.00
<i>Anopheles/Anopheles/Anopheles</i>	<i>An. sinensis</i>	15.418/194.49	1/216	1.11
<i>Anopheles/Anopheles/Anopheles</i>	<i>An. atroparvus</i>	15.458/217.57	3/170	0.78
<i>Anopheles/Nyssorhynchus/Argyritarsis</i>	<i>An. darlingi</i>	15.386/132.94	1/122	0.92
Culicinae				
<i>Culex/Culex/</i>	<i>Cx. quinquefasciatus</i>	15.587/574.57	13/28,431	51.92
<i>Aedes/Stegomyia/</i>	<i>Ae. aegypti</i>	16.655/1,342.21	122/92,934	69.24
	<i>Ae. albopictus</i>	16.665/1,868.07	196/60,322	32.29

^a Lack of partial control region.

^b Lack of control region.

<https://doi.org/10.1371/journal.pone.0204667.t003>

medium-sized and small-sized NUMTs are 12, 157 and 171, respectively. This suggests that half of the NUMTs are small-sized (< 200 bp). Most of the longer NUMTs would be disruptive to shorter NUMTs through nucleotide deletions or insertions in the evolutionary process, which is the probable reason why half of NUMTs are small [26,27].

In the 16 species of Anophelinae, the total length of NUMTs in each species ranges from 0 bp to 309 bp, and the density ranges from 0 bp/Mb to 1.83 bp/Mb of the nuclear genome. The average total length and density are 66.75 bp and 0.38 bp/Mb, respectively (Table 3 and Fig 3). In the three species of Culicinae, the total length of NUMTs in each species range from 28,431 bp to 92,934 bp, and the density range from 32.29 bp/Mb to 69.24 bp/Mb, with the average total length and density being 60,562.33 bp and 51.15 bp/Mb, respectively (Table 3 and Fig 3). The Culicinae species have greater length and density of NUMTs than the Anophelinae species. Statistical analyses showed that the total length of NUMTs is significantly correlated with genome size ($R = 0.9104$, $p = 6.31E-08$) in all 16 species investigated, which is consistent with the NUMT investigation of 85 species [57]. In addition, the density of NUMTs is significantly correlated with genome size in these mosquitoes ($R = 0.7667$, $p = 1.29E-04$). The total length and density of NUMTs are also significantly correlated ($R = 0.9219$, $p = 2.04E-08$). These results suggest that NUMTs could contribute to the expansion of genome size.

In the genetics of other insects, the total length and density of NUMTs in the *Drosophila melanogaster* (genome ~170 Mb), *Tribolium castaneum* (~150 Mb) and *Apis mellifera* (~230

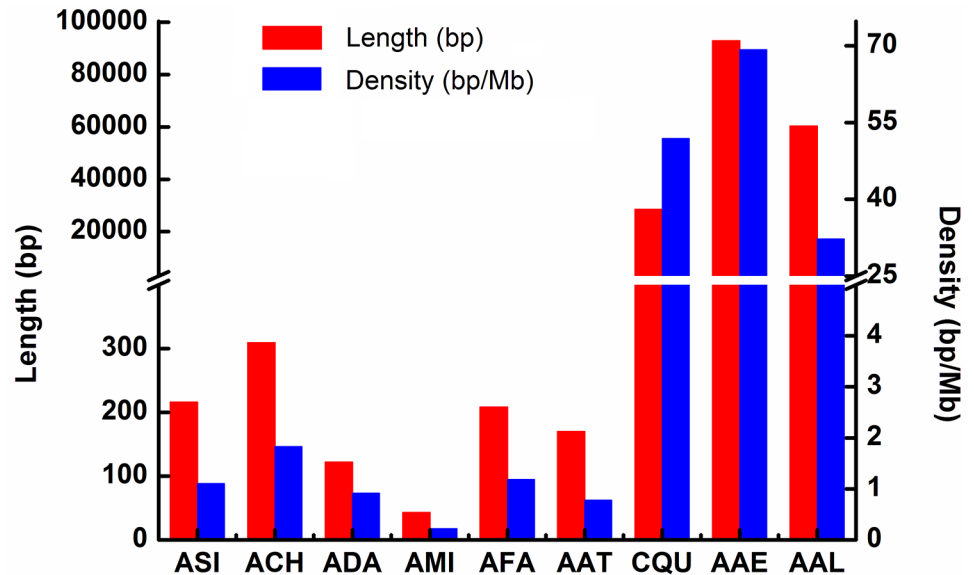


Fig 3. Length (bp) and density (bp/Mb) of nuclear mitochondrial DNA segments (NUMTs) in nine mosquito species. ASI: *An. sinensis*, ACH: *An. christyi*, ADA: *An. darlingi*, AFA: *Anopheles farauti*, AMI: *An. minimus*, AAT: *An. atroparvus*, CQU: *Cx. quinquefasciatus*, AAE: *Ae. aegypti*, AAL: *Ae. albopictus*.

<https://doi.org/10.1371/journal.pone.0204667.g003>

Mb) are 777 bp and 4.57 bp/Mb [58], 8,821 bp and 58.81 bp/Mb [40], and 275,022 bp and 1195.75 bp/Mb [40], respectively. The genome sizes in these three species are comparable with the Anophelinae species; however, the total length and density of NUMTs of the three species are larger than those in the Anophelinae species, especially for *T. castaneum* and *A. mellifera*. The genome size expansion due to mitochondrial insertion segments is variable in different insect groups. The *A. mellifera* nuclear genome recombination rate is much greater than that of *D. melanogaster* [59], suggesting that the NUMT total length and density in the nuclear genome is related to the genome recombination rate.

The largest number of NUMTs in the genomes of the nine mosquito species originated from the *COI* (39 NUMTs) gene, followed by the *16SrRNA* (31), *CytB* (27) and *ND5* (22) genes. The least number of NUMTs are from the CR (6) (Fig 4). In the six Anophelinae species, NUMTs were derived from only seven mitochondrial genes, including three *COII* in *An. atroparvus*, two *COIII* in *An. farauti*, one *ND4* in *An. sinensis*, one *ND5* in *An. christyi*, one *12SrRNA* in *An. minimus* and one *tRNA^{Cys}-tRNA^{Tyr}* in *An. darlingi*. In the three Culicinae species, NUMTs in *Ae. aegypti* and *Cx. quinquefasciatus* cover the whole mt genome. Similarly, NUMTs in *Ae. albopictus* also cover the whole mt genome, with the exception of *ND4*, *12SrRNA* and CR. These data suggest that the PCGs are transferred to the nuclear genome at a higher frequency, and the larger-sized nuclear genomes in mosquitoes have larger-sized mt genome sequences.

The largest numbers of NUMTs were derived from the *16SrRNA* gene in three mammals, *Sus scrofa* (6), *Pan troglodytes* (65) and *Homo sapiens* (53). In *Mus musculus*, the largest number of NUMTs originated from CR (11), followed by *ND2* (7) and *ND4* (6), but no NUMT was found from *tRNA^{Val}* [60]. This suggests that the NUMT origination may be different in mosquitoes and mammals. The genome sizes of *M. musculus*, *S. scrofa*, *P. troglodytes* and *H. sapiens* are 2.7 G, 2.3 G, 2.9 G and 2.9 G, and the mt genome coverage rates of NUMT are 84.7%, 32.6%, 100% and 100%, respectively [60]. The larger-sized nuclear genomes in mammals appear to have a wider range of mt genome sequence coverage rates.

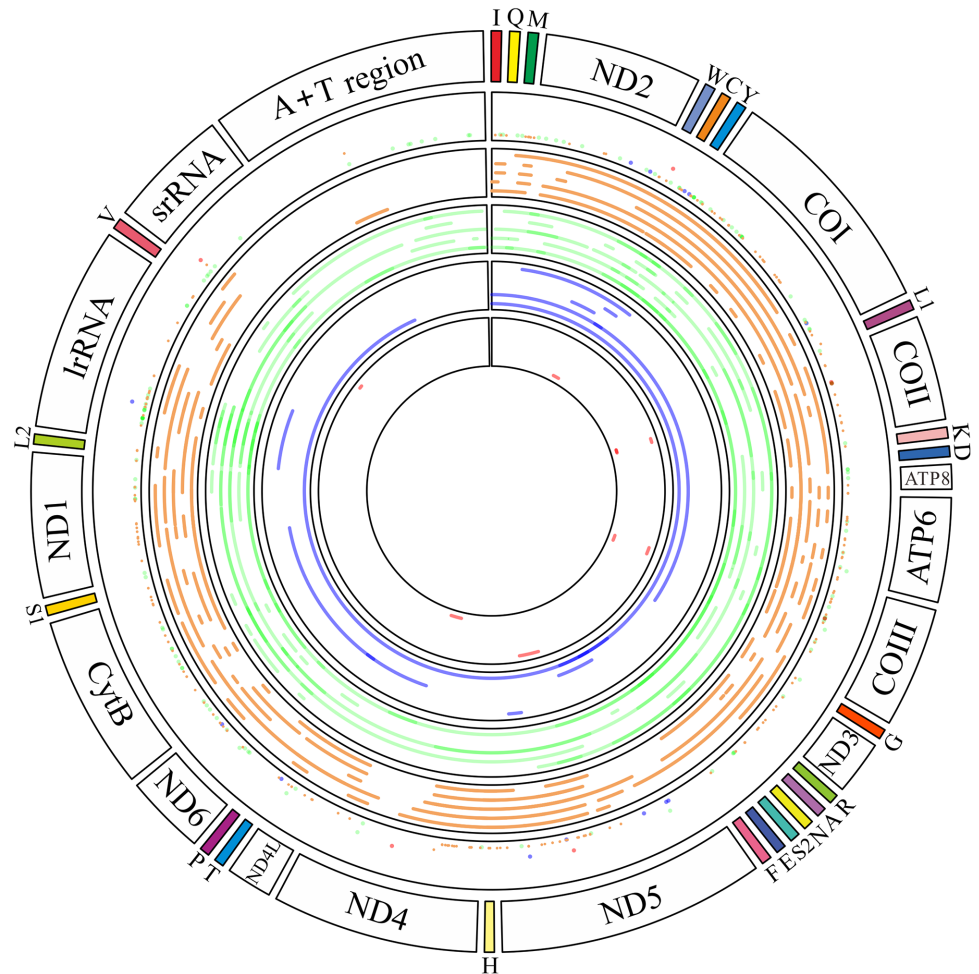


Fig 4. NUMT distribution position in nine mosquito species of mt genomes. The outermost cycle denotes the structure of the reference mt genome, and the second cycle from outside marks the starting positions of the NUMTs for *Ae. albopictus* (orange), *Ae. aegypti* (green), *Cx. quinquefasciatus* (blue) and six *Anopheles* species as a group (red). The four cycles entad show the NUMT regions of the three and six *Anopheles* species as a group, respectively.

<https://doi.org/10.1371/journal.pone.0204667.g004>

Conclusion

We studied mitochondrial genes in *An. sinensis* through analysis of the complete mt genome sequence. NUMT analysis of nineteen mosquito species led to the conclusion that the number, total length and density of NUMTs are significantly correlated with genome size. NUMTs are an important cause of nuclear genome size expansion in mosquitoes. The genome size expansion due to mitochondrial insertion segments is variable in different insect groups. PCGs are transferred to the nuclear genome at a higher frequency in mosquitoes, but the NUMT origination is quite different from mammals. Larger-sized nuclear genomes, in both mosquitoes and mammals, have a wider range of transferred mt genome sequences.

Supporting information

S1 Fig. The percentage of each amino acid in the *An. sinensis* mt genome. N-strand: majority strand; J-strand: minority strand.

(TIF)

S2 Fig. Inferred secondary structures of tRNAs in the *An. sinensis* mt genome. The tRNAs are labeled with their corresponding amino acids.

(TIF)

S1 Table. Mt genome and nuclear genome sequence information of 19 mosquito species.

(DOC)

S2 Table. Positions and lengths of NUMTs in the nuclear genomes of nine species.

(DOC)

Author Contributions

Data curation: Bo Li, Yu-Juan Zhang, Qi-Meng Mao.

Formal analysis: Yi-Ran Ding, Bo Li.

Investigation: Yi-Ran Ding.

Methodology: Yi-Ran Ding, Bin Chen.

Resources: Bin Chen.

Supervision: Bin Chen.

Writing – original draft: Yi-Ran Ding, Bin Chen.

Writing – review & editing: Bin Chen.

References

1. Chen ZT, Du YZ. First Mitochondrial Genome from Nemouridae (Plecoptera) Reveals Novel Features of the Elongated Control Region and Phylogenetic Implications. *Int J Mol Sci*. 2017; 18(5):996. <https://doi.org/10.3390/ijms18050996> PMID: 28475163
2. Zhang NX, Yu G, Li TJ, He QY, Zhou Y, Si FL, et al. The Complete Mitochondrial Genome of *Delia anti-qua* and Its Implications in Dipteran Phylogenetics. *Plos One*. 2015; 10(10): e0139736. <https://doi.org/10.1371/journal.pone.0139736> PMID: 26427045
3. Li X, Huang Y, Lei F. Comparative mitochondrial genomics and phylogenetic relationships of the Cross-optilon species (Phasianidae, Galliformes). *BMC Genomics*. 2015; 16(1):1–12. <https://doi.org/10.1186/s12864-015-1234-9> PMID: 25652939
4. Cameron SL. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. *Annu Rev Entomol*. 2014; 59(1):95–117. <https://doi.org/10.1146/annurev-ento-011613-162007> PMID: 24160435
5. Foster PG, de Oliveira TMP, Bergo ES, Conn JE, Sant'Ana DC, Nagaki SS, et al. Phylogeny of Anophelinae using mitochondrial protein coding genes. *R Soc Open Sci*. 2017; 4: 170758. <https://doi.org/10.1098/rsos.170758> PMID: 29291068
6. 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 mitochondriomics of 50 mosquitoes. *Sci Rep*. 2017; 7(1):7666. <https://doi.org/10.1038/s41598-017-07977-0> PMID: 28794438
7. David JP, Coissac E, Melodelima C, Poupardin R, Riaz MA, Chandor-Proust A, et al. Transcriptome response to pollutants and insecticides in the dengue vector *Aedes aegypti* using next-generation sequencing technology. *BMC Genomics*. 2010; 11:216. <https://doi.org/10.1186/1471-2164-11-216> PMID: 20356352
8. Chen B, Zhang YJ, He ZB, Li WS, Si FL, Tang Y, et al. De novo transcriptome sequencing and sequence analysis of the malaria vector *Anopheles sinensis* (Diptera: Culicidae). *Parasites & Vectors*. 2014; 7(1):314. <https://doi.org/10.1186/1756-3305-7-314> PMID: 25000941
9. Chen K, Wang Y, Li XY, Peng H, Ma YJ. Sequencing and analysis of the complete mitochondrial genome in *Anopheles sinensis* (Diptera: Culicidae). *Infect Dis Poverty*. 2017; 6(1):149. <https://doi.org/10.1186/s40249-017-0362-7> PMID: 28969698
10. Wu XM, Xu BY, Si FL, Li JY, Yan ZT, Yan ZW, et al. Identification of carboxylesterase genes associated with pyrethroid resistance in the malaria vector *Anopheles sinensis* (Diptera: Culicidae). *Pest Manag Sci*. 2018; 74: 159–169. <https://doi.org/10.1002/ps.4672> PMID: 28731595

11. Yan ZW, He ZB, Yan ZT, Si FL, Zhou Y, Chen B. Genome-wide and expression-profiling analyses suggest the main cytochrome P450 genes related to pyrethroid resistance in the malaria vector *Anopheles sinensis* (Diptera Culicidae). *Pest Manag Sci*. 2018; 74: 159–169.
12. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ, Lopez JV, et al. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J Mol Evol*. 1994; 39(2):174–190. PMID: [7932781](#)
13. Thorsness PE, Weber ER. Escape and Migration of Nucleic Acids between Chloroplasts, Mitochondria, and the Nucleus. *Int Rev Cytol*. 1996; 165:207–234. [https://doi.org/10.1016/S0074-7696\(08\)62223-8](https://doi.org/10.1016/S0074-7696(08)62223-8) PMID: [8900960](#)
14. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet*. 2004; 5(2):123–135. <https://doi.org/10.1038/nrg1271> PMID: [14735123](#)
15. Kuyl AC, Kuiken CL, Dekker JT, Perizonius WRK, Goudsmit J. Nuclear counterparts of the cytoplasmic mitochondrial 12S rRNA gene: a problem of ancient dna and molecular phylogenies. *J Mol Evol*. 1995; 40(6):652–657. <https://doi.org/10.1007/BF00160513> PMID: [7543951](#)
16. Zhang DX, Hewitt GM. Nuclear integrations: challenges for mitochondrial dna markers. *Trends Ecol Evol*, 1996; 11(6):247–251. [https://doi.org/10.1016/0169-5347\(96\)10031-8](https://doi.org/10.1016/0169-5347(96)10031-8) PMID: [21237827](#)
17. Song H, Buhay JE, Whiting MF, Crandall KA. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *P Natl Acad Sci USA*, 2008; 105(36):13486–13491. <https://doi.org/10.1073/pnas.0803076105> PMID: [18757756](#)
18. Zhang DX, Hewitt GM. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol*, 2010; 12(3):563–584. <https://doi.org/10.1046/j.1365-294X.2003.01773.x>
19. Leister D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet*, 2005; 21(12):655–663. <https://doi.org/10.1016/j.tig.2005.09.004> PMID: [16216380](#)
20. Sorenson MD, Quinn TW. Numts: a challenge for avian systematics and population biology. *Auk*, 1998; 115(1):214–221. <https://doi.org/10.2307/4089130>
21. Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. *Mol Bio Evol*. 2004; 21(6):1081–1084. <https://doi.org/10.1093/molbev/msh110> PMID: [15014143](#)
22. Hazkani-Covo E. Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. *Mol Bio Evol*, 2009; 26(10):2175–2179. <https://doi.org/10.1093/molbev/msp131> PMID: [19578158](#)
23. Ullrich H, Lättig K, Brennicke A, Knoop V. Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 arabidopsis thaliana ecotypes. *Plant Mol Bio*, 1997; 33(1):37–45. <https://doi.org/10.1023/A:1005720910028>
24. Martins J, Solomon SE, Mikheyev AS, Mueller UG, Ortiz A, Bacci M Jr. Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini). *Insect Mol Bio*. 2007; 16(6):777–784. <https://doi.org/10.1111/j.1365-2583.2007.00771.x> PMID: [18093006](#)
25. Bensasson D, Zhang DX, Hewitt GM. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol Bio Evol*. 2000; 17(3):406–415. <https://doi.org/10.1093/oxfordjournals.molbev.a026320> PMID: [10723741](#)
26. Sunnucks P, Hales DF. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol Bio Evol*. 1996; 13(3):510–524. <https://doi.org/10.1093/oxfordjournals.molbev.a025612> PMID: [8742640](#)
27. 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 Bio*. 2011; 41(10):770–777. <https://doi.org/10.1016/j.ibmb.2011.05.006> PMID: [21640823](#)
28. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*. 2002; 298(5591):129–149. <https://doi.org/10.1126/science.1076181> PMID: [12364791](#)
29. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, et al. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science*. 2010; 330(6000): 86–88. <https://doi.org/10.1126/science.1191864> PMID: [20929810](#)
30. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*. 2007; 316(5832): 1718–1723. <https://doi.org/10.1126/science.1138878> PMID: [17510324](#)
31. Chen XG, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. Genome sequence of the Asian Tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. *Proc Natl Acad Sci*. 2015; 112(44): 5907–5915. <https://doi.org/10.1073/pnas.1516410112> PMID: [26483478](#)

32. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science*. 2015; 347(6217): 1258522. <https://doi.org/10.1126/science.1258522> PMID: 25554792
33. Zou YL, Ding YR, Luo QC, Chen B. The extraction method of mosquito mitogenome. *Chinese Journal of Vector Biology and Control*. 2015; 26(4), 333–336.
34. Zhang NX, Zhang YJ, Yu G. Structure characteristics of the mitochondrial genomes of Diptera and design and application of universal primers for their sequencing. *Acta Entomologica Sinica*. 2013; 56(4), 398–407.
35. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 1997; 25(24): 4876–4882. <https://doi.org/10.1093/nar/25.24.4876> PMID: 9396791
36. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology & Evolution*. 2013; 30(4):2725–2729. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
37. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997; 25(5):955–964. PMID: 9023104
38. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999; 27(2):573–580. <https://doi.org/10.1093/nar/27.2.573> PMID: 9862982
39. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol*. 1995; 41(3):353–8. <https://doi.org/10.1007/BF01215182> PMID: 7563121
40. Pamilo P, Viljakainen L, Vihavainen A. Exceptionally high density of NUMTs in the honeybee genome. *Mol Biol Evol*. 2007; 24(6):1340–1346. <https://doi.org/10.1093/molbev/msm055> PMID: 17383971
41. Mitchell SE, Cockburn AF, Seawright JA. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome*. 1993; 36(36):1058–73. <https://doi.org/10.1139/g93-141>
42. Negrisoló E, Babbucci M, Patarnello T. The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondrial genomics of neuropterid insects. *BMC Genomics*. 2011; 12(1):221. <https://doi.org/10.1186/1471-2164-12-221> PMID: 21569260
43. Krzywinski J, Grushko OG, 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
44. 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(1):1–11. <https://doi.org/10.1186/1475-2875-12-64> PMID: 23405960
45. Hua YQ, Ding YR, Yan ZT, Si FL, Luo QC, Chen B. The complete mitochondrial genome of *Anopheles minimus* (Diptera: Culicidae) and the phylogenetics of known *Anopheles* mitogenomes. *Insect Sci*. 2016; 23, 353–365. <https://doi.org/10.1111/1744-7917.12326> PMID: 26852698
46. 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(1):1–13. <https://doi.org/10.1186/1475-2875-9-127> PMID: 20470395
47. Hua YQ, Yan ZT, Fu WB, He QY, Zhou Y, Chen B, et al. Sequencing and analysis of the complete mitochondrial genome in *Anopheles culicifacies* species B (Diptera: Culicidae). *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016; 27, 2909–2910. <https://doi.org/10.3109/19401736.2015.1060434> PMID: 26114319
48. Beard CB, Hamm DM, Collins FH. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol Bio*. 1993; 2(2):103–124. <https://doi.org/10.1111/j.1365-2583.1993.tb00131.x>
49. Zhang HD, Xing D, Wang G, Li CX, Zhao TY. Sequencing and analysis of the complete mitochondrial genome of *Aedes albopictus* (Diptera: Culicidae) in China. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016; 27, 2787–2788. <https://doi.org/10.3109/19401736.2015.1053067> PMID: 26114325
50. Demarisilva B, Foster PG, de Oliveira TM, Bergo ES, Sanabani SS, Pessôa R, et al. Mitochondrial genomes and comparative analyses of *Culex camposi*, *Culex coronator*, *Culex usquatus* and *Culex usquatissimus* (Diptera: Culicidae), members of the coronator group. *BMC Genomics*. 2015; 16, 831. <https://doi.org/10.1186/s12864-015-1951-0> PMID: 26489754
51. Hardy CM, Court LN, Morgan MJ. The complete mitochondrial DNA genome of *Aedes vigilax* (Diptera: Culicidae)[J]. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016; 27(4):2552–2553. <https://doi.org/10.3109/19401736.2015.1038800> PMID: 26099979

52. Wolstenholme DR. Genetic novelties in mitochondrial genomes of multicellular animals. *Curr Opin Genet Dev.* 1992; 2(6):918–925. [https://doi.org/10.1016/S0959-437X\(05\)80116-9](https://doi.org/10.1016/S0959-437X(05)80116-9) PMID: 1282405
53. Cha SY, Yoon HJ, Lee EM, Yoon MH, Hwang JS, Jin BR, et al. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene.* 2007; 392(1–2):206–220. <https://doi.org/10.1016/j.gene.2006.12.031> PMID: 17321076
54. Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta.* 1999; 1410(2):103–123. [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3) PMID: 10076021
55. Saito S, Tamura K, Aotsuka T. Replication origin of mitochondrial DNA in insects. *Genetics.* 2005; 171(4):1695–1705. <https://doi.org/10.1534/genetics.105.046243> PMID: 16118189
56. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999; 23(2):147. <https://doi.org/10.1038/13779> PMID: 10508508
57. Einat HC, Zeller RM, William M. Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *Plos Genetics.* 2010; 6(2):e1000834. <https://doi.org/10.1371/journal.pgen.1000834> PMID: 20168995
58. Bensasson D, Zhang D, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol.* 2001; 16(6):314–321. [https://doi.org/10.1016/S0169-5347\(01\)02151-6](https://doi.org/10.1016/S0169-5347(01)02151-6) PMID: 11369110
59. Beye M, Gattermeier I, Hasselmann M, Gempe T, Schioett M, Baines JF, et al. Exceptionally high levels of recombination across the honey bee genome. *Genome Res.* 2015; 16(11):1339–1344. <https://doi.org/10.1101/gr.5680406> PMID: 17065604
60. Qu HY, Ma F, Li QW. Comparative analysis of mitochondrial fragments transferred to the nucleus in vertebrate. *J Genet Genomics.* 2008; 35(8):485–490. [https://doi.org/10.1016/S1673-8527\(08\)60066-1](https://doi.org/10.1016/S1673-8527(08)60066-1) PMID: 18721785