



OPEN

## Mitochondrial genome sequencing and phylogeny of *Haemagogus albomaculatus*, *Haemagogus leucocelaenus*, *Haemagogus spegazzinii*, and *Haemagogus tropicalis* (Diptera: Culicidae)

Fábio Silva da Silva<sup>1,2</sup>, Ana Cecília Ribeiro Cruz<sup>2</sup>, Daniele Barbosa de Almeida Medeiros<sup>2</sup>, Sandro Patroca da Silva<sup>2</sup>, Márcio Roberto Teixeira Nunes<sup>3</sup>, Lívia Carício Martins<sup>2</sup>, Jannifer Oliveira Chiang<sup>2</sup>, Poliana da Silva Lemos<sup>3</sup>, Gabriel Muricy Cunha<sup>4</sup>, Renato Freitas de Araujo<sup>4</sup>, Hamilton Antônio de Oliveira Monteiro<sup>2</sup> & Joaquim Pinto Nunes Neto<sup>2</sup>✉

The genus *Haemagogus* (Diptera: Culicidae) comprises species of great epidemiological relevance, involved in transmission cycles of the *Yellow fever virus* and other arboviruses in South America. So far, only *Haemagogus janthinomys* has complete mitochondrial sequences available. Given the unavailability of information related to aspects of the evolutionary biology and molecular taxonomy of this genus, we report here, the first sequencing of the mitogenomes of *Haemagogus albomaculatus*, *Haemagogus leucocelaenus*, *Haemagogus spegazzinii*, and *Haemagogus tropicalis*. The mitogenomes showed an average length of 15,038 bp, average AT content of 79.3%, positive AT-skews, negative GC-skews, and comprised 37 functional subunits (13 PCGs, 22 tRNA, and 02 rRNA). The PCGs showed ATN as start codon, TAA as stop codon, and signs of purifying selection. The tRNAs had the typical leaf clover structure, except *tRNA<sup>Ser1</sup>*. Phylogenetic analyzes of Bayesian inference and Maximum Likelihood, based on concatenated sequences from all 13 PCGs, produced identical topologies and strongly supported the monophyletic relationship between the *Haemagogus* and *Conopostegus* subgenera, and corroborated with the known taxonomic classification of the evaluated taxa, based on external morphological aspects. The information produced on the mitogenomes of the *Haemagogus* species evaluated here may be useful in carrying out future taxonomic and evolutionary studies of the genus.

The genus *Haemagogus* Williston, 1896 (Diptera: Culicidae), belongs to the Aedini tribe and comprises 28 species of mosquitoes with occurrence records restricted to the Neotropical region, and widely distributed throughout Central and South America<sup>1</sup>. Currently, it includes two subgenera: *Conopostegus* Dyar, 1925, with four species, and *Haemagogus* Williston, 1896, with 24 species<sup>1–3</sup>. The species belonging to the latter subgeneric category are subdivided into three sections: *Albomaculatus* (14 species), *Splendens* (09 species), and *Tropicalis* (01 species)<sup>1</sup>. In Brazil, the occurrence of nine species of *Haemagogus* has been reported, with representatives for both subgenera and Sects. <sup>4</sup>. They are wild mosquitoes, with daytime habits, not commonly found outside the forest environment<sup>2</sup>. However, some species had shown a great adaptive capacity, due to environmental degradation conditions, invading houses, especially those located near areas of altered forest<sup>5,6</sup>.

<sup>1</sup>Graduate Program in Parasitology Biology in the Amazon, Center of Biological and Health Sciences, State University of Pará, Belém 66095-662, Brazil. <sup>2</sup>Arbovirology and Hemorrhagic Fevers Section, Evandro Chagas Institute – IEC/MS/SVS, Ananindeua 67030-000, Brazil. <sup>3</sup>Center for Technological Innovation, Evandro Chagas Institute – IEC/MS/SVS, Ananindeua 67030-000, Brazil. <sup>4</sup>Directorate of Epidemiological Surveillance, Bahia State Department of Health, Salvador 40279-700, Brazil. ✉email: joaquimneto@iec.gov.br

*Haemagogus* has great medical-epidemiological relevance, due to the involvement of some species in the transmission and maintenance of important arboviruses. In South America, *Haemagogus janthinomys* Dyar, 1925 is the main vector of the *Yellow Fever virus* (YFV) in the wild cycle and has a widespread distribution in the Brazilian territory<sup>4,7,8</sup>. Moreover, it is also the main vector for the *Mayaro virus* (MAYV), an alphavirus endemic in the Amazon rainforest<sup>9,10</sup>. *Haemagogus leucocelaenus* Dyar & Shannon, 1924 is also an important vector in the transmission of YFV, mainly in the South and Southeast of Brazil<sup>11,12</sup>. Other species are also related to secondary YFV transmission in the country, such as *Haemagogus albomaculatus* Theobald, 1903<sup>13,14</sup>, *Haemagogus spegazzinii* Br ethes, 1912<sup>15</sup>, and *Haemagogus tropicalis* Cerqueira & Antunes, 1938<sup>16</sup>. This latter species is restricted to the Brazilian Amazon region<sup>1,17,18</sup>.

The use of genomic sequencing tools in studies of evolutionary and molecular taxonomy, in particular, of arthropods of epidemiological importance, has significantly contributed to the provision of critical information useful for vector research and control strategies, allowing not only an improved visualization of the evolutionary and phylogenetic patterns of these organisms, but also contributing directly to the subsidy regarding co-evolutionary properties between these and the infectious agents that they can harbor and potentially transmit<sup>19–25</sup>.

The mitochondrial genome (mitogenome) in most metazoan organisms is a small double-stranded DNA molecule, ranging from 15 to 20 kb in length and contain 37 functional subunits, of which 13 are protein-coding genes (PCGs), 22 are transfer RNA (tRNA), and two are ribosomal RNA (*rRNA*<sup>12S</sup> and *rRNA*<sup>16S</sup>). Besides, it also has a region rich in Adenine and Thymine (A + T), associated with the control of the molecule's replicative processes<sup>26</sup>. Due to factors such as the absence of recombinant processes and the high rate of evolution and accumulation of mutations concerning the nuclear genome, the mitogenome stands out as one of the main markers used for conducting evolutionary and taxonomic studies of vertebrates and invertebrates<sup>27–30</sup>. The use of mitochondrial sequences in evolutionary studies, in particular of insects, has produced significant results in Diptera<sup>31–33</sup>, Hemiptera<sup>34</sup>, Hymenoptera<sup>35</sup>, and Lepidoptera<sup>36</sup>.

Despite the great medical-epidemiological importance of *Haemagogus*, there is still a great unavailability of data and information on the evolutionary biology and molecular taxonomy aspects of this genus. The only species of this genus, so far, that has a complete sequencing data for the mitogenome is *Hg. janthinomys*<sup>37</sup>. In addition, only partial sequences of the *rRNA*<sup>16S</sup> and *COI* genes of *Hg. janthinomys*, *Hg. leucocelaenus*, *Hg. spegazzinii*, *Haemagogus lucifer* Howard, Dyar & Knab, 1913, and *Haemagogus equinus* Theobald, 1903 are available on the Genbank database.

Considering the great potential of mitochondrial genes in conducting taxonomic and phylogenetic studies, we performed, for the first time, the characterization of mitogenomes, using the High Throughput Sequencing (HTS) of the species *Hg. albomaculatus*, *Hg. leucocelaenus*, *Hg. spegazzinii*, and *Hg. tropicalis*. The four mitogenomes obtained were analyzed, compared to those of other species of the Culicidae family available on GenBank database, and the phylogeny was reconstructed based on the concatenation of all 13 PCGs by Bayesian inference and Maximum Likelihood methods.

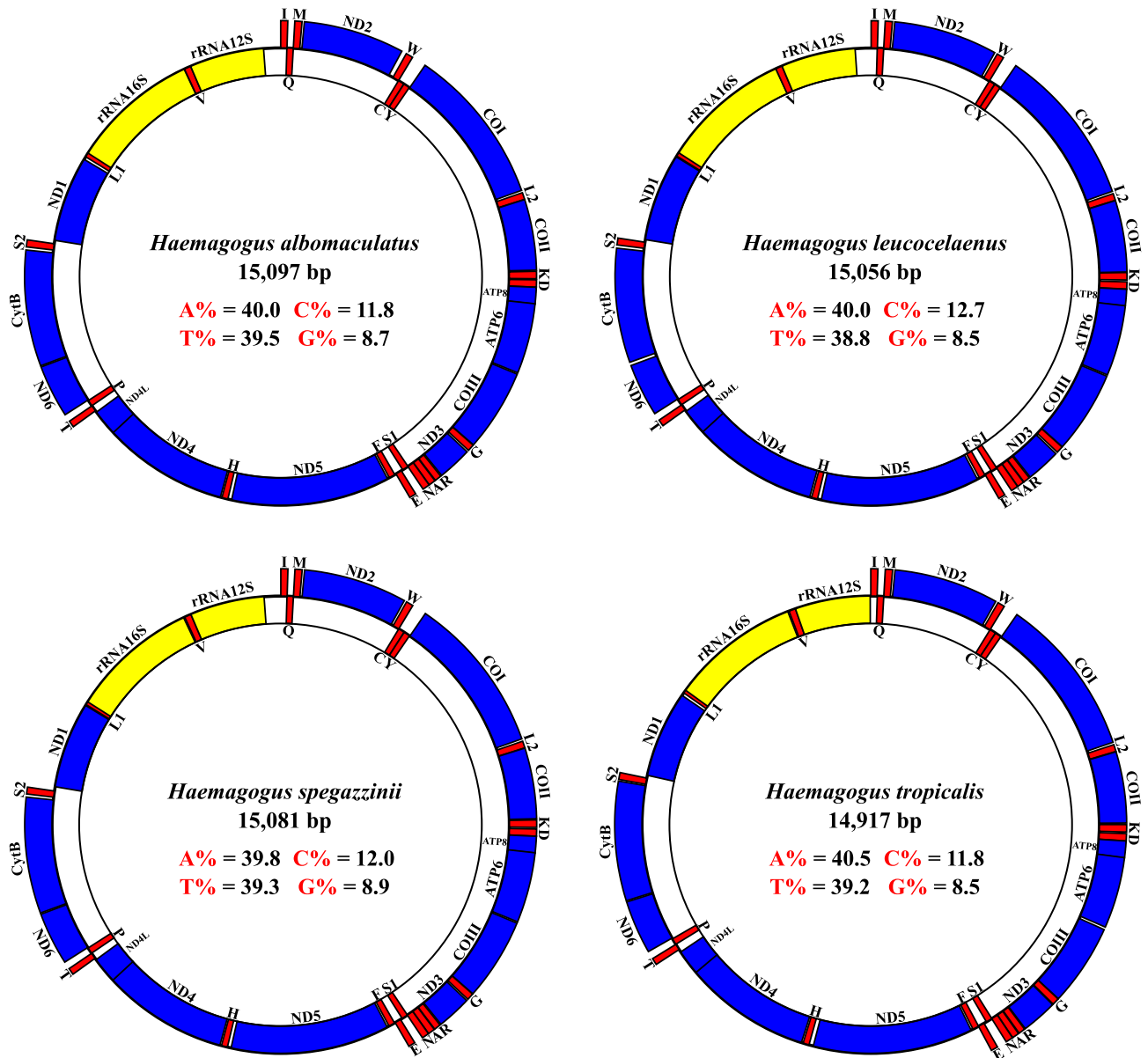
## Results and discussion

**Organization and composition of the obtained mitogenomes.** The *Haemagogus* mitogenomes obtained (Supplementary Tables 3–6) were structured into compact circular double-stranded DNA molecules, with lengths ranging from 14,917 bp (*Hg. tropicalis*) to 15,097 bp (*Hg. albomaculatus*), and with total AT contents ranging from 78.8% (*Hg. leucocelaenus*) to 79.7% (*Hg. tropicalis*). They also comprised 37 functional and conserved subunits—13 PCGs, 22 tRNAs, and 02 rRNAs—and demonstrated similar patterns of gene positioning along the J (Forward) and N (Reverse) strands (Fig. 1), similarly to that observed in studies of other species of Culicidae<sup>32,37–45</sup>, and other groups of insects<sup>31,46–48</sup> with reported mitogenomes.

Asymmetry analyzes of the total genome of each species evaluated resulted in positive AT-skews (average of 0.0108) and negative GC-skews (average of  $-0.1659$ ) (Supplementary Table 7A), showing the content of Adenine and Cytosine in greater proportions in the majority chain compared to the content of Thymine and Guanine, respectively. Those findings corroborate the pattern of nucleotide composition observed in other groups of insects<sup>31,46,49</sup>, and were similar to previous studies for *Anopheles*<sup>32,44,50,51</sup>, *Culex*, and *Lutzia*<sup>39,45</sup>.

In PCGs, the total length ranged from 10,917 bp (*Hg. albomaculatus*) to 10,950 bp (*Hg. spegazzinii*), and the AT content varied from 77.4% (*Hg. leucocelaenus*) to 78.4% (*Hg. tropicalis*) (Supplementary Table 7A). Similar to that reported for other species of Culicidae<sup>39,40,52</sup>, in this study, the AT content increased considerably at the position of the third codon in these regions, ranging from 93.5% (*Hg. leucocelaenus*) to 96.1% (*Hg. albomaculatus*) (Supplementary Table 7B). This increase in the AT content may be related to events where there is a high availability of cellular ATP, parallel to a decrease in other NTPs, thus promoting an increase in the efficiency of the transcription due to the maximization, mainly of Adenine, in the position of the third codon<sup>53</sup>. The asymmetry analysis of these concatenated regions resulted in negative AT-skews (average of  $-0.1521$ ) and positive GC-skews (average of 0.0491) (Supplementary Table 7A). The higher proportions of Thymine and Guanine in relation to Adenine and Cytosine in these regions contrasted with the values of asymmetry obtained in the evaluation of complete sequences, similar to that reported in other studies of Culicidae<sup>40,44,54</sup>. Individually assessed, PCGs presented average AT contents ranging from 71.4% (*COI*) to 86.1% (*ND6*) (Fig. 2a). Also, all PCGs had negative AT-skews (Fig. 2b) and, in most cases, negative GC-skews, except *COIII* and *ND3* (only in *Hg. spegazzinii*), *ND5*, *ND4*, *ND4L*, and *ND1* (Fig. 2c), as observed in other studies<sup>44,52</sup>.

The mitogenomes obtained here presented sets of 22 tRNAs, with approximate values in total lengths (average of 1488 bp), AT contents (average of 80.2%), and positive AT/GC-skews (Supplementary Table 7A), similar to that reported for other species of Culicidae<sup>40,44,54,55</sup>. Individually, the average AT content in these regions ranged from 69.7% (*tRNA*<sup>Arg</sup>) to 91.3% (*tRNA*<sup>Glu</sup>) (Fig. 2a), and the average lengths ranged from 64 bp (*tRNA*<sup>Arg</sup>) to 72 bp (*tRNA*<sup>Val</sup>) (Supplementary Tables 3–6). The typical cloverleaf-like structure, including four arms [Amino acid

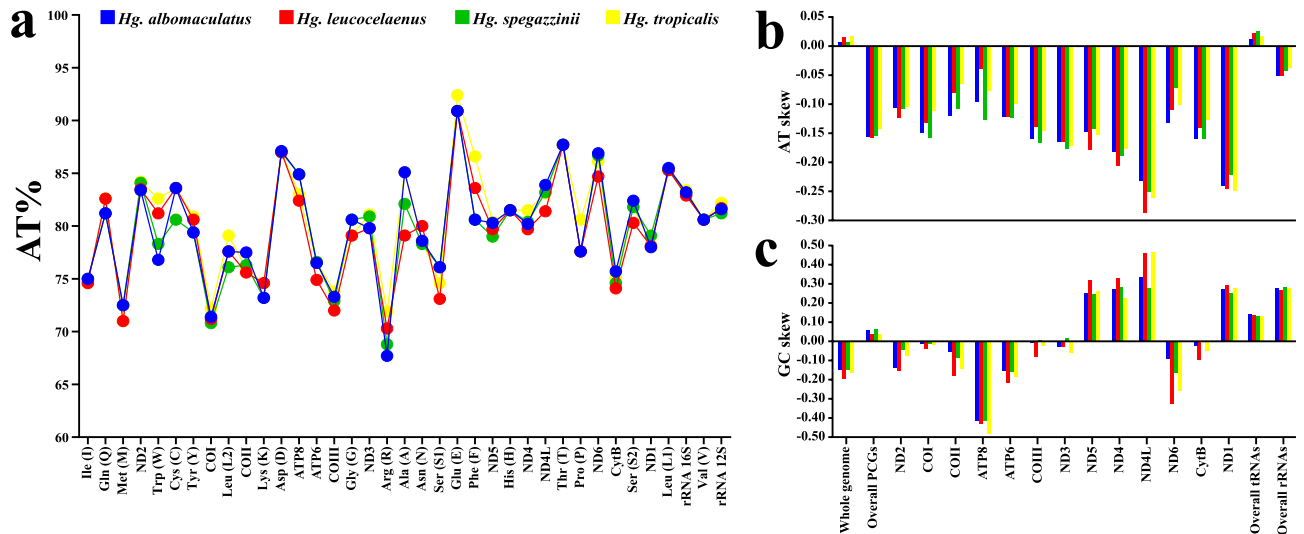


**Figure 1.** Structural representations of the *Hg. albomaculatus*, *Hg. leucocelaenus*, *Hg. spegazzinii*, and *Hg. tropicalis* mitogenomes. The internal values indicate the content of the nucleotide bases. The blue, red, and yellow blocks indicate PCGs, tRNAs, and rRNAs, respectively. Each tRNA is identified by a unique letter abbreviation. The genes arranged in the outer circle are located in the J strand (Forward), and the genes arranged in the inner circle are located in the N strand (Reverse). The structural representations of the genomes in this figure were generated using *GenomeVx* (<https://wolfe.ucd.ie/GenomeVx/>) and edited using *Inkscape* 0.92.4 (<https://inkscape.org/>).

acceptor (AA), dihydrouridine (DHU),  $T_{\psi}C$ , and Anticodons (AC)] and loops [AA, DHU,  $T_{\psi}C$ , and Variable (V)] was observed in 21 of the 22 tRNAs described for the sequenced species (Supplementary Figs. 1–4). The  $tRNA^{Ser1}$  was the only one that did not present a secondary structure similar to the others, showing the absence of DHU arm, replaced by a DHU loop, similar to that reported in previous studies of other mosquito species<sup>32,39,44,54</sup>, and other groups of insects<sup>31,48,56</sup>. Besides, between 18 and 22 mismatched base pairs were detected in the tRNAs of the mitogenomes sequenced: in *Hg. albomaculatus*, there are 17 GU pairs and 01 UU; in *Hg. leucocelaenus*, there are 20 GU pairs and 02 UU; in *Hg. spegazzinii*, there are 19 GU pairs and 01 UU; and in *Hg. tropicalis*, there are 16 GU pairs and 02 UU.

The subunits  $rRNA^{16S}$  and  $rRNA^{12S}$ , concatenated, had lengths (average of 2,153 bp) and AT contents (average of 82.6%) (Supplementary Table 7A), similar in the four mitogenomes obtained, with negative AT-skews (Fig. 2b) and positive GC-skews (Fig. 2c). Individually, they presented, respectively, average AT contents of 83.2% and 81.7%, and average lengths of 1,360 bp and 792 bp (Supplementary Tables 3–6).

The mitochondrial sequences obtained presented between 24 and 25 small intergenic regions, non-coding, and with sizes varying from 1 to 47 nucleotides, totaling in *Hg. albomaculatus* 394 non-coding nucleotides; in



**Figure 2.** Information on lengths, AT contents, and AT/GC-skews of the investigated mitogenomes. AT contents (%) (a). AT-skews (b). GC-skews (c). All graphs in this figure were generated using *GraphPad Prism 8.0.1* (<https://www.graphpad.com/scientific-software/prism/>) and edited using *Inkscape 0.92.4* (<https://inkscape.org/>).

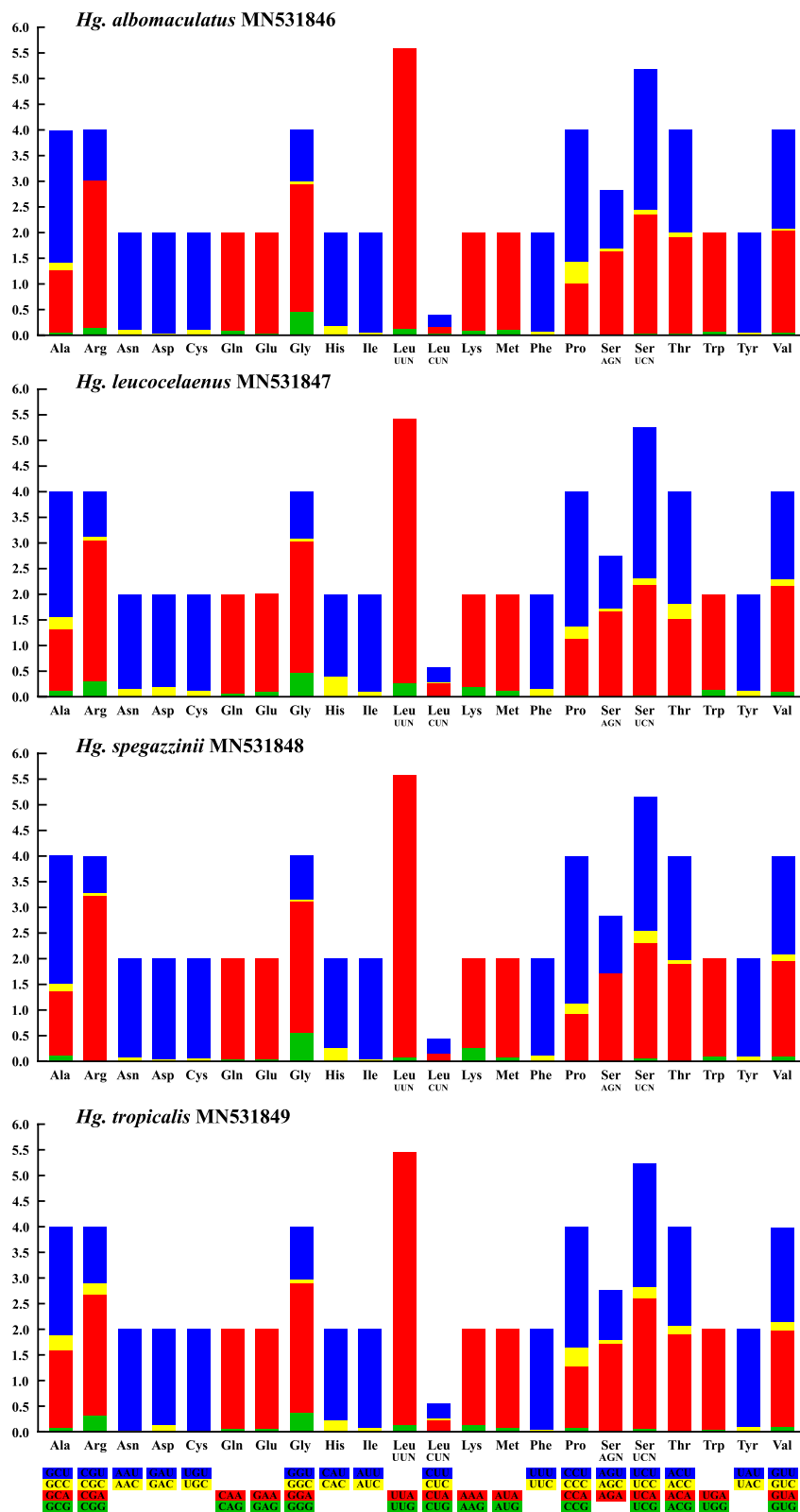
*Hg. leuocolaenus*, 351; in *Hg. spegazzinii*, 367; and in *Hg. tropicalis*, 387 (Supplementary Tables 3–6). In the first three species, there is an intergenic spacer with an average length of 174 nucleotides, located between *rRNA*<sup>12S</sup> and *tRNA*<sup>Ile</sup>. In this location, probably, the control region (A + T) of these species would be located<sup>57</sup>, as predicted in other Culicidae studies<sup>32,41,45,54</sup>, and other dipterans<sup>31,46,47</sup>. In *Hg. tropicalis*, there is an intergenic spacer of only 5 bp between the two subunits mentioned.

In this study, no sequencing data were obtained for the control region (A + T) of the analyzed species. In metazoans, this region is associated with the initial processes of mitogenome transcription and replication, presenting, in particular in invertebrates, variable lengths due to factors such as high rates of nucleotide substitution, insertions/deletions, and a varied number of repetitions (homopolymers)<sup>57–59</sup>. Likewise, Lemos et al.<sup>37</sup>, sequencing the mitogenome of *Hg. janthinomys* obtained a nucleotide sequence (length of 14,937 bp) with the absence of this region. It is essential to adopt methods that aim to include information about the control region of the species analyzed in future studies. Techniques such as conventional PCR and sequencing by the Sanger method are relevant alternatives that can be used in an attempt to obtain these data, as they allow selective amplification of specific mitogenome regions, with satisfactory results in similar studies<sup>60–62</sup>.

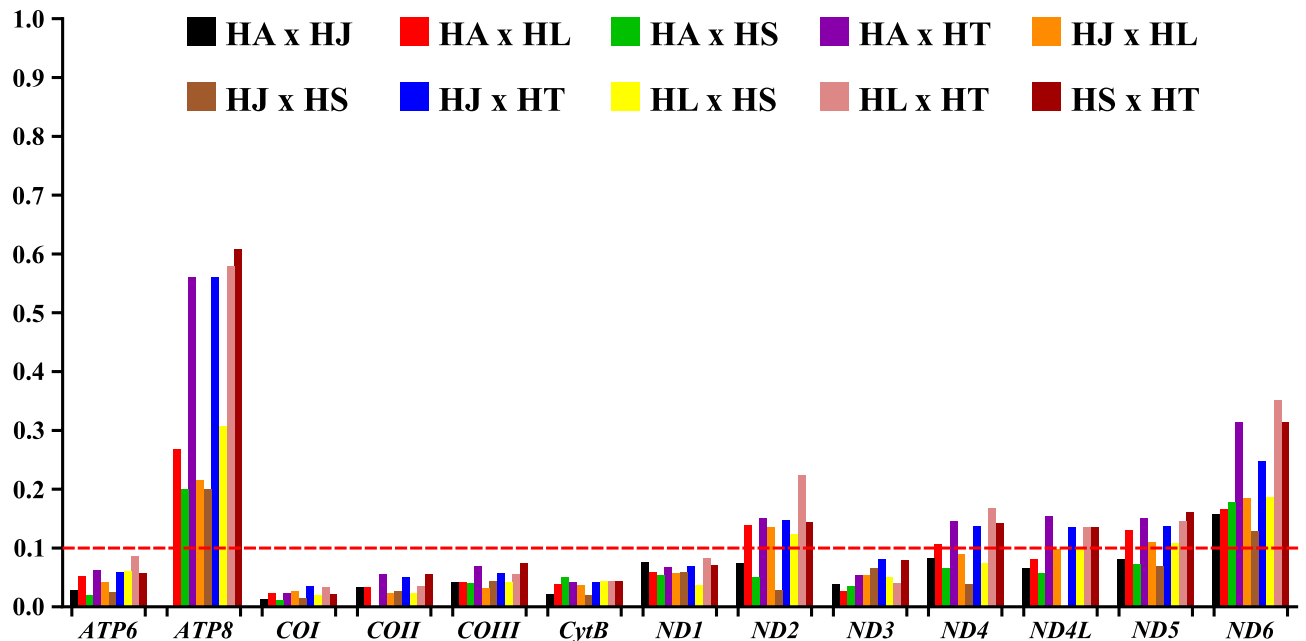
**Characteristics of protein-coding genes (PCGs).** In the mitogenomes obtained in this study, nine PCGs (*ATP6*, *ATP8*, *COI*, *COII*, *COIII*, *CytB*, *ND2*, *ND3*, and *ND6*) showed a sense of transcription on J strand (Forward), and four (*ND1*, *ND4*, *ND4L*, and *ND5*) on N strand (Reverse). Except for *ATP6* (which used TTG as the start codon), the other PCGs, in the four mitogenomes, used the ATN standard as the start codon (Supplementary Tables 3–6), similar to that reported in other studies with Culicidae<sup>32,39,45,54</sup>.

In the genes *ND1*, *ND2*, *ND3*, and *ND6*, all species are using ATA as the start codon. In *Hg. albomaculatus*, *Hg. leuocolaenus*, and *Hg. spegazzinii*, the sets of PCGs [*ATP8*, *COI*, *CytB* and *ND4*], [*COII* and *COIII*], and [*ND4L* and *ND5*] used, respectively, the start codons ATT, ATG, and ATC. In *Hg. tropicalis*, the sets of PCGs [*ATP8*, *COI*, and *COII*], [*COIII*, *CytB*, *ND4*, and *ND4L*], and [*ND5*] they also used, respectively, the start codons ATT, ATG, and ATC. The complete stop codon TAA was common to all PCGs. It is estimated that the expression of complete TAA stop codons can be related to post-transcriptional polyadenylation events<sup>63</sup>, and although it was not detected in this study, the recording of incomplete stop codons (T/TA) is common in insects<sup>32,45,55,58,64</sup>.

Excluding stop codons, there are 3,639 codons in use in *Hg. albomaculatus*, 3,647 in *Hg. leuocolaenus*, 3,650 in *Hg. spegazzinii*, and 3,646 in *Hg. tropicalis* mitogenomes. The analysis of the RSCU (Supplementary Table 8) showed that almost all codons are present in the sequenced mitogenomes, with the exception of AGG, which synthesizes the Serine amino acid. In addition, the codons with Adenine and/or Thymine (Uracil) in the third position were much more used in comparison with other synonymous codons, with Cytosine and/or Guanine in the third position (Fig. 3). This can be seen, for example, in the use of the amino acid Leucine (L), which presented the codons with the highest and lowest frequency of use in the four newly-sequenced mitogenomes: while UUA had an average RSCU value of 5.37, CUG, which synthesizes the same amino acid, had an average RSCU of 0.01. Thus, among the four investigated mitogenomes, the most frequently used codons were, in decreasing order, UUA (L) (average RSCU value 5.37), CGA (A) (2.80), UCU (S) (2.68), CCU (P) (2.61), and GGA (G) (2.54), while CUC (L) (RSCU value 0.01), ACG (T) (0.01), and CUG (L) (0.01) were rarely used. It was observed that all codons with RSCU  $\geq 1$  were of the NNU and/or NNA type, corroborating with other studies, in which this tendency to use codons seems to follow a frequent pattern, as observed in Culicidae<sup>39,40,44,45,50,54</sup>, and other groups of insects<sup>33,36,56</sup>. Among the four sequenced mitogenomes, a total of 20 different amino acids are coded, with Leucine being the most frequent amino acid (15.9%), and Cysteine the least frequent (1.01%).



**Figure 3.** Relative synonymous codon usage (RSCU) of the sequenced mitogenomes. RSCU values are represented on the y-axis, and families of synonymous codons and their respective amino acids are indicated on the x-axis. All graphs in this figure were generated using *GraphPad Prism 8.0.1* (<https://www.graphpad.com/scientific-software/prism/>) and edited using *Inkscape 0.92.4* (<https://inkscape.org/>).

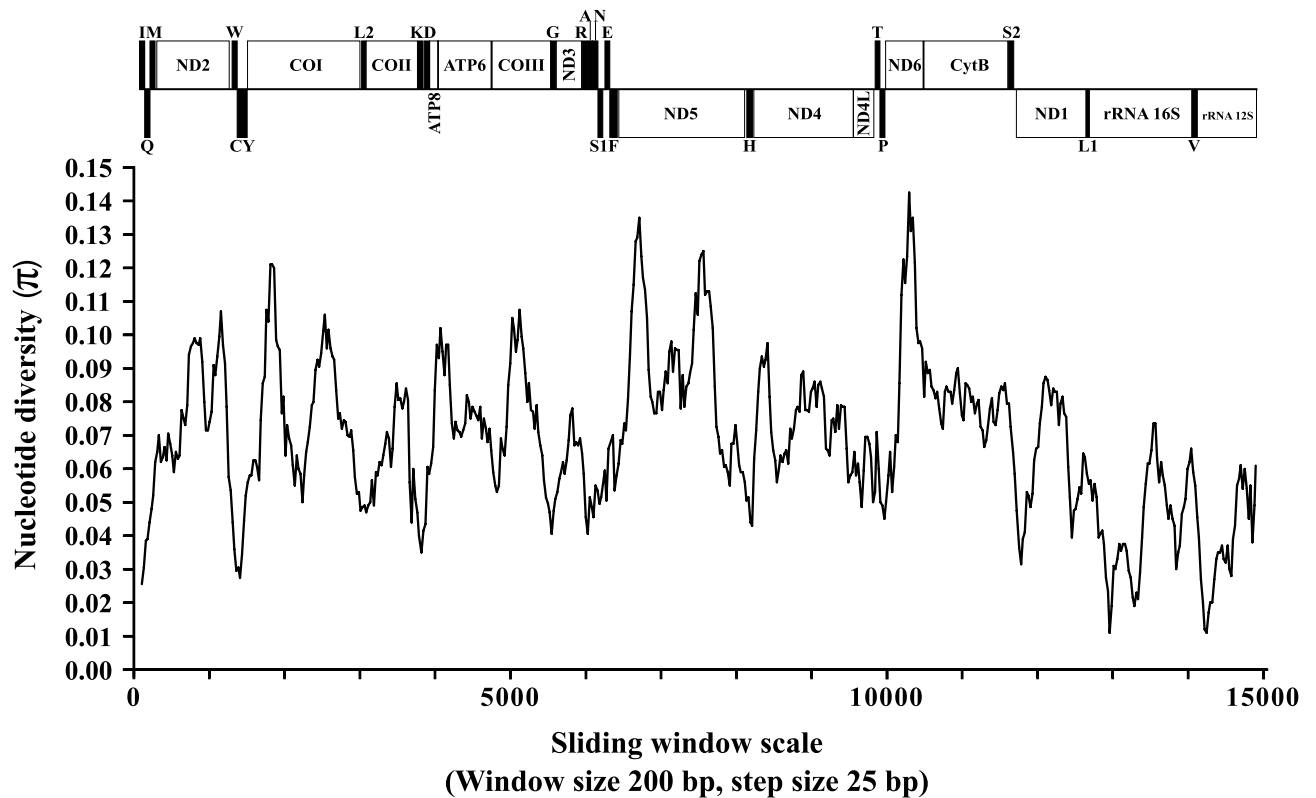


**Figure 4.** Proportions between rates non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) nucleotide substitutions ( $d_N/d_S$ ) in 05 *Haemagogus* mitogenomes. The ratios were calculated in paired analyzes between the 13 PCGs of the sequenced mitogenomes, together with the data already available for *Hg. janthinomys*. The  $d_N/d_S$  ratios are plotted on the y-axis, and the PCGs are plotted on the x-axis. Abbreviations: HA (*Hg. albomaculatus*), HJ (*Hg. janthinomys*), HL (*Hg. leucocelaenus*), HS (*Hg. spegazzinii*), and HT (*Hg. tropicalis*). The graph in this figure was generated using *GraphPad Prism 8.0.1* (<https://www.graphpad.com/scientific-software/prism/>) and edited using *Inkscape 0.92.4* (<https://inkscape.org/>).

**Evolutionary selection analysis.** In order to evaluate the pressure of evolutionary selection acting on the PCGs of the sequenced mitogenomes in this study, together with *Hg. janthinomys* (GenBank ID: NC28025), pairwise analyzes were performed to estimate the proportions between the existing non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions. The results obtained indicate that the 13 PCGs in the analyzed mitogenomes are globally evolving under negative selection pressure (purification) ( $d_N/d_S < 1$ ), with ratios ranging from  $0.0115 \pm 0.0347$  in *COI* to  $0.0 \pm 0.6079$  in *ATP8* (Fig. 4), displaying the following order: *COI* < *COII* < *CytB* < *ATP6* < *COIII* < *ND3* < *ND1* < *ND4L* < *ND4* < *ND5* < *ND2* < *ND6* < *ATP8*. The purifying selection was particularly strong ( $d_N/d_S < 0.1$ ) in the first seven coding regions of the order presented, with greater emphasis on the genes of complex III (*CytB*) and IV (*COI*, *COII*, and *COIII*) on mitogenomes. At the same time, the complex I genes (*NADH*), together with *ATP8*, showed higher  $d_N/d_S$  proportions compared to the other genes, which indicates the presence of less conservative evolutionary restrictions in these regions, which showed a relaxed purifying selection. The rates of synonymous substitutions were significantly higher compared to non-synonymous substitutions in all PCGs (mainly in the *COI*) of the *Haemagogus* mitogenomes evaluated here. The results obtained corroborate the pattern observed in previous studies, which also demonstrated the existence of a heterogeneity between the evolution rates in the different complexes that encode proteins of the mitochondrial genome, as well as the predominance of a strong active purifying selection in these regions<sup>32,39,55,65</sup>.

**Nucleotide diversity analysis.** We performed here a sliding window analysis based on the complete alignment of the mitogenomes sequenced in this study, together with the data already available in *Hg. janthinomys* (GenBank ID: NC28025), in order to verify which regions of the mitogenome have greater nucleotide divergence, allowing the identification of potential markers that can be used in future evolutionary investigations of the *Haemagogus* genus. The results demonstrated a diversity of nucleotides ( $\pi$ ) ranging from 0.0110 to 0.1425 between the sequences analyzed (Fig. 5). Among the PCGs (all showed  $\pi \geq 0.06$ ), the highest values of diversity ( $\pi \geq 0.11$ ) were for *ND6* ( $\pi = 0.0450 \pm 0.1425$ ), *ND5* ( $\pi = 0.0430 \pm 0.1350$ ), and *COI* ( $\pi = 0.0275 \pm 0.1210$ ). Just below ( $0.11 > \pi \geq 0.10$ ), were *COIII* ( $\pi = 0.0405 \pm 0.1075$ ), *ND2* ( $\pi = 0.0305 \pm 0.1070$ ), and *ATP6* ( $\pi = 0.0550 \pm 0.1020$ ). Among tRNAs, the highest diversity value ( $0.09 > \pi \geq 0.08$ ) was for *tRNA<sup>Ser2</sup>* ( $\pi = 0.0315 \pm 0.0855$ ), and for rRNAs, the highest diversity value achieved ( $0.08 > \pi \geq 0.06$ ) went to *rRNA<sup>16S</sup>* ( $\pi = 0.0110 \pm 0.0735$ ), followed by *rRNA<sup>12S</sup>* ( $\pi = 0.0110 \pm 0.0610$ ).

Notably, PCGs showed the highest nucleotide diversity values, compared to those obtained for tRNAs and rRNAs. Among these, the 5' proximal ends of the *ND5* and *COI* genes showed two of the most divergent points, similar to those observed in other other descriptive studies focusing on the mitogenome of other mosquitoes species<sup>39,40,52</sup>. In particular, the *COI* gene is widely used in phylogenetic studies, being used as a universal barcode for species identification<sup>66</sup>. Two of the main advantages of using this gene are: the high robustness and availability of universal primers that allow to recover the 5' end of the representatives of most animal phyla<sup>67</sup>, and, being



**Figure 5.** Nucleotide diversity among the mitogenomes of the four species evaluated, with *Hg. janthinomys* (GenBank ID: NC28025). The  $\pi$  values were calculated from a sliding window analysis of 200 bp in 25 bp steps, and are represented on the y-axis. The length values of the aligned sequence are represented on the x-axis. The limits of each gene are indicated in the representation above: vertical black bars indicate tRNAs and white rectangles indicate PCGs and rRNAs. The graph of this figure was generated using *DnaSP v.6.12.03* (<https://www.ub.edu/dnasp/>) and edited using *Inkscape 0.92.4* (<https://inkscape.org/>).

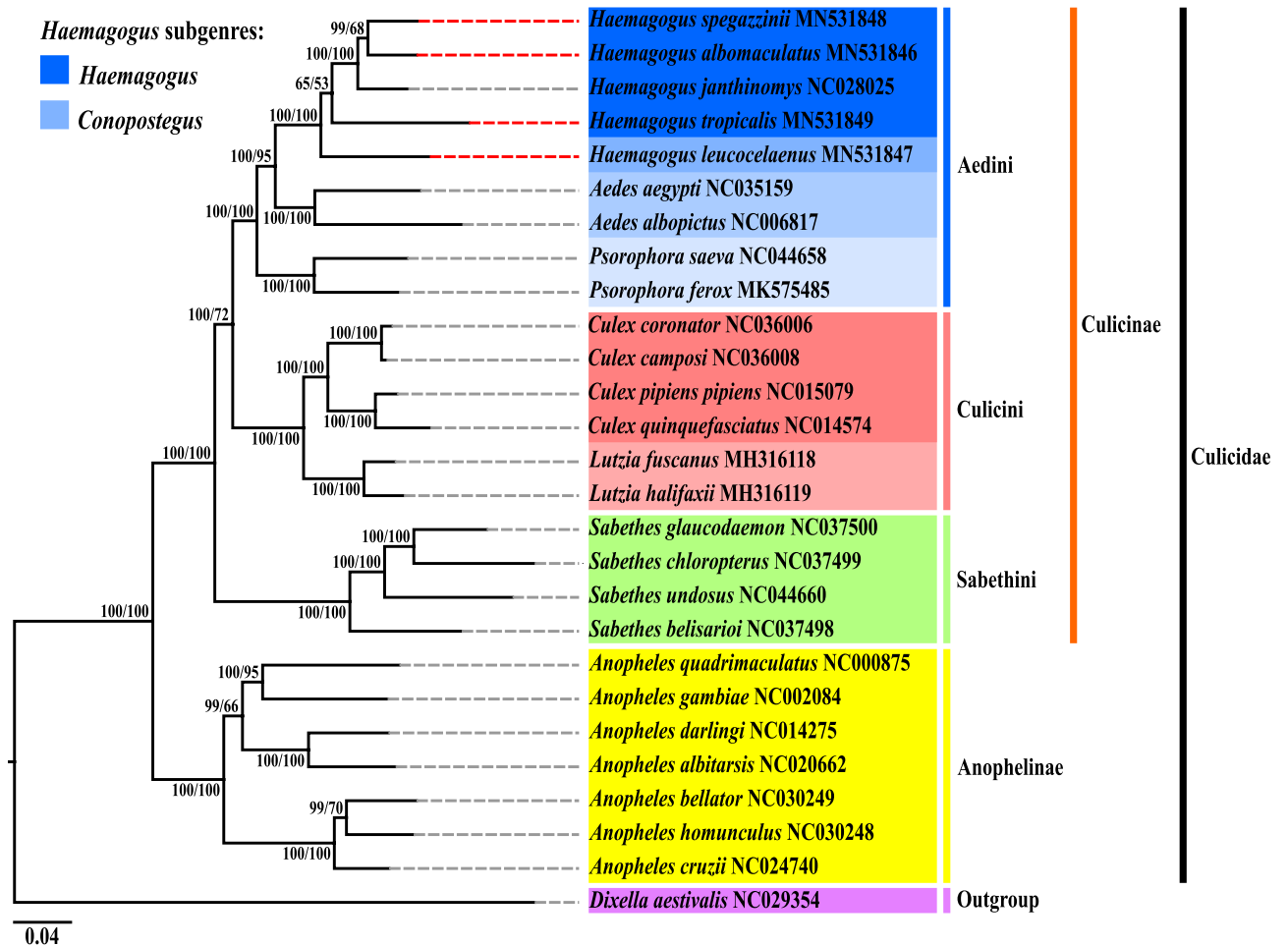
present in hundreds of copies per cell, like other PCGs, this gene has a high rate of base replacement, especially in the third codon region. However, it is observed that changes in its amino acid sequence seem to occur more slowly than in any other mitochondrial gene, a factor that directly contributes to the resolution of deeper taxonomic affinities<sup>66,68</sup>. In mosquitoes, species identification is difficult in some cases due to the great similarity of the external anatomical characteristics. In this situation, sometimes only the observation of structures of the male genitalia allows a reliable identification<sup>69</sup>. In this sense, the alternative use of the *COI* in the identification of species of great epidemiological importance has shown satisfactory results<sup>70–75</sup>.

The results obtained here suggest that PCGs are the most suitable regions to be used as molecular markers to elucidate the phylogenetic relationships between sequenced species of *Haemagogus*. More specifically, the combined results of the sliding window analysis and the  $d_n/d_s$  ratios suggest that the *COI*, *ND5*, *ND6*, and *ATP8* genes represent potential evolutionary markers in future studies on *Haemagogus* genus, based on the highest values of nucleotide diversity (in *COI*, *ND5*, and *ND6*), in the strong purifying selection that operates in these regions, especially in *COI*, and in the record of the highest accumulation rates of non-synonymous substitutions in *ATP8* and *ND6*, compared to other PCGs. However, these results need to be evaluated in future studies with the inclusion of a greater number of mitochondrial sequences from other species of the genus investigate.

**Phylogenetic analysis.** The nucleotide sequences alignment containing all 13 PCGs of 27 taxa (04 of them with mitogenomes sequenced in this study, and another 23 mitogenomes with data available in the GenBank database), showed an average nucleotide distance of 0.16%, with percentages ranging from 0.01 to 0.28%. The *Hg. janthinomys* (GenBank ID: NC028025) was the taxon with the lowest average nucleotide distance (0.07%), and *Dixella aestivalis* (GenBank ID: NC029354) was the taxon with the highest average nucleotide distance (0.26%), when compared to the newly-sequenced species (Supplementary Table 9).

The best nucleotide replacement model defined under the Akaike information criterion (AIC) for the reconstruction of phylogenies by Bayesian inference and Maximum Likelihood methods was the *General Time Reversible with Gamma Distribution (GTR + G)*. The phylogenetic analyzes by the two methods produced topologies of identical trees (Fig. 6; Supplementary Figs. 5, 6), with an initial structure of 26 taxa in a large monophyletic group, corresponding to the Culicidae family, anchored externally by the taxon *Dixella aestivalis* (Dixidae). The definition of the external group in the two phylogenies occurred in an automated way (*midpoint*)<sup>76</sup>.

The phylogeny obtained showed that Culicidae contained two well-related clades (BP/BPP = 100%/100%), corresponding to the subfamilies Anophelinae and Culicinae. Anophelinae, containing seven taxa, demonstrated



**Figure 6.** Phylogenetic reconstruction by the Bayesian Inference and Maximum Likelihood methods, based on the concatenation of the 13 PCGs of the *Haemagogus* species sequenced in this study and 23 other taxa with data available on GenBank database. The subsequent Bayesian probabilities (BP), and the values of support for bootstrapping (BPP) for 1,000 replicates, respectively, are shown on the left, in each node. The complementary lines dotted in red indicate the newly-sequenced species. The GenBank database accession numbers and references of mitogenomes sequences of the taxa used are listed in Supplementary Table 2. The topologies were visualized using *FigTree v.1.4.4* (<https://tree.bio.ed.ac.uk/software/figtree/>) and edited using *Inkscape 0.92.4* (<https://inkscape.org/>).

a monophyletic conformation previously predicted in morphological and molecular studies, focusing on the mitogenomes of species of the *Anopheles* genus<sup>51,77–79</sup>. Likewise, it was observed in Culicinae a monophyletic arrangement formed by 19 taxa distributed in three subclades, which corroborated with studies of morphological and molecular taxonomy, also based on the sequencing of the mitogenome of species belonging to the tribes Sabethini<sup>42,43</sup>, Culicini<sup>39,40,45</sup>, and, in particular, Aedini<sup>2,37,80</sup>, which in the phylogeny reconstructed, included three closely related groupings, corresponding to the *Haemagogus*, *Aedes*, and *Psorophora* genera.

Currently, the *Haemagogus* genus, focus of this study, comprises 28 species, distributed in two subgenera: *Conopostegus*, with 04 species, and *Haemagogus*, with 24 species. The current internal taxonomic classification of this genus, based on morphological aspects, is mainly due to the review studies carried out by Zavortink<sup>3</sup>, who proposed the relocation of *Hg. leucocelaenus* and three other species representatives of the subgenus *Conopostegus*, previously classified in *Aedes*, to the *Haemagogus* genus, and by Arnell<sup>1</sup>, who supported the relocation proposed by Zavortink<sup>3</sup> and proposed the merging of all other 24 species, previously classified into three subgenera (*Haemagogus*, *Stegoconops* Lutz, 1905, and *Longipalpijer* Levi-Castillo, 1951) in a single subgenus (*Haemagogus*). In addition, Arnell<sup>1</sup> also proposed to subdivide the species of the newly merged subgenus *Haemagogus* into three sections: Albomaculatus, with 14 species, would contain all those previously belonging to the old subgenera *Stegoconops* and *Longipalpijer*; Splendens, with 09 species, would contain those previously belonging to the old subgenus *Haemagogus*; and Tropicalis, which would be monotypic, comprising only the species *Hg. tropicalis*, which has morphological characters from the first section, in the adult phase, and the second, in the larval development phase.

The phylogeny of *Haemagogus* spp. resulted in a monophyletic subclade, in accordance with the taxonomic classification known for the genus<sup>1–3,80</sup>. The generated topology allowed the taxa classification into two subgenera, equally to the current taxonomic classification based on morphology: *Hg. leucocelaenus*, the first species



among the taxa evaluated to compose the evolutionary lineage of the grouping, belongs to the subgenus *Conopostegus*, while the other taxa in the above branches belong to the subgenus *Haemagogus*. For this last subgeneric category, it was also possible to classify taxa into two species sections: *Hg. albomaculatus*, *Hg. spegazzinii*, and *Hg. janthinomys* belong to the Albomaculatus section, and just below *Hg. tropicalis* belongs to the monotypic section Tropicalis<sup>1</sup>. The low nucleotide distance observed among *Haemagogus* sequences (which ranged from 0.05 between *Hg. albomaculatus* and *Hg. spegazzinii* to 0.11 between *Hg. leucoelaenus* and *Hg. Tropicalis*—Supplementary Table 9) supports the monophilia observed in the structuring of the genus subclade.

The results of the reconstructed phylogeny strongly corroborated the review study carried out by Zavortink<sup>3</sup>, where this subgenus maintained many primitive morphological characters and would probably be closer to the ancestral lineage of the genus. In addition, more recent studies have made observations on *Hg. leucoelaenus*, that demonstrated, in addition to the morphological similarity, ecological behaviors similar to the species of *Aedes*<sup>6,11,12</sup>. Therefore, based on the mitochondrial genes used and the number of available taxa, it was possible to confirm the current evolutionary position of this species and its subgenus.

*Haemagogus tropicalis* presented itself as the first species in the evolutionary lineage of the group of taxa belonging to the subgenus *Haemagogus*. Despite the taxonomic agreement, the probability of anchoring this taxon in relation to the upper grouping was low in the two phylogenetic analysis methods applied (BP/BPP = 65%/53%), considering the number of available taxa and included in the analysis. It is important to note that, in this study, mitogenome sequencing data were obtained only from representatives of the sections Albomaculatus and Tropicalis of the subgenus *Haemagogus*, lacking genomic data from representatives of the Splendens section. The future inclusion of more taxa may likely influence the phylogenetic relations of the mitochondrial genes used, however, it will be of great relevance for the evolutionary biology studies of the genus investigated.

To address the effectiveness of the genes *ATP8*, *COI*, *ND5*, and *ND6* in the reconstruction of the phylogenetic relationships of the newly-sequenced species, an additional phylogenetic analysis was performed adopting the same parameters established for the phylogeny based on the use of 13 PCGs (Supplementary Figs. 7, 8). The topologies generated for the *Haemagogus* species grouping were similar to those obtained in the 13 PCGs.

## Methods

**Sampling and extraction of total DNA.** The *Haemagogus* samples used in this study came from field expeditions carried out in the state of Bahia (Northeast Brazil), in the regions of Jaborandi (S 13° 58' 92" W 44° 53' 36") in 2013, and Coribe (S 13° 82' 94" W 44° 44' 92"), and in the state of Pará (Northern Brazil), in the regions of Canãa dos Carajás (S 06° 49' 75" W 49° 87' 83") in 2016 (MMA/ICMbio/Carajás National Forest/Direct authorization n°021/2018), and Ilha do Combu (S 01° 51' 97" W 48° 49' 28") in 2018 (MMA/ICMbio/SISBIO authorization n°60595-1) (Supplementary Table 1). The specimens were identified taxonomically using gender-specific dichotomous keys<sup>2,4,81</sup>, organized in batches containing mosquitoes of the same species, and stored at -70°C for conservation. Subsequently, the mosquito batches were macerated with 3 mm stainless steel beads using TissueLyser II (Qiagen), and the genomic DNA of each batch of species was extracted, using the DNeasy Blood & Tissue kit (Qiagen), following the recommendations established by the manufacturer.

**DNA quantification, library construction and sequencing.** The genomic DNA extracted from each batch of species was quantified using the Qubit 2.0 fluorometer (Life Technologies), with the Qubit dsDNA Hs Assay kit (Invitrogen), following the manufacturer instructions. The genomic libraries were constructed using the Ion Xpress Plus Fragment Library Kit (Thermo Fischer Scientific) in the AB Library Builder system (Life Technologies). In this step, the 200 bp library fragments were individually linked to barcode sequences for sample identification, using the Ion Xpress Barcode DNA Adaptors 1–32 Kit (Thermo Fischer Scientific). The prepared genomic libraries were subjected to a size selection (200 bp) by electrophoresis on 2% agarose gel, using the E-Gel SizeSelect (Invitrogen), aiming at the removal of short sequences, and then were quantified by *qPCR* on LightCycler 480 (Roche), using the Ion Library TaqMan quantitation kit (Thermo Fischer Scientific). Subsequently, the quantified fragments were amplified, through the *emPCR*, using the Ion PGM Hi-Q View OT2 kit (Thermo Fischer Scientific), in the Ion OneTouch 2 system (Thermo Fischer Scientific). Finally, the libraries were enriched with the Ion Xpress template kit (Thermo Fischer Scientific), deposited in the Ion 318 Chip Kit v2 BC (Thermo Fischer Scientific), and sequenced in the Ion Torrent PGM system (Thermo Fischer Scientific), using Ion PGM HI-Q View Sequencing (Thermo Fischer Scientific), following the recommendations established by the manufacturer.

**Mitochondrial genome assembly.** The output files (Raw data) corresponding to the sequencing reads for each of the species evaluated were subjected to genomic assembly procedures by the De Novo method, using the IDBA UD v.1.1.1<sup>82</sup> and SPAdes v.3.10.1<sup>83</sup>. Then, the selection of the regions corresponding to the mitogenome of each species was performed using DIAMOND<sup>84</sup>, these regions were visualized using MEGAN6<sup>85</sup>. The contigs with match to mitogenome was comparing with the available information of the complete sequencing of *Hg. janthinomys* (GenBank ID: NC28025), and inspected manually using Geneious v.9.1.8<sup>86</sup>.

**Sequence analysis.** The obtained sequences were annotated using the online tool *MITOchondrial genome annotation Server*<sup>87</sup> (MITOS—<https://mitos.bioinf.uni-leipzig.de/>), which also provided the prediction of the secondary structures of the tRNAs. Geneious v.9.1.8<sup>86</sup> was used to cure sequences manually, and graphical maps of the mitogenomes were generated using the online tool *GenomeVx*<sup>88</sup> (<https://wolfe.ucd.ie/GenomeVx/>). The nucleotide content values for each sequence were obtained using MEGA X<sup>89</sup>. The composition bias based on the asymmetry values (Whole genome, concatenated PCGs, individualized PCGs, concatenated tRNAs, and concatenated rRNAs) was estimated using the formulas: AT-skew = (A% - T%) / (A% + T%), and GC-skew = (G% - C%) /

(G% + C%)<sup>90</sup>. A sliding window of 200 bp in steps 25 bp was performed using DnaSP v.6.12.03<sup>91</sup> to estimate the diversity of nucleotides between the sequences obtained in this study, together with the data already available for *Hg. janthinomys* (GenBank ID: NC28025). In the PCGs, the relative use of synonymous codons (RSCU) was estimated using MEGA X<sup>89</sup>, and the pairwise comparison of the proportions of non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions ( $d_N/d_S$ ) were calculated using the DnaSP v.6.12.03<sup>91</sup>. In this last analysis, the reasons obtained are used to estimate whether the protein-coding genes are under negative selection (purification) ( $d_N/d_S < 1$ ), positive selection (adaptive) ( $d_N/d_S > 1$ ), or neutral evolution.

**Phylogenetic analysis.** The sequences containing the 13 PCGs of 27 mitogenomes (including the four mitogenomes sequenced in this study, and 23 mitogenomes available in the GenBank database—Supplementary Table 2) were extracted together using Geneious v.9.1.8<sup>86</sup>, and were aligned using the MAFFT algorithm<sup>92</sup>, being grouped into a single reading file. The nucleotide distances between the sequences used were obtained based on the alignment performed, using the Maximum Composite Likelihood model to build the distance matrix in the MEGA X<sup>89</sup>. The best nucleotide replacement model for phylogeny reconstruction was defined under the Akaike information criterion (AIC), using jModelTest v.2.1.7<sup>93</sup>. Bayesian inference was performed using MrBayes v.3.2.7a<sup>94</sup> in two independent and simultaneous runs, with four chains each (three hot chains and one cold chain), established for 5,000,000 generations with a sampling of trees every 1000 steps, with relative burning of 25%. The Maximum Likelihood analysis was performed using RaxML v.8.2.11<sup>95</sup>, with bootstrapping values defined for 1,000 repetitions. The obtained topologies were visualized in FigTree v.1.4.4<sup>96</sup>.

### Data availability

All data generated during this study are available as tables and figures included in this published article and its Supplementary Information files. The GenBank database accession numbers for the four mitochondrial genomes sequenced in this study are MN531846 (*Hg. albomaculatus*), MN531847 (*Hg. leucocelaenus*), MN531848 (*Hg. spegazzinii*), and MN531849 (*Hg. tropicalis*).

Received: 3 July 2020; Accepted: 14 September 2020

Published online: 12 October 2020

### References

1. Arnell, J. H. Mosquito studies (Diptera, Culicidae). XXXII. A revision of the genus *Haemagogus*. *Contrib. Am. Entomol. Inst.* **10**, 1–174 (1973).
2. Forattini, O. P. *Culicidologia Médica. Identificação, Biologia, Epidemiologia. Vol 2.* (edusp, 2002).
3. Zavortink, T. J. Mosquito studies (Diptera, Culicidae). XXVIII. The new world species formerly placed in *Aedes* (Finlaya). *Contrib. Am. Entomol. Inst.* **8**, 1–206 (1972).
4. Marcondes, C. & Alencar, J. Revisão de mosquitos *Haemagogus* Williston (Diptera: Culicidae) do Brasil. *Rev. Biomed* **21**, 221–238 (2010).
5. Camargo-Neves, V. L. F. *et al.* Entomological investigation of a sylvatic yellow fever area in São Paulo State, Brazil. *Cad. Saude Publica* **21**, 1278–1286 (2005).
6. Alencar, J. *et al.* Diversity of yellow fever mosquito vectors in the Atlantic forest of Rio de Janeiro, Brazil. *Rev. Soc. Bras. Med. Trop.* **49**, 351–356 (2016).
7. Vasconcelos, P. F. C. Yellow fever in Brazil: thoughts and hypotheses on the emergence in previously free areas. *Rev. Saude Publica* **44**, 1144–1149 (2010).
8. Monath, T. P. & Vasconcelos, P. F. C. Yellow fever. *J. Clin. Virol.* **64**, 160–173 (2015).
9. Hoch, A. L., Peterson, N. E., LeDuc, J. W. & Pinheiro, F. P. An outbreak of Mayaro Virus disease in Belterra, Brazil. III. Entomological and ecological studies. *Am. J. Trop. Med. Hyg.* **30**, 689–698 (1981).
10. Azevedo, R. S. S. *et al.* Mayaro fever virus, Brazilian amazon. *Emerg. Infect. Dis.* **15**, 1830–1832 (2009).
11. Vasconcelos, P. F. C. *et al.* Isolations of yellow fever virus from *Haemagogus leucocelaenus* in Rio Grande do Sul State, Brazil. *Trans. R. Soc. Trop. Med. Hyg.* **97**, 60–62 (2003).
12. Cardoso, J. C. *et al.* Yellow fever virus in *Haemagogus leucocelaenus* and *Aedes serratus* mosquitoes, Southern Brazil, 2008. *Emerg. Infect. Dis.* **16**, 1918–1924 (2010).
13. Travassos-da-Rosa, A. P. A., Vasconcelos, P. F. C., Hervé, J. P. & Travassos-da-Rosa, J. F. S. Febre amarela silvestre no Estado do Pará-Brasil, 1984. *Bol. Epidem. FSESP.* **16**, 97–104 (1984).
14. Hervé, J. P., Dégallier, N., Filho, G. C. S. & Travassos-da-Rosa, A. P. A. Ecologia da Febre amarela silvestre no Brasil. *Rev. Fund. SESP.* **31**, 13–134 (1986).
15. Liria, J. & Navarro, J. Modelo de nicho ecológico en *Haemagogus* Williston (Diptera: Culicidae), vectores del virus de la fiebre amarilla. *Rev. Biomed* **21**, 149–161 (2010).
16. Vasconcelos, P. F. C. *et al.* Epidemic of jungle yellow fever in Brazil, 2000: implications of climatic alterations in disease spread. *J. Med. Virol.* **65**, 598–604 (2001).
17. Alencar, J. *et al.* New records of *Haemagogus* (*Haemagogus*) from Northern and Northeastern Brazil (Diptera: Culicidae, Aedini). *Zootaxa* **68**, 65–68 (2008).
18. Alencar, J. *et al.* Scanning electron microscopy of the egg of *Haemagogus tropicalis*. *J. Am. Mosq. Control Assoc.* **24**, 16–20 (2008).
19. Yan, G., Chadee, D. D. & Severson, D. W. Evidence for genetic hitchhiking effect associated with insecticide resistance in *Aedes aegypti*. *Genetics* **148**, 793–800 (1998).
20. DeSalle, R., Egan, M. G. & Siddall, M. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1905–1916 (2005).
21. Lobo, F. P. *et al.* Virus-host coevolution: Common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS ONE* **4**, e6282 (2009).
22. Barzon, L., Lavezzo, E., Militello, V., Toppo, S. & Palù, G. Applications of next-generation sequencing technologies to diagnostic virology. *Int. J. Mol. Sci.* **12**, 7861–7884 (2011).
23. Tabachnick, W. J. Nature, nurture and evolution of intra-species variation in mosquito arbovirus transmission competence. *Int. J. Environ. Res. Public Health* **10**, 249–277 (2013).
24. Sim, S., Jupatanakul, N. & Dimopoulos, G. Mosquito immunity against arboviruses. *Viruses* **6**, 4479–4504 (2014).
25. Severson, D. W. & Behura, S. K. Genome investigations of vector competence in *Aedes aegypti* to inform novel arbovirus disease control approaches. *Insects* **7**, 58 (2016).

26. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
27. Brown, W. M., George, M. & Wilson, A. C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci.* **76**, 1967–1971 (1979).
28. Moritz, C., Dowling, T. E. & Brown, W. M. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* **18**, 269–292 (1987).
29. Simon, C. *et al.* Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701 (1994).
30. Avise, J. C. Phylogeography: retrospect and prospect. *J. Biogeogr.* **36**, 3–15 (2009).
31. Jiang, F. *et al.* The first complete mitochondrial genome of *Dacus longicornis* (Diptera: Tephritidae) using next-generation sequencing and mitochondrial genome phylogeny of Dacini tribe. *Sci. Rep.* **6**, 36426 (2016).
32. Hao, Y. J. *et al.* Complete mitochondrial genomes of *Anopheles stephensi* and *An. dirus* and comparative evolutionary mitochondrialomics of 50 mosquitoes. *Sci. Rep.* **7**, 7666 (2017).
33. Pu, D. Q. *et al.* Mitochondrial genomes of the hoverflies *Episyrphus balteatus* and *Eupeodes corollae* (Diptera: Syrphidae), with a phylogenetic analysis of Muscomorpha. *Sci. Rep.* **7**, 44300 (2017).
34. Pita, S. *et al.* Complete mitochondrial genome of *Triatoma infestans* (Hemiptera, Reduviidae, Triatominae), main vector of Chagas disease. *Infect. Genet. Evol.* **54**, 158–163 (2017).
35. Chen, L. *et al.* Extensive gene rearrangements in the mitochondrial genomes of two egg parasitoids, *Trichogramma japonicum* and *Trichogramma ostrinae* (Hymenoptera: Chalcidoidea: Trichogrammatidae). *Sci. Rep.* **8**, 7034 (2018).
36. Sun, Y. X. *et al.* Characterization of the complete mitochondrial genome of *Leucoma salicis* (Lepidoptera: Lymantriidae) and comparison with other lepidopteran insects. *Sci. Rep.* **6**, 39153 (2016).
37. Lemos, P. S. *et al.* Characterization of mitochondrial genome of *Haemagogus janthinomys* (Diptera: Culicidae). Mitochondrial DNA Part A DNA mapping. *Seq. Anal.* **28**, 50–51 (2017).
38. Behura, S. K. *et al.* Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. *Insect Biochem. Mol. Biol.* **41**, 770–777 (2011).
39. Demari-Silva, B. *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* **16**, 831 (2015).
40. Luo, Q. C. *et al.* The mitochondrial genomes of *Culex tritaeniorhynchus* and *Culex pipiens pallens* (Diptera: Culicidae) and comparison analysis with two other *Culex* species. *Parasites Vectors* **9**, 406 (2016).
41. Chen, K., Wang, Y., Li, X. Y., Peng, H. & Ma, Y. J. Sequencing and analysis of the complete mitochondrial genome in *Anopheles sinensis* (Diptera: Culicidae). *Infect. Dis. Poverty* **6**, 149 (2017).
42. Aragão, A. O. *et al.* Description and phylogeny of the mitochondrial genome of *Sabethes chloropterus*, *Sabethes glaucodaemon* and *Sabethes belisarioi* (Diptera: Culicidae). *Genomics* **111**, 607–611 (2019).
43. Lorenz, C., Alves, J. M. P., Foster, P. G., Sallum, M. A. M. & Suesdek, L. First record of translocation in Culicidae (Diptera) mitogenomes: evidence from the tribe Sabethini. *BMC Genomics* **20**, 721 (2019).
44. Martínez-Villegas, L. *et al.* Characterization of the complete mitogenome of *Anopheles aquasalis*, and phylogenetic divergences among *Anopheles* from diverse geographic zones. *PLoS ONE* **14**, e0219523 (2019).
45. Sun, L. *et al.* The complete mt genomes of *Lutzia halifaxia*, *Lt. fuscatus* and *Culex pallidothorax* (Diptera: Culicidae) and comparative analysis of 16 *Culex* and *Lutzia* mt genome sequences. *Parasites Vectors* **12**, 368 (2019).
46. Wang, K. *et al.* The complete mitochondrial genome of the *Atylotus miser* (Diptera: Tabanomorpha: Tabanidae), with mitochondrial genome phylogeny of lower Brachycera (Orthorrhapha). *Gene* **586**, 184–196 (2016).
47. Yong, H. S., Song, S. L., Lim, P. E. & Eamsobhana, P. Complete mitochondrial genome of *Zeugodacus tau* (Insecta: Tephritidae) and differentiation of *Z. tau* species complex by mitochondrial cytochrome c oxidase subunit I gene. *PLoS ONE* **12**, e0189325 (2017).
48. Dai, L. S., Kausar, S., Abbas, M. N. & Wang, T. T. Complete sequence and characterization of the *Ectropis oblique* mitochondrial genome and its phylogenetic implications. *Int. J. Biol. Macromol.* **107**, 1142–1150 (2018).
49. Wei, S. J. *et al.* New views on strand asymmetry in insect mitochondrial genomes. *PLoS ONE* **5**, e12708 (2010).
50. Oliveira, T. M. P. *et al.* Mitochondrial genomes of *Anopheles (Kerteszia)* (Diptera: Culicidae) from the Atlantic forest, Brazil. *J. Med. Entomol.* **53**, 790–797 (2016).
51. Peng, X. Y., Zhou, P., Duan, X. Y. & Qian, Z. Q. The mitochondrial genomes of twelve *Anopheles* mosquitoes (Diptera: Culicidae) and their phylogenetic implications. *Conserv. Genet. Resour.* **8**, 387–390 (2016).
52. Krzywinski, J. *et al.* Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex. *Mol. Phylogenet. Evol.* **58**, 469–477 (2011).
53. Sun, Z. *et al.* Comparison of base composition and codon usage in insect mitochondrial genomes. *Genes Genomics* **31**, 65–71 (2009).
54. Ding, Y. R., Li, B., Zhang, Y. J., Mao, Q. M. & Chen, B. Complete mitogenome of *Anopheles sinensis* and mitochondrial insertion segments in the nuclear genomes of 19 mosquito species. *PLoS ONE* **13**, e0204667 (2018).
55. Moreno, M. *et al.* Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malar. J.* **9**, 127 (2010).
56. Chen, Y. H., Huang, D. Y., Wang, Y. L., Zhu, C. D. & Hao, J. S. The complete mitochondrial genome of the endangered Apollo butterfly, *Parnassius apollo* (Lepidoptera: Papilionidae) and its comparison to other Papilionidae species. *J. Asia. Pac. Entomol.* **17**, 663–671 (2014).
57. Zhang, D. X. & Hewitt, G. M. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* **25**, 99–120 (1997).
58. Wolstenholme, D. R. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* **141**, 173–216 (1992).
59. Inohira, K., Hara, T. & Matsuura, E. T. Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. *Mol. Biol. Evol.* **14**, 814–822 (1997).
60. Vila, M. & Björklund, M. The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). *J. Mol. Evol.* **58**, 280–290 (2004).
61. Xu, J. & Fonseca, D. M. One-way sequencing of multiple amplicons from tandem repetitive mitochondrial DNA control region. *Mitochondrial DNA* **22**, 155–158 (2011).
62. Zhu, H. M. *et al.* Phylogeny of certain members of *Hyrceanus* group (Diptera: Culicidae) in China based on mitochondrial genome fragments. *Infect. Dis. Poverty* **8**, 91 (2019).
63. Ojala, D., Montoya, J. & Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470–474 (1981).
64. Dai, L. *et al.* Characterization of the complete mitochondrial genome of *Cerura menciana* and comparison with other lepidopteran insects. *PLoS ONE* **10**, e0132951 (2015).
65. Meiklejohn, C. D., Montooth, K. L. & Rand, D. M. Positive and negative selection on the mitochondrial genome. *Trends Genet.* **23**, 259–263 (2007).
66. Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **270**, 313–321 (2003).
67. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).

68. Lynch, M. & Jarrell, P. E. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics* **135**, 1197–1208 (1993).
69. Laurito, M. & Almirón, W. R. Phylogenetic relationships of *Culex* (*Culex*) species (Diptera, Culicidae) from Argentina based on morphological characters. *Zootaxa* **3652**, 117–155 (2013).
70. Cywinska, A., Hunter, F. F. & Hebert, P. D. N. Identifying Canadian mosquito species through DNA barcodes. *Med. Vet. Entomol.* **20**, 413–424 (2006).
71. Murugan, K. *et al.* DNA barcoding and molecular evolution of mosquito vectors of medical and veterinary importance. *Parasitol. Res.* **115**, 107–121 (2016).
72. Torres-Gutierrez, C. *et al.* Mitochondrial COI gene as a tool in the taxonomy of mosquitoes *Culex* subgenus *Melanoconion*. *Acta Trop.* **164**, 137–149 (2016).
73. Fang, Y., Shi, W. Q. & Zhang, Y. Molecular phylogeny of *Anopheles hyrcanus* group (Diptera: Culicidae) based on mtDNA COI. *Infect. Dis. Poverty* **6**, 61 (2017).
74. Laurito, M., Ayala, A. M., Almirón, W. R. & Gardenal, C. N. Molecular identification of two *Culex* (*Culex*) species of the neotropical region (Diptera: Culicidae). *PLoS ONE* **12**, e0173052 (2017).
75. Kirchgatter, K. *et al.* Phylogeny of *Anopheles* (*Kerteszia*) (Diptera: Culicidae) using mitochondrial genes. *Insects* **11**, 324 (2020).
76. Hess, P. N. & Russo, C. A. M. An empirical test of the midpoint rooting method. *Biol. J. Linn. Soc.* **92**, 669–674 (2007).
77. Harbach, R. E. The phylogeny and classification of *Anopheles*. *Anopheles mosquitoes—new insights into Malar. Vectors* **1**, 1. <https://doi.org/10.5772/54695> (2013).
78. Hua, Y. Q. *et al.* The complete mitochondrial genome of *Anopheles minimus* (Diptera: Culicidae) and the phylogenetics of known *Anopheles* mitogenomes. *Insect Sci.* **23**, 353–365 (2016).
79. Foster, P. G. *et al.* Phylogeny of anophelinae using mitochondrial protein coding genes. *R. Soc. Open Sci.* **4**, 170758 (2017).
80. Reinert, J. F., Harbach, R. E. & Kitching, I. J. Phylogeny and classification of tribe Aedini (Diptera: Culicidae). *Zool. J. Linn. Soc.* **157**, 700–794 (2009).
81. Liria, J. & Navarro, J. C. Clave fotográfica para hembras de *Haemagogus* Williston 1896 (Diptera: Culicidae) de Venezuela, con nuevo registro para el país. *B. Malaritol. Salud. Amb.* **49**, 283–292 (2009).
82. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
83. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
84. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
85. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
86. Kears, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
87. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
88. Conant, G. C. & Wolfe, K. H. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* **24**, 861–862 (2008).
89. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
90. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
91. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
92. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
93. Darrriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
94. Ronquist, F. *et al.* MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
95. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
96. Rambaut, A. FigTree v.1.4.4. (2020). <https://tree.bio.ed.ac.uk/software/figtree/>.

## Acknowledgements

This research was supported by the Evandro Chagas Institute (IEC/MS/SVS), where it was developed, and by the State University of Pará, which awarded the scholarship through the Coordination for the Improvement of Higher Education Personnel (CAPES/1743623). The authors express their gratitude to Dra. Flavia Barreto dos Santos (Oswaldo Cruz Foundation, Rio de Janeiro) for a critical review of the English language and valuable comments on this study.

## Author contributions

J.P.N.N. conceived, designed, and together with A.C.R.C. and D.B.A.M., coordinated the study. J.P.N.N., A.C.R.C., M.R.T.N., L.C.M., and H.A.O.M. provided laboratory equipment to carry out this study. L.C.M. released *Hg. leucocelaenus* samples. J.O.C. released *Hg. tropicalis* samples. G.M.C. and R.F.A. released *Hg. albomaculatus* and *Hg. spegazzinii* samples. F.S.S., S.P.S., and P.S.L. performed the genomic DNA extraction and sequencing procedures. F.S.S. and S.P.S. performed the the assembly of mitogenomes and data analysis. F.S.S. carried out the bibliographic research, prepared the figures, supplementary information, and wrote the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-73790-x>.

**Correspondence** and requests for materials should be addressed to J.P.N.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020