Decoding rRNA sequences for improved metagenomics of sylvatic mosquito species

Cassandra Koh[1], Lionel Frangeul[1], Hervé Blanc[1], Carine Ngoagouni[2], Sébastien Boyer[3], Philippe Dussart[4], Nina Grau[5], Romain Girod[5], Jean-Bernard Duchemin[6] and Maria-Carla Saleh[1*]

[1] Viruses and RNA Interference Unit, Institut Pasteur, Université Paris Cité, CNRS UMR3569, Paris F-75015, France

[2] Medical Entomology Laboratory, Institut Pasteur de Bangui, Bangui PO Box 923, Central African Republic

[3] Medical and Veterinary Entomology Unit, Institut Pasteur du Cambodge, Phnom Penh 12201, Cambodia

[4] Virology Unit, Institut Pasteur du Cambodge, Phnom Penh 12201, Cambodia

[5] Medical Entomology Unit, Institut Pasteur de Madagascar, Antananarivo 101, Madagascar

[6] Vectopôle Amazonien Emile Abonnenc, Institut Pasteur de la Guyane, Cayenne 97306, French Guiana

* To whom correspondence should be addressed.

Present address: Philippe Dussart, Institut Pasteur de Madagascar, Antananarivo 101, Madagascar; Nina Grau, Sciences Economiques et Sociales de la Santé et Traitement de l'Information Médicale, Faculté de Médecine, Marseille 13005, France

Corresponding author: Maria-Carla Saleh, carla.saleh@pasteur.fr, +33 1 45 68 85 47.

ABSTRACT

As mosquito-borne virus epidemics are often preceded by undetected spillover events, surveillance and virus discovery studies in non-urban mosquitoes informs pre-emptive and responsive public health measures. RNA-seq metagenomics is a popular methodology but it is constrained by overabundant rRNA. The lack of reference sequences for most mosquito species is a major impediment against physical and computational removal of rRNA reads.

We describe a strategy to assemble novel rRNA sequences from mosquito specimens, producing an unprecedented dataset of 234 full-length 28S and 18S rRNA sequences of 33 medically important species from countries with known histories of mosquito-borne virus circulation (Cambodia, the Central African Republic, Madagascar, and French Guiana). We also evaluate the utility of rRNA sequences as molecular barcodes relative to the mitochondrial *cytochrome* c *oxidase I* (COI) gene. We show that rRNA sequences can be used for species identification when COI sequences are ambiguous or unavailable, revealing evolutionary relationships concordant with contemporary mosquito systematics.

This expansion of the rRNA reference library improves mosquito RNA-seq metagenomics by permitting the optimization of species-specific rRNA depletion protocols for a broader species range and streamlined species identification by rRNA barcoding. In addition, rRNA barcodes could serve as an additional tool for mosquito taxonomy and phylogeny.

Keywords: surveillance, RNA-seq, ribosomal RNA, barcode, metagenomics, mosquito

INTRODUCTION

Mosquitoes top the list of vectors for arthropod-borne diseases, being implicated in the transmission of many human pathogens responsible for arboviral diseases, malaria, and lymphatic filariasis (WHO, 2017). Mosquito-borne viruses circulate in sylvatic or urban transmission cycles driven by different mosquito species with their own distinct host preferences. Although urban mosquito species are chiefly responsible for amplifying epidemics in dense human populations, non-urban (sylvatic and peri-urban) mosquitoes maintain the transmission of these viruses among forest-dwelling animal reservoir hosts and are implicated in spillover events when humans enter their ecological niches (Valentine et al., 2019). Given that mosquito-borne virus emergence is preceded by such spillover events, continuous surveillance and virus discovery in non-urban mosquitoes is integral to designing effective public health measures to pre-empt or respond to mosquito-borne viral epidemics.

Metagenomics on field specimens is the most powerful method in the toolkit for understanding mosquito-borne disease ecology through the One Health lens (Webster et al., 2016). With next-generation sequencing becoming more accessible, such studies have provided unprecedented insights into the interfaces among mosquitoes, their environment, and their animal and human hosts. However, working with lesser studied mosquito species poses several problems.

First, metagenomics studies based on RNA-seq are bedevilled by overabundant ribosomal RNAs (rRNAs). These non-coding RNA molecules comprise at least 80% of the total cellular RNA population (Gale & Crampton, 1989). Due to their length and their abundance, they are a sink for precious next generation sequencing reads, decreasing the sensitivity of pathogen detection unless depleted during library preparation. Yet the most common rRNA depletion protocols require prior knowledge of rRNA sequences of the species of interest as they involve hybridizing antisense oligos to the rRNA molecules prior to removal by ribonucleases (Fauver et al., 2019; Phelps et al., 2021) or by bead capture (Kukutla et al., 2013). Presently, reference sequences for rRNAs are limited to only a handful of species from three genera: *Aedes*, *Culex*, and *Anopheles* (Ruzzante et al., 2019). The lack of reliable rRNA depletion methods could deter mosquito metagenomics studies from expanding their sampling diversity, resulting in a gap in our knowledge of mosquito vector ecology. The inclusion of lesser studied yet medically relevant non-urban species is therefore imperative.

Second, species identification based on morphology is notoriously complicated between members of species subgroups. This is especially the case among *Culex* subgroups. Sister species are often

69 sympatric and show at least some competence for a number of viruses, such as Japanese
70 encephalitis virus, St Louis encephalitic virus, and Usutu virus (Nchoutpouen et al., 2019). Although
71 they share many morphological traits, each of these species have distinct ecologies and host
72 preferences, thus the challenge of correctly identifying vector species can affect epidemiological risk
73 estimation for these diseases (Farajollahi et al., 2011). DNA molecular markers are often employed to
74 a limited degree of success to distinguish between sister species (Batovska et al., 2017; Zittra et al.,
75 2016).

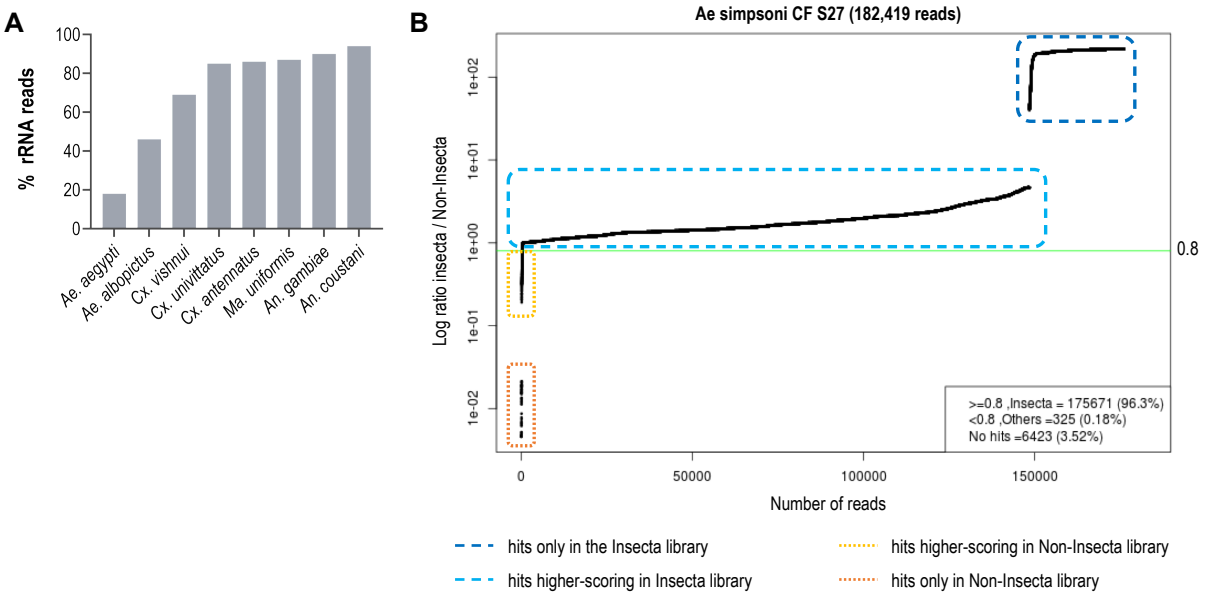
76 To address the lack of full-length rRNA sequences in public databases, we sought to determine
77 the 28S and 18S rRNA sequences of a diverse set of Old and New World non-urban (sylvatic and
78 peri-urban) mosquito species from four countries representing three continents: Cambodia, the
79 Central African Republic, Madagascar, and French Guiana. These countries, due to their proximity to
80 the equator, contain high mosquito biodiversity (Foley et al., 2007) and have had long histories of
81 mosquito-borne virus circulation. Increased and continued surveillance of local mosquito species
82 could lead to valuable insights on mosquito virus biogeography. Using a unique score-based read
83 filtration strategy to remove interfering non-mosquito rRNA reads for accurate *de novo* assembly, we
84 produced a dataset of 234 novel full-length 28S and 18S rRNA sequences from 33 mosquito species,
85 30 of which have never been recorded before.

86 We also explored the functionality of 28S and 18S rRNA sequences as molecular barcodes by
87 comparing their performance to that of the mitochondrial *cytochrome* c *oxidase subunit I* (COI) gene
88 for molecular taxonomic and phylogenetic investigations. The COI gene is the most widely used DNA
89 marker for molecular species identification and forms the basis of the Barcode of Life Data System
90 (BOLD) (Hebert et al., 2003; Ratnasingham & Hebert, 2007). Presently, full-length rRNA sequences
91 are much less represented compared to other molecular markers. We hope that our sequence
92 dataset, with its species diversity and eco-geographical breadth, and the assembly strategy we
93 describe would further facilitate the use of rRNA as barcodes. In addition, this dataset enables the
94 design of species-specific oligos for cost-effective rRNA depletion for a broader range of mosquito
95 species and streamlined molecular species diagnostics during RNA-seq.

96 **RESULTS**

97 **Poor rRNA depletion using a non-specific depletion method**

98 During library preparations of mosquito samples for RNA-seq, routinely used methods for depleting
99 rRNA are commercial kits optimised for human or mice samples (Belda et al., 2019; Bishop-Lilly et al.,
100 2010; Chandler et al., 2015; N. Kumar et al., 2012; Weedall et al., 2015; Zakrzewski et al., 2018) or
101 through 80–100 base pair antisense probe hybridisation followed by ribonuclease digestion (Fauver et
102 al., 2019; Phelps et al., 2021). In cases where the complete reference rRNA sequence of the target
103 species is not known, oligos would be designed based on the rRNA sequence of the closest related
104 species (25, this study). These methods should deplete the conserved regions of rRNA sequences.
105 However, the variable regions remain at abundances high enough to compromise RNA-seq output. In
106 our hands, we have found that using probes designed for the *Ae. aegypti* rRNA sequence followed by
107 RNase H digestion according to the protocol published by Morlan *et al.* (2012) produced poor
108 depletion in *Ae. albopictus*, and in Culicine and Anopheline species (Figure 1A). Additionally, the lack
109 of reference rRNA sequences compromises the *in silico* clean-up of remaining rRNA reads from
110 sequencing data, as reads belonging to variable regions would not be removed. To solve this and to
111 enable RNA-seq metagenomics on a broader range of mosquito species, we performed RNA-seq to
112 generate reference rRNA sequences for 33 mosquito species representing 10 genera from
113 Cambodia, the Central African Republic, Madagascar, and French Guiana. Most of these species are
114 associated with vector activity for various pathogens in their respective ecologies (Table 1).



115

116 **Figure 1.** (**A**) Proportion of rRNA reads found in mosquito specimen pools of 5 individuals depleted by
117 probe hybridisation followed by RNase H digestion. Probes were antisense to *Ae. aegypti* rRNA

118 sequences. (**B**) Read vs. score ratio plot of a representative specimen “Ae simpsoni CS S27”. Green

119 line indicates 0.8 cut-off where only reads above this threshold are used in rRNA assembly.

120 **Table 1.** List of mosquito species represented in this study and their vector status.

Mosquito taxonomy*	Origin**	Collection site (ecosystem type)	Vector for***	Reference
Aedes (Fredardsius) vittatus	CF	rural (village)	DENV, ZIKV, CHIKV, YFV	(Diallo et al., 2020)
Aedes (Ochlerotatus) scapularis	GF	rural (village)	YFV	(Vasconcelos et al., 2001)
Aedes (Ochlerotatus) serratus	GF	rural (village)	YFV, OROV	(Cardoso et al., 2010; Gaillet et al., 2021)
Aedes (Stegomyia) aegypti	CF	urban	DENV, ZIKV, CHIKV, YFV	(Kraemer et al., 2019)
Aedes (Stegomyia) albopictus	CF, KH	rural (village, nature reserve)	DENV, ZIKV, CHIKV, YFV, JEV	(Auerswald et al., 2021; Kraemer et al., 2019)
Aedes (Stegomyia) simpsoni	CF	rural (village)	YFV	(Mukwaya et al., 2000)
Anopheles (Anopheles) baezai	KH	rural (nature reserve)	unreported	–
Anopheles (Anopheles) coustani	MG, CF	rural (village)	RVFV, malaria	(Mwangangi et al., 2013; Nepomichene et al., 2018; Ratovonjato et al., 2011)
Anopheles (Cellia) funestus	MG, CF	rural (village)	ONNV, malaria	(Lutomiah et al., 2013)
Anopheles (Cellia) gambiae	MG, CF	rural (village)	ONNV, malaria	(Brault et al., 2004)
Anopheles (Cellia) squamosus	MG	rural (village)	RVFV, malaria	(Ratovonjato et al., 2011; Stevenson et al., 2016)
Coquillettidia (Rhynchotaenia) venezuelensis	GF	rural (village)	OROV	(Gaillet et al., 2021)
Culex (Culex) antennatus	MG	rural (village)	RVFV	(Nepomichene et al., 2018; Ratovonjato et al., 2011)
Culex (Culex) duttoni	CF	rural (village)	unreported	–
Culex (Culex) neavei	MG	rural (village)	USUV	(Nikolay et al., 2011)
Culex (Culex) orientalis	KH	rural (nature reserve)	JEV	(Kim et al., 2015)
Culex (Culex) perexiguus	MG	rural (village)	WNV	(Vázquez González et al., 2011)
Culex (Culex) pseudovishnui	KH	rural (nature reserve)	JEV	(Auerswald et al., 2021; Maquart et al., 2021)
Culex (Culex) quinquefasciatus	MG, CF, KH	rural (village, nature reserve)	ZIKV, JEV, WNV, DENV, SLEV, RVFV, *Wuchereria bancrofti*	(Bhattacharya et al., 2016; Maquart et al., 2021; Ndiaye et al., 2016)
Culex (Culex) tritaeniorhynchus	MG, KH	rural (village, nature reserve)	JEV, WNV, RVFV	(Auerswald et al., 2021; Maquart et al., 2021)
Culex (Melanoconion) spissipes	GF	rural (village)	VEEV	(Weaver et al., 2004)
Culex (Melanoconion) portesi	GF	rural (village)	VEEV, TONV	(Talaga et al., 2021; Weaver et al., 2004)
Culex (Melanoconion) pedroi	GF	rural (village)	EEEV, VEEV, MADV	(Talaga et al., 2021; M. J. Turell et al., 2008)
Culex (Oculeomyia) bitaeniorhynchus	MG, KH	rural (village, nature reserve)	JEV	(Auerswald et al., 2021; Maquart et al., 2021)
Culex (Oculeomyia) poicilipes	MG	rural (village)	RVFV	(Ndiaye et al., 2016)
Eretmapodites intermedius	CF	rural (village)	unreported	–
Limatus durhamii	GF	rural (village)	ZIKV	(Barrio-Nuevo et al., 2020)

Mansonia (Mansonia) titillans	GF	rural (village)	VEEV, SLEV	(Hoyos-López et al., 2015; Michael J. Turell, 1999)
Mansonia (Mansonioides) indiana	KH	rural (nature reserve)	JEV	(Arunachalam et al., 2004)
Mansonia (Mansonioides) uniformis	MG, CF, KH	rural (village, nature reserve)	WNV, RVFV, *Wuchereria bancrofti*	(Lutomiah et al., 2013; Maquart et al., 2021; Ughasi et al., 2012)
Mimomyia (Etorleptiomyia) mediolineata	MG	rural (village)	unreported	–
Psorophora (Janthinosoma) ferox	GF	rural (village)	ROCV	(Mitchell et al., 1986)
Uranotaenia (Uranotaenia) geometrica	GF	rural (village)	unreported	–

* () indicates subgenus

** Origin countries are listed as their ISO alpha-2 codes: Central African Republic, CF; Cambodia, KH; Madagascar, MG; French Guiana, GF.

** dengue virus, DENV; Zika virus, ZIKV; chikungunya virus, CHIKV; Yellow Fever virus, YFV; Oropouche virus, OROV; Japanese encephalitis virus, JEV; Rift Valley Fever virus, RVFV; O'Nyong Nyong virus, ONNV; Usutu virus, USUV; West Nile virus, WNV; Saint Louis encephalitis virus, SLEV; Venezuelan equine encephalitis virus, VEEV; Tonate virus, TONV; Eastern equine encephalitis virus, EEEV; Madariaga virus, MADV; Rocio virus, ROCV.

rRNA reads filtering and sequence assembly

Assembling Illumina reads to reconstruct rRNA sequences from total mosquito RNA is not a straightforward task. Apart from host rRNA, total RNA samples also contain rRNA from other organisms associated with the host (microbiota, external parasites, or ingested diet). As rRNA sequences share high homology in conserved regions, Illumina reads (150 bp) from non-host rRNA can interfere with the contig assembly of host 28S and 18S rRNA.

Our score-based filtration strategy, described in detail in Methods, allowed us to bioinformatically remove interfering rRNA reads and achieve successful *de novo* assembly of 28S and 18S rRNA sequences for all our specimens. Briefly, for each Illumina read, we computed a ratio of BLAST scores against an Insecta library over a Non-Insecta library. Reads were segregated into four categories: (i) reads mapping only to the Insecta library, (ii) reads mapping better to the Insecta relative to Non-Insecta library, (iii) reads mapping better to the Non-Insecta relative to the Insecta library, and finally (iv) reads mapping only to the Non-Insecta library (Figure 1B, Supplementary Figure S1). By applying a conservative threshold of 0.8 to account for the non-exhaustiveness of the SILVA database, we removed reads that likely do not originate from mosquito rRNA. Notably, 15 of our specimens were engorged with vertebrate blood, a rich source of non-mosquito rRNA (Supplementary Table S1). The successful assembly of complete 28S and 18S rRNA sequences

146 demonstrates that this strategy performs as expected even with high amounts of non-host rRNA
147 reads. This is particularly important in studies on field-captured mosquitoes as females are often
148 sampled already having imbibed a blood meal or captured using the human landing catch technique.
149 We encountered challenges for three specimens morphologically identified as *Ma. africana*
150 (Specimen ID 33-35). COI amplification by PCR did not produce any product, hence COI barcoding
151 could not be used to confirm species identity. In addition, SPAdes was only able to assemble partial
152 length contigs, despite the high number of reads with high scores against the Insecta library. Among
153 other *Mansonia* specimens, the partial length contigs shared the highest similarity with contigs
154 obtained from sample "Ma uniformis CF S51". We then performed a guided assembly using the 28S
155 and 18S sequences of this specimen as references, which successfully produced full-length contigs.
156 In two of these specimens (Specimen ID 34 and 35), our assembly initially produced two sets of 28S
157 and 18S rRNA sequences, one of which was similar to mosquito rRNA with low coverage and another
158 with ten-fold higher coverage and 95% nucleotide sequence similarity to a water mite of genus
159 *Horreolanus* known to parasitize mosquitoes. Our success in obtaining rRNA sequences for mosquito
160 and water mite shows that our strategy can be applied to metabarcoding studies where the input
161 material comprises multiple insect species, provided that appropriate reference sequences of the
162 target species or of a close relative are available.
163 Altogether, we were able to assemble 122 28S and 114 18S full-length rRNA sequences for 33
164 mosquito species representing 10 genera sampled from four countries across three continents. This
165 dataset contains, to our knowledge, the first records for 30 mosquito species and for seven genera:
166 *Coquillettidia*, *Mansonia*, *Limatus*, *Mimomyia*, *Uranotaenia*, *Psorophora,* and *Eretmapodites*.
167 Individual GenBank accession numbers for these sequences and specimen information are listed in
168 Supplementary Table S1.

169 **Comparative phylogeny of novel rRNA sequences relative to existing records**

170 To verify the assembly accuracy of our rRNA sequences, we constructed a comprehensive
171 phylogenetic tree from the 28S rRNA sequences generated from our study alongside those publicly
172 available from GenBank (Figure 2). We applied a search criterion for GenBank sequences with at
173 least 95% coverage of our sequence lengths (~4000 bp), aiming to represent as many species or
174 genera as possible. Although we rarely found records for the same species included in our study, the
175 resulting tree showed that our 28S sequences generally clustered according to their respective

species and subgenera, supported by moderate to good bootstrap support at terminal nodes. Species taxa generally formed monophyletic clades, with the exception of *An. gambiae* and *Cx. quinquefasciatus*. *An. gambiae* 28S rRNA sequences formed a clade with closely related sequences from *An. arabiensis*, *An. merus*, and *An. coluzzii,* suggesting unusually high interspecies homology for Anophelines or other members of subgenus *Cellia.* Meanwhile, *Cx. quinquefasciatus* 28S rRNA sequences formed a taxon paraphyletic to sister species *Cx. pipiens*.

1 Cx perexiguus MG S60
1 Cx perexiguus MG S76
Cx perexiguus MT808440.1
1 Cx neavei MG S84
2 Cx neavei MG S84
1 Cx sp.4 MG S62
1 Cx antennatus MG S58
1 Cx antennatus MG S57
1 Cx antennatus MG S59
1 Cx sp.3 MG S61, S65
1 Cx sp.3 MG S63
1 Cx duttoni CF S52
1 Cx quinquefasciatus MG S75
1 Cx quinquefasciatus MG S74
Cx quinquefasciatus XR 005279053.1
Cx quinquefasciatus XR 005279034.1
Cx quinquefasciatus XR 005279033.1
Cx pipiens MT808425.1
Cx pipiens XR 005606618.1
1 Cx orientalis KH S14
1 Cx tritaeniorhynchus MG S115
1 Cx tritaeniorhynchus KH S23
1 Cx pseudovishnui KH S7
1 Cx pseudovishnui KH S8
1 Cx pseudovishnui KH S9
1 Cx tritaeniorhynchus KH S21
1 Cx tritaeniorhynchus KH S22
1 Cx bitaeniorhynchus KH S18
1 Cx bitaeniorhynchus KH S19, 2 S20
1 Cx bitaeniorhynchus KH S20
2 Cx poicilipes MG S82
1 Cx poicilipes MG S82
Cx theileri MT808435.1
Cx modestus MT808439.1
1 Cx sp.2 CF S48
1 Cx sp.2 CF S49
1 Cx sp.2 CF S50
1 Cx sp.1 KH S13
Cx impudicus MT808438.1
1 Cx portesi GF S95
1 Cx portesi GF S94
1 Cx portesi GF S96
1 Cx pedroi GF S107
1 Cx pedroi GF S108
1 Cx pedroi GF S106
1 Cx spissipes GF S98
1 Cx spissipes GF S99
1 Cx spissipes GF S97
Cu longiareolata MT808423.1
Cu annulata MT808429.1
1 Ma uniformis MG S77–S79
1 Ma uniformis KH S11, S15–S17
1 Ma uniformis CF S51
1 Ma indiana KH S10, S12
1 Ma sp.1 CF S33
1 Ma sp.3 CF S35
1 Ma sp.2 CF S34
2 Ma sp.2 CF S34
1 Ma sp.4 GF S103
1 Ma sp.4 GF S104
1 Ma titillans GF S105
Cq richiardii MT672341.1
Cq richiardii MT808427.1
Cq buxtoni MT808430.1
1 Cq venezuelensis GF S92
1 Cq venezuelensis GF S91
1 Cq venezuelensis GF S93
2 Cq venezuelensis GF S93
1 Ae scapularis GF S86
1 Ae scapularis GF S87
1 Ae scapularis GF S85
2 Ae scapularis GF S85
1 Ae serratus GF S88
1 Ae serratus GF S89
1 Ae serratus GF S90
Oc sticticus MW677941.1
Oc sticticus MT808442.1
Oc caspius MT808431.1
Oc detritus MT808422.1
Oc pulcritarsis MT808426.1
Ae geniculatus MT808437.1
Oc rusticus MT808432.1
1 Ae vittatus CF S31
1 Ae vittatus CF S30
1 Ae vittatus CF S32
Ae vexans MT808428.1
1 Ae aegypti CF S24
1 Ae aegypti CF S25
1 Ae aegypti CF S26
1 Ae simpsoni CF S27
1 Ae simpsoni CF S28
1 Ae simpsoni CF S29
1 Ae albopictus KH S3
1 Ae albopictus CF S36
Ae albopictus MT808421.1
Ae albopictus L22060.1
1 Ae albopictus CF S37
1 Ae albopictus CF S38
1 Ae albopictus KH S2
1 Ae albopictus KH S1
1 Er intermedius CF S54
1 Er intermedius CF S55
1 Er intermedius CF S56
1 Li durhamii GF S100–S102
1 Ps ferox GF S109
1 Ps ferox GF S111
1 Ps ferox GF S110
Ur unguiculata MT808444.1
1 Ur geometrica GF S112
1-2 Ur geometrica GF S114
1 Ur geometrica GF S113
1 Mi mediolineata MG S83
1 Mi sp.1 MG S64
1 An coustani CF S39, S41
1 An coustani CF S40
1 An coustani MG S66
1 An coustani MG S67
1 An baezai KH S4
1 An baezai KH S6
1 An baezai KH S5
An sacharovi MT808434.1
An messeae MW677939.1
An melanoon MT808424.1
An algeriensis MT808436.1
An claviger MT808441.1
An albimanus XR 004845806.1
An albimanus XR 004845855.1
An albimanus L78065.1
1 An funestus CF S44
1 An funestus MG S69, S70
1 An funestus CF S42
1 An funestus CF S43
An superpictus MT808443.1
An stephensi XR 004908154.1
An stephensi XR 004907438.1
1 An squamosus MG S68
An merus XR 006006278.1
An arabiensis XR 005740189.1
An arabiensis XR 005740497.1
An arabiensis XR 005740437.1
An merus XR 006005901.1
1 An gambiae CF S45
An merus XR 006005999.1
An coluzzii XR 005752429.1
An arabiensis XR 005740974.1
1 An gambiae CF S47
1 An gambiae MG S71
1 An gambiae CF S46
1 An gambiae MG S72, S73
2 An gambiae CF S45
An coluzzii XR 005752533.1
An coluzzii XR 005752529.1
An coluzzii XR 005752343.1
1 Ho sp.1 CF S34
2 Ho sp.2 CF S34
subgenus Culex
subgenus Oculeomyia
subgenus Melanoconion
subgenus Mansonoides
subgenus Coquillettidia
subgenus Rhynchotaenia
subgenus Ochlerotatus
subgenus Stegomyia
subgenus Anopheles
subgenus Nyssorhynchus
subgenus Cellia
0.05

182

183 **Figure 2.** Phylogenetic tree based on 28S sequences generated from this study and from GenBank
184 (3900 bp) as inferred using maximum-likelihood method and constructed to scale in MEGA X (S.
185 Kumar et al., 2018). Values at each node indicate bootstrap support (%) from 500 replications. For
186 sequences from this study, each specimen label contains information on taxonomy, origin (in 2-letter
187 country codes), and specimen ID. Labels in bold indicate GenBank sequences with accession
188 numbers shown. Label colours indicate genera: *Culex* in coral, *Anopheles* in purple, *Aedes* in dark
189 blue, *Mansonia* in dark green, *Culiseta* in maroon, *Limatus* in light green, *Coquillettidia* in light blue,
190 *Psorophora* in yellow, *Mimomyia* in teal, *Uranotaenia* in pink and *Eretmapodites* in brown. Scale bar
191 at 0.05 is shown.

192 28S sequence-based phylogenetic reconstructions (Figure 2, with GenBank sequences;
193 Supplementary Figure S3, this study only) showed marked incongruence to that of 18S sequences
194 (Figure 3). Although all rRNA trees show the bifurcation of *Culicidae* into subfamilies *Anophelinae* and
195 *Culicinae*, the recovered intergeneric phylogenetic relationships vary between the 28S and 18S trees
196 and are weakly supported. The 18S tree also exhibited several taxonomic anomalies: (i) the lack of
197 definitive clustering by species within the *Culex* subgenus (ii) the lack of distinction between 18S
198 sequences of *Cx*. *pseudovishnui* and *Cx*. *tritaeniorhynchus*; (iii) the placement of Ma sp. 3 CF S35
199 within a *Culex* clade; and (iv) the lack of a monophyletic *Mimomyia* clade. The topology of the 18S
200 tree suggests higher sequence divergence between the two *Cx. quinquefasciatus* taxa and between
201 the two *Mimomyia* taxa than in the 28S trees. However, 28S and 18S rRNA sequences are encoded
202 by linked loci in rDNA clusters and should not be analysed separately.

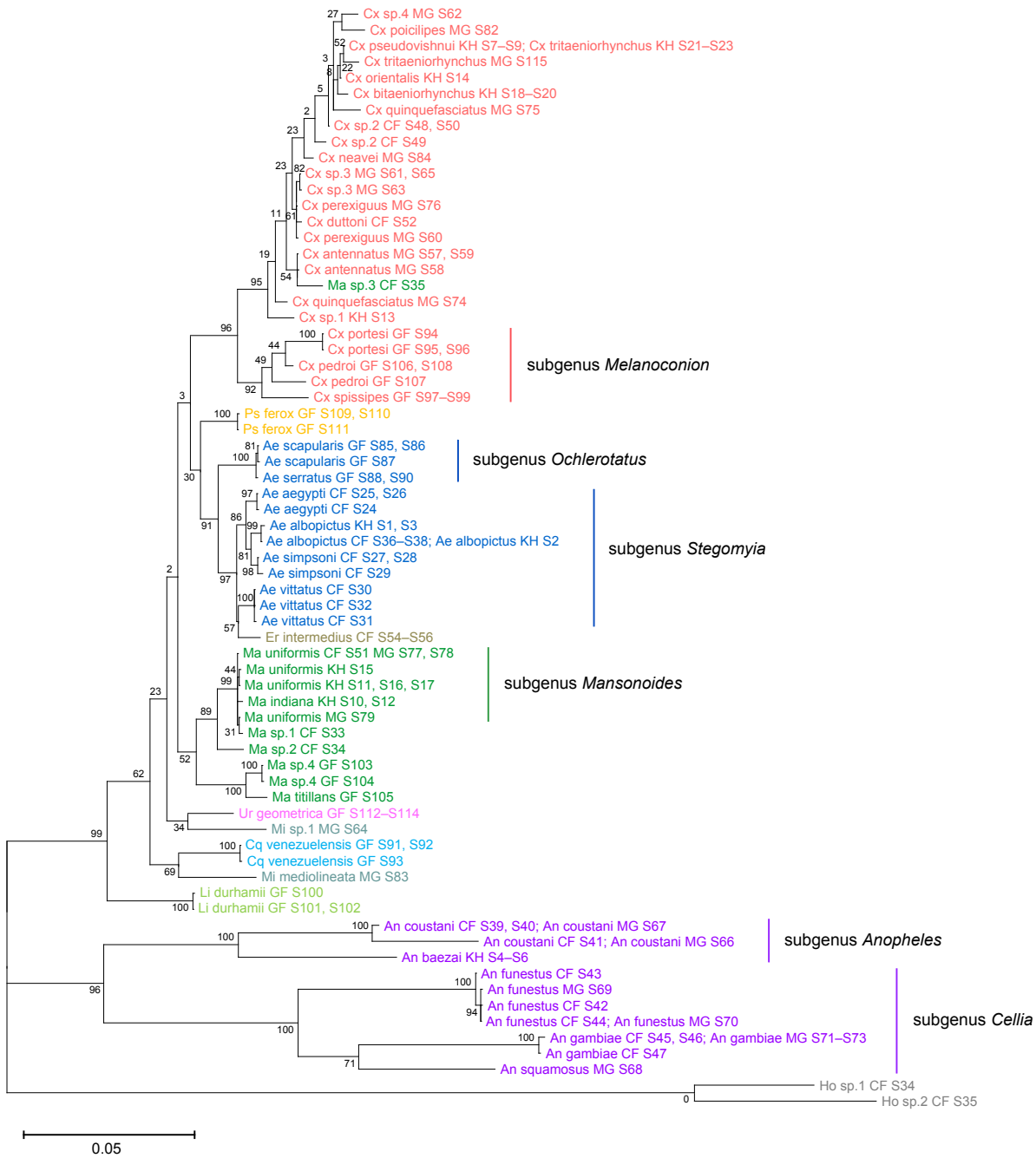


Figure 3. Phylogenetic tree based on 18S sequences (1900 bp) as inferred using maximum-likelihood method and constructed to scale in MEGA X (S. Kumar et al., 2018). Values at each node indicate bootstrap support (%) from 500 replications. Each specimen label contains information on taxonomy, origin (in 2-letter country codes), and specimen ID. Label colours indicate genera: *Culex* in coral, *Anopheles* in purple, *Aedes* in dark blue, *Mansonia* in dark green, *Limatus* in light green, *Coquillettidia* in light blue, *Psorophora* in yellow, *Mimomyia* in teal, *Uranotaenia* in pink and *Eretmapodites* in brown. Scale bar at 0.05 is shown.

211 Indeed, when concatenated 28S+18S rRNA sequences were generated from the same specimens
212 (Figure 4), the phylogenetic tree resulting from these sequences more closely resembles the 28S tree
213 (Figure 2) with regard to the basal position of the *Mimomyia* clade within the *Culicinae* subfamily with
214 good bootstrap support in either tree (84% in 28S tree, 100% in concatenated 28S+18S tree). For
215 internal nodes, bootstrap support values were higher in the concatenated tree compared to the 28S
216 tree. Interestingly, the 28S+18S tree formed an *Aedini* tribe-clade encompassing taxa from genera
217 *Psorophora, Aedes,* and *Eretmapodites*, possibly driven by the inclusion of 18S sequences.
218 Concatenation also resolved the anomalies found in the 18S tree and added clarity to the close
219 relationship between *Culex* and *Mansonia* taxa. Of note, relative to the 28S tree (Figure 2) the *Culex*
220 and *Mansonia* genera are no longer monophyletic in the concatenated 28S+18S tree (Figure 4).
221 Genus *Culex* is paraphyletic with respect to subgenus *Mansonoides* of genus *Mansonia* (Figure 2).
222 *Ma*. *titillans* and Ma sp. 4, which we suspect to be *Ma. pseudotitillans,* always formed a distinct branch
223 in 28S or 18S phylogenies, thus possibly representing a clade of subgenus *Mansonia*.
224 The concatenated 28S+18S tree recapitulates what is classically known about the systematics of
225 our specimens, namely (i) the early divergence of subfamily *Anophelinae* from subfamily *Culicinae*, (ii)
226 the division of genus *Anopheles* into two subgenera, *Anopheles* and *Cellia,* (iii) the division of genus
227 *Aedes* into subgenera *Stegomyia* and *Ochlerotatus*, (iv) the divergence of monophyletic subgenus
228 *Melanoconion* within the *Culex* genus (Harbach, 2007; Harbach & Kitching, 2016).

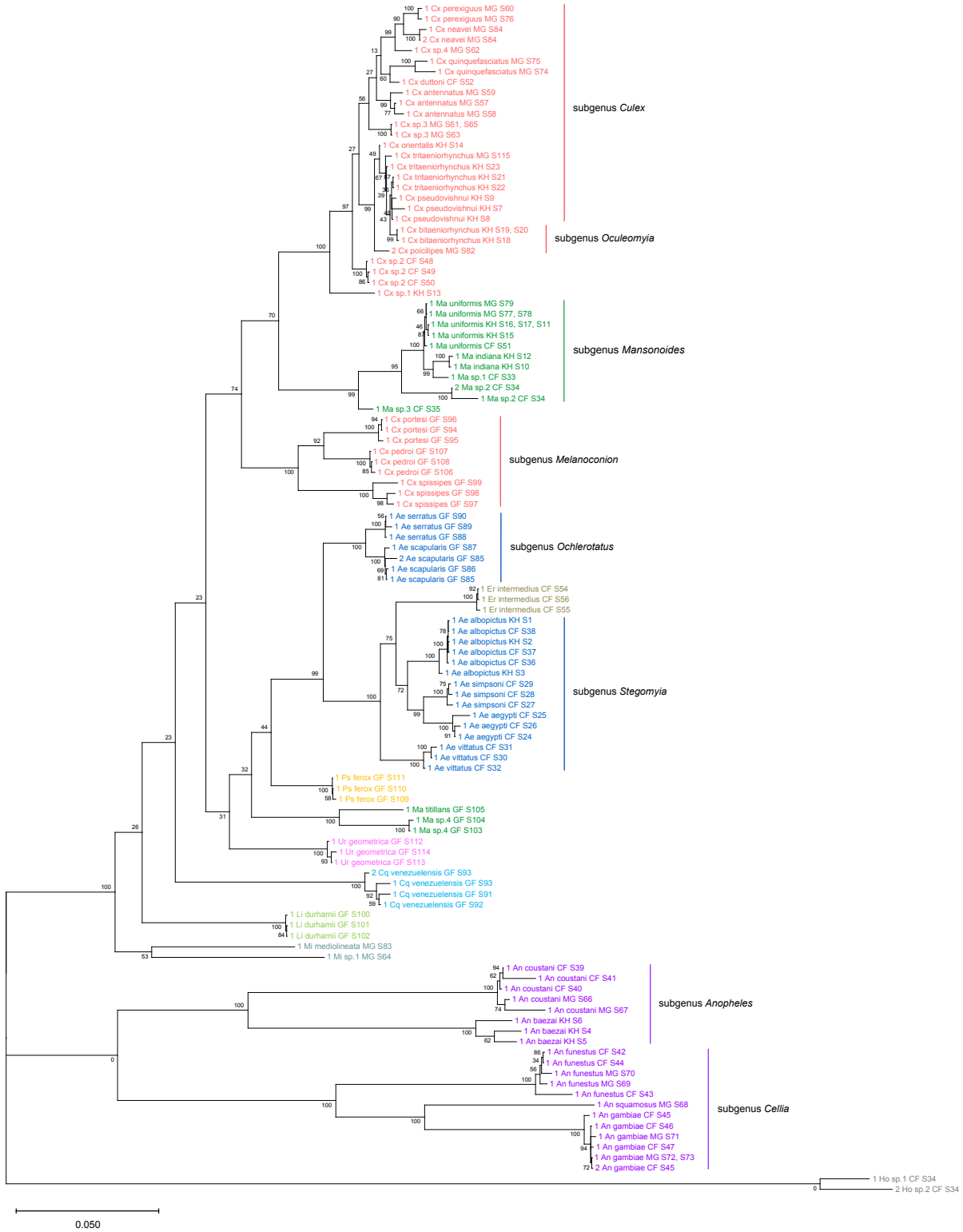


Figure 4. Phylogenetic tree based on concatenated 28S and 18S sequences generated from this study (3900+1900 bp) as inferred using maximum-likelihood method and constructed to scale in MEGA X (S. Kumar et al., 2018). Values at each node indicate bootstrap support (%) from 500 replications. For sequences from this study, each specimen label contains information on taxonomy, origin (as indicated in 2-letter country codes), and specimen ID. Label colours indicate genera: *Culex*

235 in coral, *Anopheles* in purple, *Aedes* in dark blue, *Mansonia* in dark green, *Culiseta* in maroon,
236 *Limatus* in light green, *Coquillettidia* in light blue, *Psorophora* in yellow, *Mimomyia* in teal, *Uranotaenia*
237 in pink and *Eretmapodites* in brown. Scale bar at 0.05 is shown.

238 **rRNA as a molecular marker for taxonomy and phylogeny**

239 We sequenced a 621 bp region of the COI gene to confirm morphological species identification of our
240 specimens and to compare the functionality of rRNA and COI sequences as molecular markers for
241 taxonomic and phylogenetic investigations. COI sequences were able to unequivocally determine the
242 species identity in most specimens except for the following cases. *An. coustani* COI sequences from
243 our study regardless of specimen origin shared remarkably high nucleotide similarity (>98%) with
244 several other *Anopheles* species such as *An. rhodesiensis*, *An. rufipes*, *An. ziemanni*, *An. tenebrosus*,
245 although *An. coustani* remained the most frequent and closest match. In the case of *Ae. simpsoni*,
246 three specimens had been morphologically identified as *Ae. opok* although their COI sequences
247 showed 97–100% similarity to that of *Ae. simpsoni.* As GenBank held no records of *Ae. opok* COI at
248 the time of this study, we instead aligned the putative *Ae. simpsoni* COI sequences against two sister
249 species of *Ae. opok*: *Ae. luteocephalus* and *Ae. africanus.* We found they shared only 90% and 89%
250 similarity, respectively. Given this significant divergence, we concluded these specimens to be *Ae.*
251 *simpsoni*. Ambiguous results were especially frequent among *Culex* specimens belonging to the *Cx.*
252 *pipiens* or *Cx. vishnui* subgroups, where the query sequence differed with either of the top two hits by
253 a single nucleotide. For example, between *Cx. quinquefasciatus* and *Cx. pipiens* of the *Cx. pipiens*
254 subgroup, and between *Cx. vishnui* and *Cx. tritaeniorhynchus* of the *Cx. vishnui* subgroup.

255 Among our three specimens of *Ma. titillans*, two appeared to belong to a single species that is
256 different but closely related to *Ma. titillans*. We surmised that these specimens could instead be *Ma.*
257 *pseudotitillans* based on morphological similarity but were not able to verify this by molecular means
258 as no COI reference sequence is available for this species. These specimens are hence putatively
259 labelled as “Ma sp.4”.

260 Phylogenetic reconstruction based on the COI sequences showed clustering of all species taxa
261 into distinct clades, underlining the utility of the COI gene in molecular taxonomy (Figure 5)(Hebert et
262 al., 2003; Ratnasingham & Hebert, 2007). However, species delineation among members of *Culex*
263 subgroups were not as clear cut, although sister species were correctly placed as sister taxa. This is

264 comparable to the 28S+18S tree (Figure 4) and is indicative of lower intraspecies distances relative to
265 interspecies distances.

266 To evaluate the utility of 28S and 18S rRNA sequences for molecular taxonomy, we used the
267 28S+18S rRNA tree to discern the identity of six specimens for which COI barcoding could not be
268 performed. These specimens include three unknown *Mansonia* species (Specimen ID 33–35), a *Ma.*
269 *uniformis* (Specimen ID 51), an *An. gambiae* (Specimen ID 47), and a *Ur. geometrica* (Specimen ID
270 113). Their positions in the 28S+18S rRNA tree relative to adjacent taxa confirms the morphological
271 identification of all six specimens to the genus level and, for three of them, to the species level (Figure
272 4).

273 The phylogenetic relationships indicated by the COI tree compared to the 28S+18S tree present
274 only a few points of congruence. COI-based phylogenetic inference indeed showed clustering of
275 generic taxa into monophyletic clades albeit with very weak bootstrap support, except for genera
276 *Culex* and *Mansonia* (Figure 5). Contrary to the 28S+18S tree (Figure 4), *Culex* subgenus
277 *Melanoconion* was depicted as a polyphyletic taxon with *Cx. spissipes* being a part of the greater
278 *Culicini* clade with members from subgenera *Oculeomyia* and *Culex* while *Cx. pedroi* and *Cx. portesi*
279 formed a distantly related clade. Among the *Mansonia* specimens, the two unknown Ma sp.4
280 specimens were not positioned as the nearest neighbours of *Ma. titillans* and instead appeared to
281 have diverged earlier from most of the other taxa from the *Culicidae* family. Notably, the COI
282 sequences of genus *Anopheles* is not basal to the other members of *Culicidae* and is instead shown
283 to be sister to Culex COI sequences (8% bootstrap support). This is a direct contrast to what is
284 suggested by the rRNA phylogenies (Figures 2–4), which suggests *Culex* rRNA sequences to be
285 among the most recently diverged. Bootstrap support for the more internal nodes of the COI trees
286 were remarkably low compared to those of rRNA-based trees.

287 In all rRNA trees, it is clear that the interspecific and intersubgeneric evolutionary distances within
288 the genus *Anopheles* are high relative to any other genera, indicating a greater degree of divergence.
289 This is evidenced by the longer branch lengths connecting Anopheline species-clades to the node of
290 the most recent common ancestor for subgenera *Anopheles* and *Cellia* (Figures 2-4, Supplementary
291 Figure S3). This feature is not evident in the COI tree, where the Anopheline interspecies distances
292 are comparable to those within the *Culex*, *Aedes*, and *Mansonia* taxa.

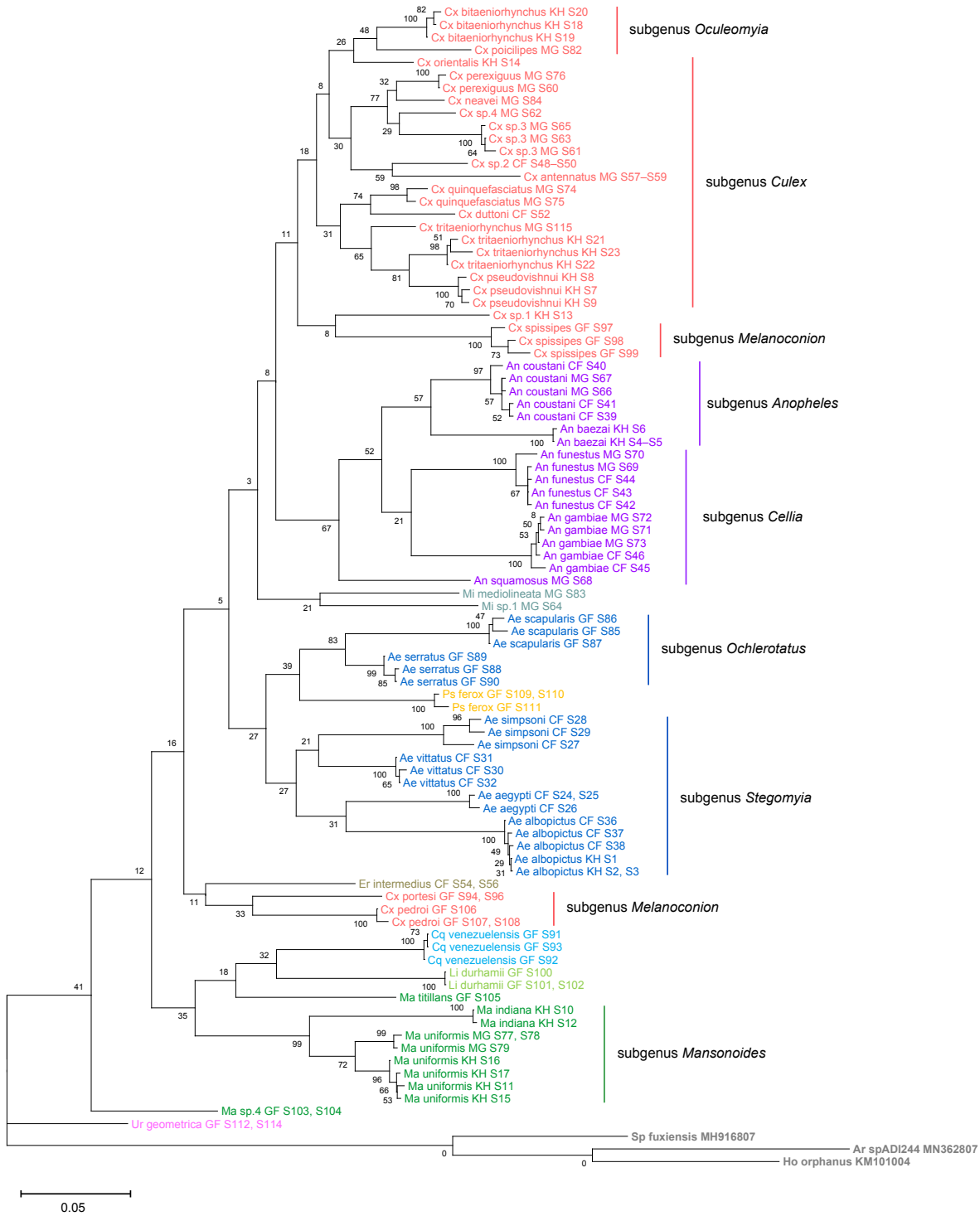


Figure 5. Phylogenetic tree based on COI sequences (621–699 bp) as inferred using maximum-likelihood method and constructed to scale in MEGA X (S. Kumar et al., 2018). Values at each node indicate bootstrap support (%) from 500 replications. Each specimen label contains information on taxonomy, origin (as indicated in 2-letter country codes), and specimen ID. Label colours indicate genera: *Culex* in coral, *Anopheles* in purple, *Aedes* in dark blue, *Mansonia* in dark green, *Limatus* in

299 light green, *Coquillettidia* in light blue, *Psorophora* in yellow, *Mimomyia* in teal, *Uranotaenia* in pink
300 and *Eretmapodites* in brown. Scale bar at 0.05 is shown.

301 **On *Culex* subgroups**

302 *Culex* (subgenus *Culex*) specimens of this study comprise several closely related sister species
303 belonging to the *Cx. vishnui* and *Cx. univittatus* subgroups, which are notoriously difficult to
304 differentiate based on morphology. Accordingly, in the 28S+18S rRNA (Figure 4) and COI (Figure 5)
305 trees these species and their known sister species were clustered together within the *Culex*
306 (subgenus *Culex*) clade: *Cx. tritaeniorhynchus* with *Cx. pseudovishnui* (*Cx. vishnui* subgroup); *Cx.*
307 *perexiguus* with *Cx. neavei* (*Cx. univittatus* subgroup).

308 The use of COI barcoding to distinguish between members of the *Culex* subgroups was limited.
309 For example, for the two *Cx. quinquefasciatus* samples in our taxonomic assemblage (Specimen ID
310 74 and 75), BLAST analyses of their COI sequences revealed they are a single nucleotide away from
311 *Cx. pipiens* or *Cx. quinquefasciatus* COI sequences (Supplementary Table S2). In the 28S rRNA tree
312 with GenBank sequences (Figure 2), two GenBank sequences of *Cx. pipiens* sequences formed a
313 clade sister to another containing three *Cx. quinquefasciatus* GenBank sequences and the "Cx
314 quinquefasciatus MG S74" sequence with 78% bootstrap support. This is in accordance with other
315 studies examining mitochondrial sequences (Sun et al., 2019) and morphological attributes (Harbach
316 et al., 2017). This shows that the 28S rRNA sequence can distinguish the two species and confirms
317 that "Cx quinquefasciatus MG S74" is indeed a *Cx. quinquefasciatus* specimen. However, "Cx
318 quinquefasciatus MG S75" is shown to be basal from other sequences within this *Cx. pipiens*
319 subgroup-clade with 100% bootstrap support. Given that *Cx. quinquefasciatus* and *Cx. pipiens* are
320 known to interbreed, it is plausible that this individual is a hybrid of the two species (Farajollahi et al.,
321 2011).

322 **DISCUSSION**

323 RNA-seq metagenomics on field-captured non-urban mosquitoes is a valuable tool for pre-empting
324 mosquito-borne virus emergences through surveillance and virus discovery. However, the lack of
325 reference rRNA sequences hinders good oligo-based depletion and efficient clean-up of RNA-seq
326 data. Additionally, *de novo* assembly of rRNA sequences is complicated due to regions that are highly
327 conserved across all distantly related organisms that could be present in a single specimen, i.e.,

microbiota, parasites, or vertebrate blood meal. Hence, we sought to establish a method to bioinformatically filter out non-host rRNA reads for the accurate assembly of novel 28S and 18S rRNA reference sequences.

We found that phylogenetic reconstructions based on 28S sequences or concatenated 28S+18S rRNA sequences were able to correctly cluster mosquito taxa according to species and corroborate current mosquito classification. This demonstrates that our bioinformatics methodology reliably generates bona fide 28S and 18S rRNA sequences, even in specimens parasitized by water mites or engorged with vertebrate blood. Further, we were able to use 28S+18S rRNA taxonomy for molecular species identification when COI sequences were unavailable or ambiguous, thus supporting the use of rRNA sequences as a barcode. They have the advantage of circumventing the need to additionally isolate and sequence DNA from specimens, as RNA-seq reads can be directly mapped against reference sequences. Post-depletion, in our hands there are sufficient numbers of remaining reads (5–10% of reads per sample) for assembly of complete rRNA contigs (unpublished data).

Phylogenetic inferences based on 28S or 18S rRNA sequences alone do not recover the same interspecific relationships. Relative to 28S sequences, we observed more instances where multiple specimens have near-identical 18S rRNA sequences. This can occur for specimens belonging to the same species, but also for conspecifics sampled from different geographic locations, such as *An. coustani*, *An. gambiae*, or *Ae. albopictus.* More rarely, specimens from the same species subgroup, such as *Cx. pseudovishnui* and *Cx. tritaeniorhynchus*, also shared 18S rRNA sequences. This was surprising given that the 18S rRNA sequences in our dataset is 1900 bp long. Concatenation of 28S and 18S rRNA sequences resolved this issue, enabling species delineation even among sister species of *Culex* subgroups, with which morphological identification meets its limits.

In Cambodia and other parts of Asia, the *Cx. vishnui* subgroup includes *Cx*. *tritaeniorhynchus*, *Cx*. *vishnui*, and *Cx*. *pseudovishnui*, which are important vectors of JEV (Maquart & Boyer, 2022). The former two were morphologically identified in our study but later revealed by COI barcoding to be a sister species. Discerning sister species of the *Cx. pipiens* subgroup is further complicated by interspecific breeding, with some populations showing genetic introgression to varying extents (Cornel et al., 2003). The seven sister species of this subgroup are practically indistinguishable based on morphology and require molecular methods to discern (Farajollahi et al., 2011; Zittra et al., 2016). Indeed, the 621 bp COI sequence amplified in our study did not contain enough nucleotide

358 divergence to allow clear identification, given that the COI sequence of *Cx. quinquefasciatus*
359 specimens differed from that of *Cx. pipiens* by a single nucleotide. Batovska et al. (2017) found that
360 even the Internal Transcribed Spacer 2 (ITS2) rDNA region, another common molecular marker,
361 could not differentiate the two species. Other DNA molecular markers such as nuclear Ace-2 or CQ11
362 genes (Aspen & Savage, 2003; Zittra et al., 2016) or *Wolbachia pipientis* infection status (Cornel et
363 al., 2003) are typically employed in tandem. In our study, 28S rRNA sequence-based phylogeny
364 (Figure 2) validated the identity of specimen "Cx quinquefasciatus MG S74" and suggested that
365 specimen "Cx quinquefasciatus MG S75" might have been a *pipiens-quinquefasciatus* hybrid. These
366 examples demonstrate how 28S rRNA sequences, concatenated with 18S or alone, contain enough
367 resolution to differentiate between *Cx. pipiens* and *Cx. quinquefasciatus*. rRNA-based phylogeny thus
368 allows for more accurate species identification and ecological observations in the context of disease
369 transmission. Additionally, tracing the genetic flow across hybrid populations within the *Cx. pipiens*
370 subgroup can inform estimates of vectorial capacity for each species. As only one or two members
371 from the *Cx. pipiens* and *Cx. vishnui* subgroups were represented in our taxonomic assemblage, an
372 explicit investigation including all member species of these subgroups in greater sample numbers is
373 warranted to further test the degree of accuracy with which 28S and 18S rRNA sequences can
374 delineate sister species.
375 Our study included French Guianese *Culex* species *Cx. spissipes* (group Spissipes), *Cx. pedroi*
376 (group Pedroi), and *Cx. portesi* (group Vomerifer). These species belong to the New World subgenus
377 *Melanoconion*, section Spissipes, with well-documented distribution in North and South Americas
378 (Sirivanakarn, 1982) and are vectors of encephalitic alphaviruses EEEV and VEEV among others
379 (Talaga et al., 2021; M. J. Turell et al., 2008; Weaver et al., 2004). Indeed, our rooted rRNA and COI
380 trees showed the divergence of the three *Melanoconion* species from the major *Culex* clade
381 comprising species broadly found across Africa and Asia (Auerswald et al., 2021; Farajollahi et al.,
382 2011; Nchoutpouen et al., 2019; Takhampunya et al., 2011). The topology of the concatenated
383 28S+18S tree places the *Cx. portesi* and *Cx. pedroi* species-clades as sister groups (92% bootstrap
384 support), with *Cx. spissipes* as a basal group within the *Melanoconion* clade (100% bootstrap
385 support). This corroborates the systematics elucidated by Navarro and Weaver (2004) using the ITS2
386 marker, and those by Sirivanakarn (1982) and Sallum and Forattini (1996) based on morphology.
387 Curiously, in the COI tree, *Cx. spissipes* sequences were clustered with unknown species Cx. sp1,

388 forming a clade sister to another containing other *Culex* (*Culex*) and *Culex* (*Oculeomyia*) species,
389 albeit with very low bootstrap support. Previous phylogenetic studies based on the COI gene have
390 consistently placed *Cx. spissipes* or the Spissipes group basal to other groups within the
391 *Melanoconion* subgenus (Torres-Gutierrez et al., 2016, 2018). However, these studies contain only
392 *Culex* (*Melanoconion*) species in their assemblage, apart from *Cx. quinquefasciatus* to act as an
393 outgroup. This clustering of *Cx. spissipes* with non-*Melanoconion* species in our COI phylogeny could
394 be an artefact of a much more diversified assemblage rather than a true phylogenetic link.
395 Taking advantage of our multi-country sampling, we examined whether rRNA or COI phylogeny
396 can be used to discriminate conspecifics originating from different geographies. Our assemblage
397 contains five of such species: *An. coustani*, *An. funestus*, *An. gambiae*, *Ae. albopictus*, and *Ma.*
398 *uniformis*. Among the rRNA trees, the concatenated 28S+18S tree and 28S tree were able to
399 discriminate between *Ma. uniformis* specimens from Madagascar, Cambodia, and the Central African
400 Republic, and between *An. coustani* specimens from Madagascar and the Central African Republic
401 (100% bootstrap support). In the COI tree, only *Ma. uniformis* was resolved into geographical clades
402 comprising specimens from Madagascar and specimens from Cambodia (72% bootstrap support). No
403 COI sequence was obtained from one *Ma. uniformis* from the Central African Republic. The 28S+18S
404 rRNA sequences ostensibly provided more population-level genetic information than COI sequences
405 alone with better support. The use of rRNA sequences in investigating the biodiversity of mosquitoes
406 should therefore be explored with a more comprehensive taxonomic assemblage.
407 The phylogenetic reconstructions based on rRNA or COI sequences in our study are hardly
408 congruent, but two principal differences stand out. First, the COI phylogeny does not recapitulate the
409 early divergence of *Anophelinae* from *Culicinae* (Figure 5). This is at odds with other studies
410 estimating mosquito divergence times based on mitochondrial genes (Logue et al., 2013; Lorenz et
411 al., 2021) or nuclear genes (Reidenbach et al., 2009). The second notable feature in the rRNA trees is
412 the remarkably large interspecies and intersubgeneric evolutionary distances within genus *Anopheles*
413 relative to *Culicinae* genera (Figures 2–4, Supplementary Figure S4) but this is not apparent in the
414 COI tree. The hyperdiversity among *Anopheles* taxa may be attributed to the earlier diversification of
415 the *Anophelinae* subfamily in the early Cretaceous period compared to that of the *Culicinae* subfamily,
416 a difference of at least 40 million years (Lorenz et al., 2021). The differences in rRNA and COI tree
417 topologies indicate a limitation in using COI alone to determine evolutionary relationships. Importantly,

418 drawing phylogenetic conclusions from short DNA barcodes such as COI has been cautioned against
419 due to its weak phylogenetic signal (Hajibabaei et al., 2006). The relatively short length of our COI
420 sequences (621–699 bp) combined with the 100-fold higher nuclear substitution rate of mitochondrial
421 genomes relative to nuclear genomes (Arctander, 1995) could result in homoplasy (Danforth et al.,
422 2005), making it difficult to clearly discern ancestral sequences and correctly assign branches into
423 lineages, as evidenced by the poor nodal bootstrap support at genus-level branches. Indeed, in the
424 study by Lorenz et al. (2021), a phylogenetic tree constructed using a concatenation of all 13 protein-
425 coding genes of the mitochondrial genome was able to resolve ancient divergence events. This
426 affirms that while COI sequences can be used to reveal recent speciation events, longer or multi-gene
427 molecular markers are necessary for studies into deeper evolutionary relationships (Danforth et al.,
428 2005).

429 In contrast to Anophelines where 28S rRNA phylogenies illustrated higher interspecies divergence
430 compared to COI phylogeny, two specimens of an unknown *Mansonia* species, “Ma sp.4 GF S103”
431 and “Ma sp.4 GF S104”, provided an example where interspecies relatedness based on their COI
432 sequences is greater than that based on their rRNA sequences in relation to “Ma titillans GF S105”.
433 While all rRNA trees (Figure 2–4) placed “Ma titillans GF S105” as a sister taxon with 100% bootstrap
434 support, the COI tree placed Ma sp.4 basal to all other species except *Ur. geometrica*. This may hint
435 at a historical selective sweep in the mitochondrial genome, whether arising from geographical
436 separation, mutations, or linkage disequilibrium with inherited symbionts (Hurst & Jiggins, 2005),
437 resulting in the disparate mitochondrial haplogroups found in French Guyanese Ma sp.4 and *Ma.*
438 *titillans*. In addition, both haplogroups are distant from those associated with members of subgenus
439 *Mansonoides*. To note, the COI sequences of “Ma sp.4 GF S103” and “Ma sp.4 GF S104” share
440 87.12 and 87.39% nucleotide similarity, respectively, to that of “Ma titillans GF S105”. Interestingly,
441 the endosymbiont *Wolbachia pipientis* has been detected in *Ma. titillans* sampled from Brazil (De
442 Oliveira et al., 2015), which may contribute to the divergence of “Ma titillans GF S105” COI sequence
443 away from those of Ma sp.4. This highlights other caveats of using a mitochondrial DNA marker in
444 determining evolutionary relationships (Hurst & Jiggins, 2005), which nuclear markers such as 28S
445 and 18S rRNA sequences may be immune to.

446 **Conclusions**

447 RNA-seq metagenomics is a valuable tool for surveillance and virus discovery in non-urban
448 mosquitoes but it is impeded by the lack of full-length rRNA reference sequences. Here we presented
449 a rRNA sequence assembly strategy and 234 newly generated 28S and 18S mosquito rRNA
450 sequences. Our work has expanded the current rRNA reference library by providing, to our
451 knowledge, the first full-length rRNA records for 30 species in public databases and paves the way for
452 the assembly of many more. These novel rRNA sequences can improve mosquito RNA-seq
453 metagenomics by expanding reference sequence data for the optimization of species-specific oligo-
454 based rRNA depletion protocols, for streamlined species identification by rRNA barcoding and for
455 improved RNA-seq data clean-up. In addition, RNA barcodes could serve as an additional tool for
456 mosquito taxonomy and phylogeny although further studies are necessary to reveal how they
457 compare against other nuclear or mitochondrial DNA marker systems (Batovska et al., 2017; Beebe,
458 2018; Behura, 2006; Ratnasingham & Hebert, 2007; Reidenbach et al., 2009; Vezenegho et al.,
459 2022).

460 We showed that phylogenetic inferences from a tree based on 28S rRNA sequences alone or
461 concatenated 28S +18S rRNA sequences largely agree with contemporary mosquito classification
462 and can be used for species diagnostics given a reference sequence. In analysing the same set of
463 specimens by COI or rRNA sequences, we found deep discrepancies in phylogenetic inferences. We
464 conclude that while COI-based phylogeny can reveal recent speciation events, rRNA sequences may
465 be better suited for investigations of deeper evolutionary relationships as they are less prone to
466 selective sweeps and homoplasy.

467 **MATERIALS AND METHODS**

468 **Sample collection**

469 Mosquito specimens were sampled from 2019 to 2020 by medical entomology teams from the Institut
470 Pasteur de Bangui (Central African Republic, Africa; CF), Institut Pasteur de Madagascar
471 (Madagascar, Africa; MG), Institut Pasteur du Cambodge (Cambodia, Asia; KH), and Institut Pasteur
472 de la Guyane (French Guiana, South America; GF). Adult mosquitoes were sampled using several
473 techniques including CDC light traps, BG sentinels, and human-landing catches. Sampling sites are
474 non-urban locations including rural settlements in the Central African Republic, Madagascar, and
475 French Guiana and national parks in Cambodia. Mosquitoes were identified using morphological

476 identification keys on cold tables before preservation by flash freezing in liquid nitrogen and
477 transportation in dry ice to Institut Pasteur Paris for analysis. A list of the 112 mosquito specimens
478 included in our taxonomic assemblage and their related information are provided in Supplementary
479 Table S1.

480 **RNA and DNA isolation**

481 Nucleic acids were isolated from mosquito specimens using TRIzol reagent according to
482 manufacturer's protocol (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA). Single
483 mosquitoes were homogenised into 200 µL of TRIzol reagent and other of the reagents within the
484 protocol were volume-adjusted accordingly. Following phase separation, RNA were isolated from the
485 aqueous phase while DNA were isolated from the remaining interphase and phenol-chloroform phase.
486 From here, RNA is used to prepare cDNA libraries for next generation sequencing while DNA is used
487 in PCR amplification and Sanger sequencing of the mitochondrial *cytochrome* c *oxidase subunit I*
488 (COI) gene as further described below.

489 **Probe depletion of rRNA**

490 We tested a selective rRNA depletion protocol by Morlan *et al.* (2012) on several mosquito species
491 from the *Aedes*, *Culex*, and *Anopheles* genera. We designed 77 tiled 80 bp DNA probes antisense to
492 the *Ae. aegypti* 28S, 18S, and 5.8S rRNA sequences. A pool of probes at a concentration of 0.04 µM
493 were prepared. To bind probes to rRNA, 1 µL of probes and 2 µL of Hybridisation Buffer (100 mM
494 Tris-HCl and 200 mM NaCl) was added to rRNA samples to a final volume of 20 µL and subjected to
495 a slow-cool incubation starting at 95 ºC for 2 minutes, then cooling to 22 ºC at a rate of 0.1 ºC per
496 second, ending with an additional 5 minutes at 22 ºC. The resulting RNA:DNA hybrids were treated
497 with 2.5 µL Hybridase™ Thermostable RNase H (Epicentre, Illumina, Madison, Wisconsin, USA) and
498 incubated at 37 ºC for 30 minutes. To remove DNA probes, the mix was treated with 1 µL DNase I
499 (Invitrogen) and purified with Agencourt RNAClean XP Beads (Beckman Coulter, Brea, California,
500 USA). The resulting RNA is used for total RNA sequencing to check depletion efficiency.

501 **Total RNA sequencing**

502 To obtain rRNA sequences, RNA samples were quantified on a Qubit Fluorometer (Invitrogen) using
503 the Qubit RNA BR Assay kit (Invitrogen) for concentration adjustment. Non-depleted total RNA was
504 used for library preparation for next generation sequencing using the NEBNext Ultra II RNA Library

505 Preparation Kit for Illumina (New England Biolabs, Ipswich, Massachusetts, USA) and the NEBNext
506 Multiplex Oligos for Illumina (Dual Index Primers Set 1) (New England Biolabs). Sequencing was
507 performed on a NextSeq500 sequencing system (Illumina, San Diego, California, USA). Quality
508 control of fastq data and trimming of adapters were performed with FastQC and cutadapt,
509 respectively.

510 **28S and 18S rRNA assembly**

511 To obtain 28S and 18S rRNA contigs, we had to first clean our fastq library by separating the reads
512 representing mosquito rRNA from all other reads. To achieve this, we used the SILVA RNA sequence
513 database to create 2 libraries: one containing all rRNA sequences recorded under the "Insecta" node
514 of the taxonomic tree, the other containing the rRNA sequences of many others nodes distributed
515 throughout the taxonomic tree, hence named “Non-Insecta” (Quast et al., 2013). Each read was
516 aligned using the nucleotide Basic Local Alignment Search Tool (BLASTn,
517 **https://blast.ncbi.nlm.nih.gov/**) of the National Center for Biotechnology Information (NCBI) against
518 each of the two libraries and the scores of the best high-scoring segment pairs from the two BLASTns
519 are subsequently used to calculate a ratio of Insecta over Non-Insecta scores (Altschul et al., 1990).
520 Only reads with a ratio greater than 0.8 were used in the assembly. The two libraries being non-
521 exhaustive, we chose this threshold of 0.8 to eliminate only reads that were clearly of a non-insect
522 origin. Selected reads were assembled with the SPAdes assembler using the “-rna” option, allowing
523 more heterogeneous coverage of contigs and kmer lengths of 31 , 51 and 71 bases (Bankevich et al.,
524 2012). This method successfully assembled rRNA sequences for all specimens, including a parasitic
525 *Horreolanus* water mite (122 sequences for 28S and 114 sequences for 18S).

526 Initially, our filtration technique had two weaknesses. First, there is a relatively small number of
527 complete rRNA sequences in the Insecta library from SILVA. To compensate for this, we carried out
528 several filtration cycles, each time adding in the complete sequences produced in previous cycles to
529 the Insecta library. Second, when our mosquito specimens were parasitized by other insects, it was
530 not possible to bioinformatically filter out rRNA reads belonging to the parasite. For these rare cases,
531 we used the " --trusted-contigs" option of SPAdes, giving it access to the 28S and 18S sequences of
532 the mosquito closest in terms of taxonomic distance. By doing this, the assembler was able to
533 reconstruct the rRNA of the mosquito as well as the rRNA of the parasitizing insect. All assembled
534 rRNA sequences from this study have been deposited in GenBank with accession numbers

535 OM350214–OM350327 for 18S rRNA sequences and OM542339–OM542460 for 28S rRNA
536 sequences.

537 **COI amplicon sequencing**

538 The mitochondrial COI gene was amplified from DNA samples using the universal "Folmer" primer set
539 LCO1490 (5'- GGTCAACAAATCATAAAGATATTGG -3') and HCO2198 (5'-
540 TAAACTTCAGGGTGACCAAAAAATCA-3'), as per standard COI barcoding practices, producing a
541 658 bp product (Folmer et al., 1994). PCRs were performed using Phusion High-Fidelity DNA
542 Polymerase (Thermo Fisher Scientific). Every 50 µL reaction contained 10 µL of 5X High Fidelity
543 buffer, 1 µL of 10 mM dNTPs, 2.5 µL each of 10 mM forward (LCO1490) and reverse (HCO2198)
544 primer, 28.5 µL of water, 5 µL of DNA sample, and 0.5 µL of 2 U/µL Phusion DNA polymerase. A 3-
545 step cycling incubation protocol was used: 98 ºC for 30 seconds; 35 cycles of 98 ºC for 10 seconds,
546 60 ºC for 30 seconds, and 72 ºC for 15 seconds; 72 ºC for 5 minutes ending with a 4 ºC hold. PCR
547 products were size-verified using gel electrophoresis and then gel-purified using the QIAquick Gel
548 Extraction Kit (Qiagen, Hilden, Germany). Sanger sequencing of the COI amplicons were performed
549 by Eurofins Genomics, Ebersberg, Germany.

550 **COI sequence analysis**

551 Forward and reverse COI DNA sequences were end-trimmed to remove bases of poor quality (Q
552 score < 30). At the 5' ends, sequences were trimmed at the same positions such that all forward
553 sequences start with 5'- TTTTGG and all reverse sequences start with 5'- GGNTCT. Forward and
554 reverse sequences were aligned using BLAST to produce a 621 bp consensus sequence. In cases
555 where good quality sequences extends beyond 621 bp, forward and reverse sequences were
556 assembled using Pearl (**https://www.gear-genomics.com/pearl/**) and manually checked for errors
557 against trace files (Rausch et al., 2019, 2020). We successfully assembled a total of 106 COI
558 sequences. All assembled COI sequences from this study have been deposited in GenBank with
559 accession numbers OM630610–OM630715.

560 **COI validation of morphology-based species identification**

561 We analysed assembled COI sequences with BLASTn against the nucleotide collection (nr/nt)
562 database to confirm morphology-based species identification. BLAST analyses revealed 32 cases
563 where top hits indicated a different species identity, taking <95% nucleotide sequence similarity as the

564 threshold to delineate distinct species (Supplementary Table S2). In these cases, the COI sequence
565 of the specimen was then BLAST-aligned against a GenBank record representing the morphological
566 species to verify that the revised identity is a closer match by a significant margin, i.e., more than 2%
567 nucleotide sequence similarity. All species names reported hereafter reflect identities determined by
568 COI barcoding except for cases where COI-based identities were ambiguous, in which case
569 morphology-based identities were retained. In cases where matches were found within a single genus
570 but of multiple species, specimens were indicated as an unknown member of their genus (e.g., *Culex*
571 sp.). Information of the highest-scoring references for all specimens, including details of ambiguous
572 BLASTn results, are recorded in Supplementary Table S2.
573 Within our COI sequences, we found six unidentified *Culex* species (including two that matched to
574 GenBank entries identified only to the genus level), four unidentified *Mansonia* species, and one
575 unidentified *Mimomyia* species. For *An. baezai*, no existing GenBank records were found at the time
576 this analysis was performed.

577 **Phylogenetic analysis**

578 Multiple sequence alignment (MSA) were performed on assembled COI and rRNA sequences using
579 the MUSCLE software (Supplementary Files 1–4) (Edgar, 2004; Madeira et al., 2019). As shown in
580 Supplementary Figure S2, the 28S rRNA sequences contain many blocks of highly conserved
581 nucleotides, which makes the result of multiple alignment particularly obvious. We therefore did not
582 test other alignment programs. The multiple alignment of the COI amplicons is even more evident
583 since no gaps are necessary for this alignment.
584 Phylogenetic tree reconstructions were performed with the MEGA X software using the maximum-
585 likelihood method (S. Kumar et al., 2018). Default parameters were used with bootstrapping with 500
586 replications to quantify confidence level in branches. For rRNA trees, sequences belonging to an
587 unknown species of parasitic water mite (genus *Horreolanus*) found in our specimens served as an
588 outgroup taxon. In addition, we created and analysed a separate dataset combining our 28S rRNA
589 sequences and full-length 28S rRNA sequences from GenBank totalling 169 sequences from 58
590 species (12 subgenera). To serve as outgroups for the COI tree, we included sequences obtained
591 from GenBank of three water mite species, *Horreolanus orphanus* (KM101004), *Sperchon fuxiensis*
592 (MH916807), and *Arrenurus* sp. (MN362807).

593 **DECLARATIONS**

594 **Availability of data and materials**

595 Raw RNA-seq fastq sequence data are available from the corresponding author upon reasonable
596 request. Multiple sequence alignment files are included as supplementary files. All sequences
597 generated in this study have been deposited in GenBank under the accession numbers OM350214–
598 OM350327 for 18S rRNA sequences, OM542339–OM542460 for 28S rRNA sequences, and
599 OM630610–OM630715 for COI sequences (Supplementary Table S1).

600 **Competing interests**

601 The authors declare that they have no financial or non-financial competing interests.

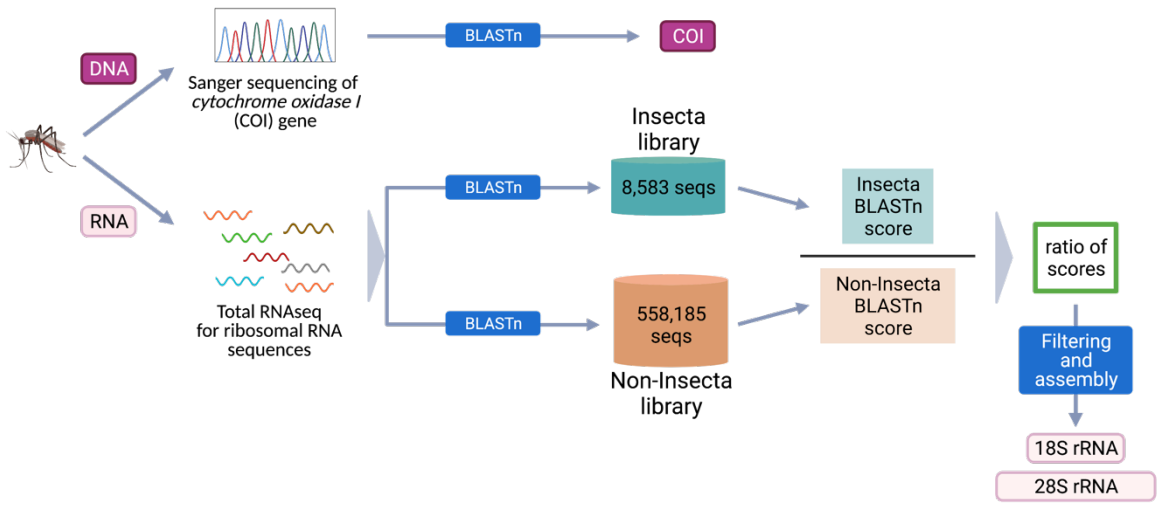
602 **Funding**

603 This work was supported by the Defence Advanced Research Projects Agency PREEMPT program
604 managed by Dr. Rohit Chitale and Dr. Kerri Dugan [Cooperative Agreement HR001118S0017] (the
605 content of the information does not necessarily reflect the position or the policy of the U.S.
606 government, and no official endorsement should be inferred).

607 **Acknowledgements**

608 We thank members of the Saleh lab for valuable discussions and to Dr Louis Lambrechts for critical
609 reading of the manuscript. We especially thank all medical entomology staff of IP Bangui, IP
610 Cambodge (Sony Yean, Kimly Heng, Kalyan Chhuoy, Sreynik Nhek, Moeun Chhum, Kimhuor Sour
611 and Pierre-Olivier Maquart), IP Madagascar, and IP Guyane for assistance in field missions,
612 laboratory work, and logistics, and Inès Partouche from IP Paris for laboratory assistance. We are
613 also grateful to Dr Catherine Dauga for advice on phylogenetic analyses, and to Amandine Guidez for
614 providing a French Guiana-specific COI reference library.

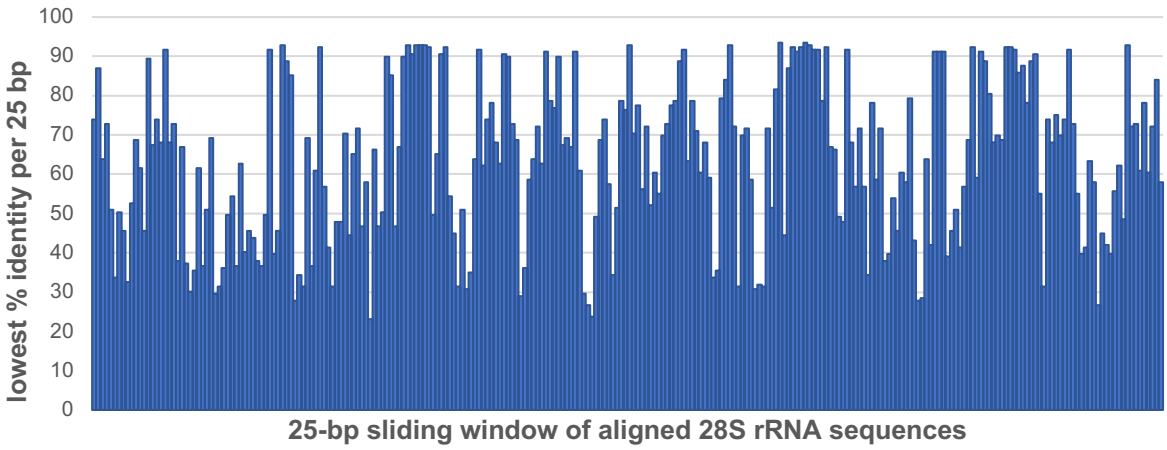
615 **SUPPLEMENTARY FILES**



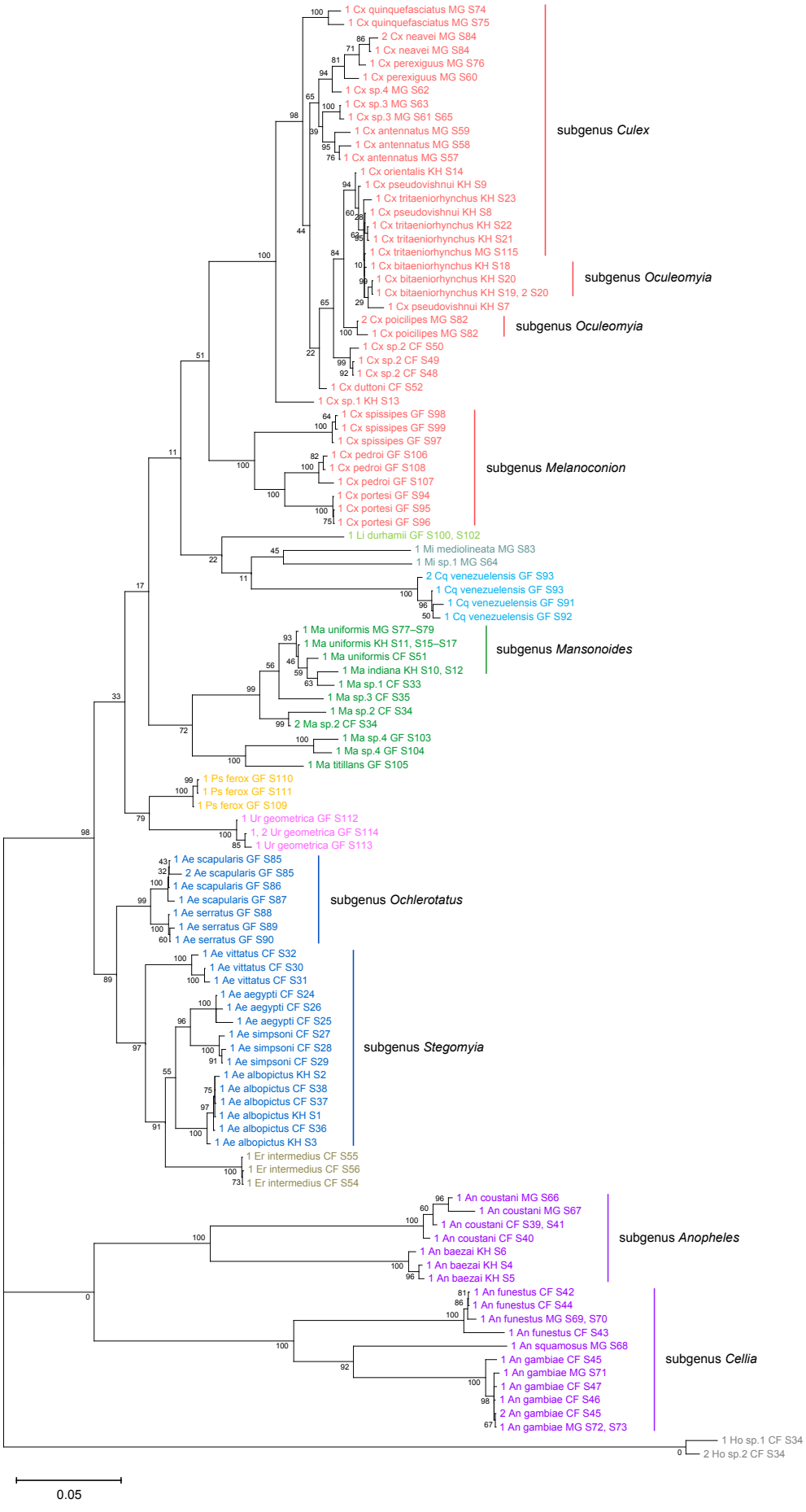
Supplementary Figure S1. Study workflow from specimens to sequences.

Supplementary Table S1. Taxonomic and sampling information on mosquito specimens and associated accession numbers of their 28S, 18S, and COI sequences (XLSX).

Supplementary Table S2. COI sequence BLAST analyses summary (XLSX).

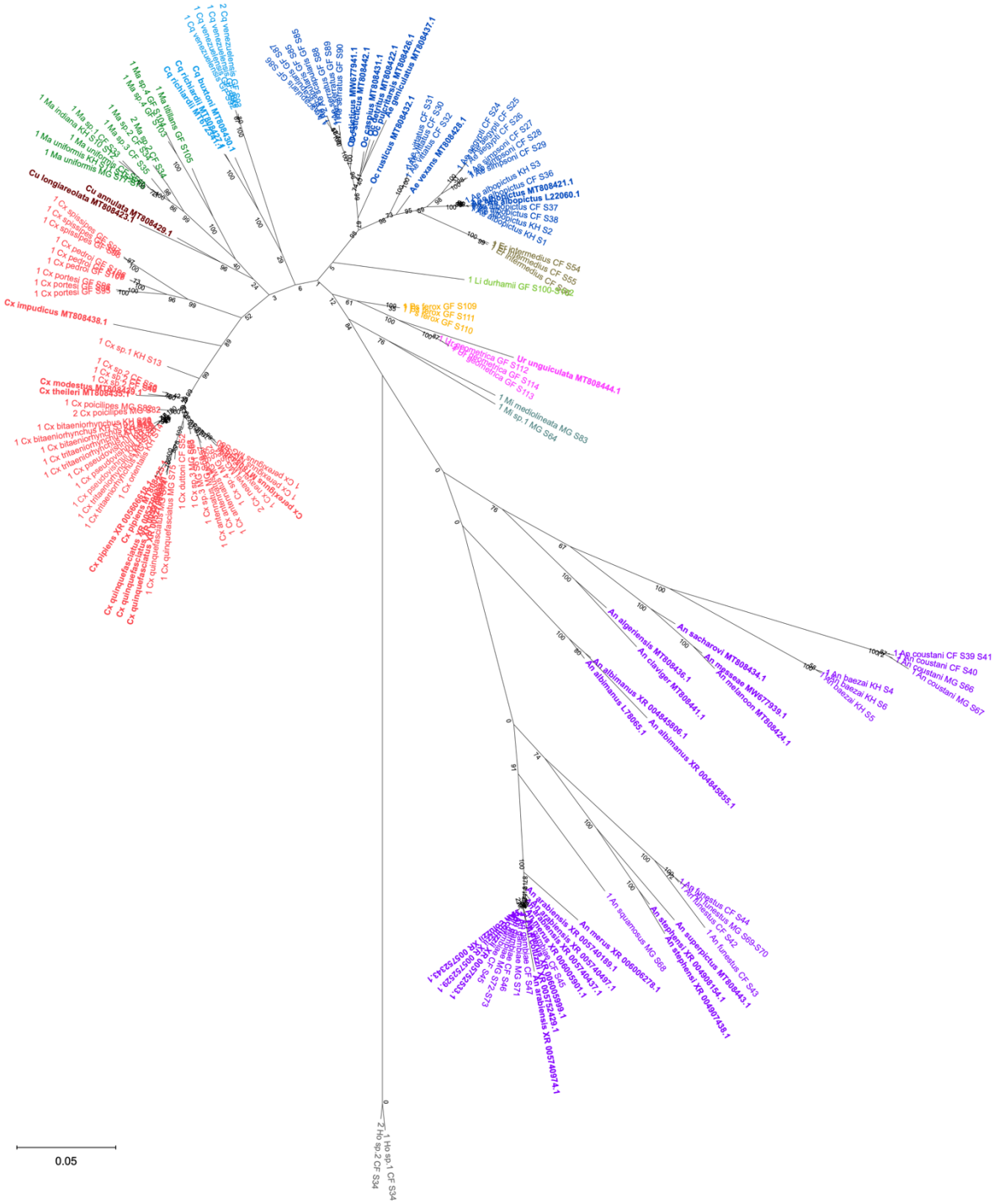


Supplementary Figure S2. Sequence conservation among 169 28S rRNA sequences obtained from this study and from GenBank combined. Multiple sequence alignment was performed on 28S rRNA sequences, 3900 bp in length. Each bar represents a 25-bp sliding window of the 28S rRNA sequence alignment where the values are the lowest percentage identity found.


1 Cx quinquefasciatus MG S74
1 Cx quinquefasciatus MG S75
2 Cx neavei MG S84
1 Cx neavei MG S84
1 Cx perexiguus MG S76
1 Cx perexiguus MG S60
1 Cx sp.4 MG S62
1 Cx sp.3 MG S63
1 Cx sp.3 MG S61 S65
1 Cx antennatus MG S59
1 Cx antennatus MG S58
1 Cx antennatus MG S57
1 Cx orientalis KH S14
1 Cx pseudovishnui KH S9
1 Cx tritaeniorhynchus KH S23
1 Cx pseudovishnui KH S8
1 Cx tritaeniorhynchus KH S22
1 Cx tritaeniorhynchus KH S21
1 Cx tritaeniorhynchus MG S115
1 Cx bitaeniorhynchus KH S18
1 Cx bitaeniorhynchus KH S20
1 Cx bitaeniorhynchus KH S19, 2 S20
1 Cx pseudovishnui KH S7
2 Cx poicilipes MG S82
1 Cx poicilipes MG S82
1 Cx sp.2 CF S50
1 Cx sp.2 CF S49
1 Cx sp.2 CF S48
1 Cx duttoni CF S52
1 Cx sp.1 KH S13
1 Cx spissipes GF S98
1 Cx spissipes GF S99
1 Cx spissipes GF S97
1 Cx pedroi GF S106
1 Cx pedroi GF S108
1 Cx pedroi GF S107
1 Cx portesi GF S94
1 Cx portesi GF S95
1 Cx portesi GF S96
1 Li durhamii GF S100, S102
1 Mi mediolineata MG S83
1 Mi sp.1 MG S64
2 Cq venezuelensis GF S93
1 Cq venezuelensis GF S93
1 Cq venezuelensis GF S91
1 Cq venezuelensis GF S92
1 Ma uniformis MG S77–S79
1 Ma uniformis KH S11, S15–S17
1 Ma uniformis CF S51
1 Ma indiana KH S10, S12
1 Ma sp.1 CF S33
1 Ma sp.3 CF S35
1 Ma sp.2 CF S34
2 Ma sp.2 CF S34
1 Ma sp.4 GF S103
1 Ma sp.4 GF S104
1 Ma titillans GF S105
1 Ps ferox GF S110
1 Ps ferox GF S111
1 Ps ferox GF S109
1 Ur geometrica GF S112
1, 2 Ur geometrica GF S114
1 Ur geometrica GF S113
1 Ae scapularis GF S85
2 Ae scapularis GF S85
1 Ae scapularis GF S86
1 Ae scapularis GF S87
1 Ae serratus GF S88
1 Ae serratus GF S89
1 Ae serratus GF S90
1 Ae vittatus CF S32
1 Ae vittatus CF S30
1 Ae vittatus CF S31
1 Ae aegypti CF S24
1 Ae aegypti CF S26
1 Ae aegypti CF S25
1 Ae simpsoni CF S27
1 Ae simpsoni CF S28
1 Ae simpsoni CF S29
1 Ae albopictus KH S2
1 Ae albopictus CF S38
1 Ae albopictus CF S37
1 Ae albopictus KH S1
1 Ae albopictus CF S36
1 Ae albopictus KH S3
1 Er intermedius CF S55
1 Er intermedius CF S56
1 Er intermedius CF S54
1 An coustani MG S66
1 An coustani MG S67
1 An coustani CF S39, S41
1 An coustani CF S40
1 An baezai KH S6
1 An baezai KH S4
1 An baezai KH S5
1 An funestus CF S42
1 An funestus CF S44
1 An funestus MG S69, S70
1 An funestus CF S43
1 An squamosus MG S68
1 An gambiae CF S45
1 An gambiae MG S71
1 An gambiae CF S47
1 An gambiae CF S46
2 An gambiae CF S45
1 An gambiae MG S72, S73
1 Ho sp.1 CF S34
2 Ho sp.2 CF S34
subgenus Culex
subgenus Oculeomyia
subgenus Oculeomyia
subgenus Melanoconion
subgenus Mansonoides
subgenus Ochlerotatus
subgenus Stegomyia
subgenus Anopheles
subgenus Cellia
0.05

626

627 **Supplementary Figure S3.** Phylogenetic tree based on 28S sequences generated from this study
628 (3900 bp) as inferred using maximum-likelihood method and constructed to scale in MEGA X (S.
629 Kumar et al., 2018) in (A) and (B) radial format. Values at each node indicate bootstrap support (%)
630 from 500 replication. Each specimen label contains information on taxonomy, origin (as indicated in 2-
631 letter country codes), and specimen ID. Label colours indicate genera: *Culex* in coral, *Anopheles* in
632 purple, *Aedes* in dark blue, *Mansonia* in dark green, *Limatus* in light green, *Coquillettidia* in light blue,
633 *Psorophora* in yellow, *Mimomyia* in teal, *Uranotaenia* in pink and *Eretmapodites* in brown. Scale bar
634 at 0.05 is shown.



Supplementary Figure S4. Phylogenetic tree based on 28S sequences generated from this study and from GenBank (3900 bp) as inferred using maximum-likelihood method and constructed to scale in MEGA X (S. Kumar et al., 2018) in radial format to illustrate the distance of Anopheline sequences relative to other Culicidae members. Values at each node indicate bootstrap support (%) from 500 replications. For sequences from this study, each specimen label contains information on taxonomy, origin (in 2-letter country codes), and specimen ID. Labels in bold indicate GenBank sequences with

accession numbers shown. Label colours indicate genera: *Culex* in coral, *Anopheles* in purple, *Aedes* in dark blue, *Mansonia* in dark green, *Culiseta* in maroon, *Limatus* in light green, *Coquillettidia* in light blue, *Psorophora* in yellow, *Mimomyia* in teal, *Uranotaenia* in pink and *Eretmapodites* in brown. Scale bar at 0.05 is shown.

Supplementary File 1: Multiple sequence alignment of 169 28S rRNA sequences from this study and from GenBank (FASTA).

Supplementary File 2: Multiple sequence alignment of 122 28S rRNA sequences, including two sequences from *Horreolanus sp.* (FASTA).

Supplementary File 3: Multiple sequence alignment of 114 18S rRNA sequences, including two sequences from *Horreolanus sp*. (FASTA).

Supplementary File 4: Multiple sequence alignment of 106 COI sequences (FASTA).

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arctander, P. (1995). Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proceedings of the Royal Society B: Biological Sciences*, *262*(1363), 13–19. https://doi.org/10.1098/rspb.1995.0170

Arunachalam, N., Philip Samuel, P., Hiriyan, J., Thenmozhi, V., & Gajanana, A. (2004). Japanese encephalitis in Kerala, South India: Can Mansonia (Diptera: Culicidae) play a supplemental role in transmission? *Journal of Medical Entomology*, *41*(3), 456–461. https://doi.org/10.1603/0022-2585-41.3.456

Aspen, S., & Savage, H. M. (2003). Polymerase chain reaction assay identifies North American members of the Culex pipiens complex based on nucleotide sequence differences in the acetylcholinesterase gene Ace.2. *Journal of the American Mosquito Control Association*, *19*(4), 323–328.

Auerswald, H., Maquart, P. O., Chevalier, V., & Boyer, S. (2021). Mosquito vector competence for japanese encephalitis virus. *Viruses*, *13*(6), 1154. https://doi.org/10.3390/v13061154

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M.,

671 Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G.,
672 Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its
673 applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.
674 https://doi.org/10.1089/cmb.2012.0021
675 Barrio-Nuevo, K. M., Cunha, M. S., Luchs, A., Fernandes, A., Rocco, I. M., Mucci, L. F., DE Souza, R.
676 P., Medeiros-Sousa, A. R., Ceretti-Junior, W., & Marrelli, M. T. (2020). Detection of Zika and
677 dengue viruses in wildcaught mosquitoes collected during field surveillance in an environmental
678 protection area in São Paulo, Brazil. *PLoS ONE*, *15*(10), e0227239.
679 https://doi.org/10.1371/journal.pone.0227239
680 Batovska, J., Cogan, N. O. I., Lynch, S. E., & Blacket, M. J. (2017). Using next-generation sequencing
681 for DNA barcoding: Capturing allelic variation in ITS2. *G3: Genes, Genomes, Genetics*, *7*(1),
682 19–29. https://doi.org/10.1534/G3.116.036145/-/DC1
683 Beebe, N. W. (2018). DNA barcoding mosquitoes: Advice for potential prospectors. *Parasitology*,
684 *145*(5), 622–633. https://doi.org/10.1017/S0031182018000343
685 Behura, S. K. (2006). Molecular marker systems in insects: current trends and future avenues.
686 *Molecular Ecology*, *15*(11), 3087–3113. https://doi.org/10.1111/J.1365-294X.2006.03014.X
687 Belda, E., Nanfack-Minkeu, F., Eiglmeier, K., Carissimo, G., Holm, I., Diallo, M., Diallo, D., Vantaux,
688 A., Kim, S., Sharakhov, I. V., & Vernick, K. D. (2019). De novo profiling of RNA viruses in
689 Anopheles malaria vector mosquitoes from forest ecological zones in Senegal and Cambodia.
690 *BMC Genomics*, *20*(1), 664. https://doi.org/10.1186/s12864-019-6034-1
691 Bhattacharya, S., Basu, P., & Sajal Bhattacharya, C. (2016). The Southern House Mosquito, Culex
692 quinquefasciatus: profile of a smart vector. *Journal of Entomology and Zoology Studies JEZS*,
693 *4*(2), 73–81.
694 Bishop-Lilly, K. A., Turell, M. J., Willner, K. M., Butani, A., Nolan, N. M. E., Lentz, S. M., Akmal, A.,
695 Mateczun, A., Brahmbhatt, T. N., Sozhamannan, S., Whitehouse, C. A., & Read, T. D. (2010).
696 Arbovirus detection in insect vectors by Rapid, high- throughput pyrosequencing. *PLoS*
697 *Neglected Tropical Diseases*, *4*(11), e878. https://doi.org/10.1371/journal.pntd.0000878
698 Brault, A. C., Foy, B. D., Myles, K. M., Kelly, C. L. H., Higgs, S., Weaver, S. C., Olson, K. E., Miller, B.
699 R., & Powers, A. M. (2004). Infection patterns of o'nyong nyong virus in the malaria-transmitting
700 mosquito, Anopheles gambiae. *Insect Molecular Biology*, *13*(6), 625–635.

701 https://doi.org/10.1111/j.0962-1075.2004.00521.x
702 Cardoso, J. da C., de Almeida, M. A. B., dos Santos, E., da Fonseca, D. F., Sallum, M. A. M., Noll, C.
703 A., Monteiro, H. A. d. O., Cruz, A. C. R., Carvalho, V. L., Pinto, E. V., Castro, F. C., Neto, J. P.
704 N., Segura, M. N. O., & Vasconcelos, P. F. C. (2010). Yellow fever virus in Haemagogus
705 leucocelaenus and Aedes serratus mosquitoes, Southern Brazil, 2008. *Emerging Infectious*
706 *Diseases*, *16*(12), 1918–1924. https://doi.org/10.3201/eid1612.100608
707 Chandler, J. A., Liu, R. M., & Bennett, S. N. (2015). RNA Shotgun Metagenomic Sequencing of
708 Northern California (USA) Mosquitoes Uncovers Viruses, Bacteria, and Fungi. *Frontiers in*
709 *Microbiology*, *6*, 185. https://doi.org/10.3389/fmicb.2015.00185
710 Cornel, A. J., Mcabee, R. D., Rasgon, J., Stanich, M. A., Scott, T. W., & Coetzee, M. (2003).
711 Differences in Extent of Genetic Introgression between Sympatric Culex pipiens and Culex
712 quinquefasciatus (Diptera: Culicidae) in California and South Africa. *Journal of Medical*
713 *Entomology*, *40*(1), 36–51. https://doi.org/10.1603/0022-2585-40.1.36
714 Danforth, B. N., Lin, C. P., & Fang, J. (2005). How do insect nuclear ribosomal genes compare to
715 protein-coding genes in phylogenetic utility and nucleotide substitution patterns? *Systematic*
716 *Entomology*, *30*(4), 549–562. https://doi.org/10.1111/J.1365-3113.2005.00305.X
717 De Oliveira, C. D., Gonçalves, D. S., Baton, L. A., Shimabukuro, P. H. F., Carvalho, F. D., & Moreira,
718 L. A. (2015). Broader prevalence of Wolbachia in insects including potential human disease
719 vectors. *Bulletin of Entomological Research*, *105*(3), 305–315.
720 https://doi.org/10.1017/S0007485315000085
721 Diallo, D., Fall, G., Diagne, C. T., Gaye, A., Ba, Y., Dia, I., Faye, O., & Diallo, M. (2020). Concurrent
722 amplification of Zika, chikungunya, and yellow fever virus in a sylvatic focus of arboviruses in
723 Southeastern Senegal, 2015. *BMC Microbiology*, *20*, 181. https://doi.org/10.1186/s12866-020-
724 01866-9
725 Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space
726 complexity. *BMC Bioinformatics*, *5*, 113. https://doi.org/10.1186/1471-2105-5-113
727 Farajollahi, A., Fonseca, D. M., Kramer, L. D., & Marm Kilpatrick, A. (2011). “Bird biting” mosquitoes
728 and human disease: A review of the role of Culex pipiens complex mosquitoes in epidemiology.
729 *Infection, Genetics and Evolution*, *11*(7), 1577–1585. doi: 10.1016/j.meegid.2011.08.013
730 Fauver, J. R., Akter, S., Morales, A. I. O., Black, W. C., Rodriguez, A. D., Stenglein, M. D., Ebel, G.

731 D., & Weger-Lucarelli, J. (2019). A reverse-transcription/RNase H based protocol for depletion of
732 mosquito ribosomal RNA facilitates viral intrahost evolution analysis, transcriptomics and
733 pathogen discovery. *Virology*, *528*, 181–197. https://doi.org/10.1016/j.virol.2018.12.020
734 Foley, D. H., Rueda, L. M., & Wilkerson, R. C. (2007). Insight into Global Mosquito Biogeography from
735 Country Species Records. *Journal of Medical Entomology*, *44*(4), 554–567.
736 https://doi.org/10.1093/JMEDENT/44.4.554
737 Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of
738 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular*
739 *Marine Biology and Biotechnology*, *3*(5), 294–299.
740 Gaillet, M., Pichard, C., Restrepo, J., Lavergne, A., Perez, L., Enfissi, A., Abboud, P., Lambert, Y.,
741 Ma, L., Monot, M., Demar, M., Djossou, F., Servas, V., Nacher, M., Andrieu, A., Prudhomme, J.,
742 Michaud, C., Rousseau, C., Jeanne, I., … Rousset, D. (2021). Outbreak of oropouche virus in
743 french guiana. *Emerging Infectious Diseases*, *27*(10), 2711–2714.
744 https://doi.org/10.3201/eid2710.204760
745 Gale, K., & Crampton, J. (1989). The ribosomal genes of the mosquito, Aedes aegypti. *European*
746 *Journal of Biochemistry*, *185*(2), 311–317. https://doi.org/10.1111/j.1432-1033.1989.tb15117.x
747 Hajibabaei, M., Singer, G. A. C., & Hickey, D. A. (2006). Benchmarking DNA barcodes: An
748 assessment using available primate sequences. *Genome*, *49*(7), 851–854.
749 https://doi.org/10.1139/G06-025/SUPPL_FILE/G06-025B.PDF
750 Harbach, R. E. (2007). The Culicidae (Diptera): A review of taxonomy, classification and phylogeny.
751 *Zootaxa*, *1668*(1), 591–638. https://doi.org/10.11646/zootaxa.1668.1.28
752 Harbach, R. E., Culverwell, C. L., & Kitching, I. J. (2017). Phylogeny of the nominotypical subgenus of
753 Culex (Diptera: Culicidae): insights from analyses of anatomical data into interspecific
754 relationships and species groups in an unresolved tree. *Systematics and Biodiversity*, *15*(4),
755 296–306. https://doi.org/10.1080/14772000.2016.1252439
756 Harbach, R. E., & Kitching, I. J. (2016). The phylogeny of Anophelinae revisited: Inferences about the
757 origin and classification of Anopheles (Diptera: Culicidae). *Zoologica Scripta*, *45*(1), 34–47.
758 https://doi.org/10.1111/zsc.12137
759 Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through
760 DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321.

761 https://doi.org/10.1098/rspb.2002.2218
762 Hoyos-López, R., Soto, S. U., Rúa-Uribe, G., & Gallego-Gómez, J. C. (2015). Molecular identification
763 of saint louis encephalitis virus genotype IV in Colombia. *Memorias Do Instituto Oswaldo Cruz*,
764 *110*(6), 719–725. https://doi.org/10.1590/0074-02760280040
765 Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population,
766 phylogeographic and phylogenetic studies: The effects of inherited symbionts. *Proceedings of*
767 *the Royal Society B: Biological Sciences*, *272*, 1525–1534.
768 https://doi.org/10.1098/rspb.2005.3056
769 Kim, H., Cha, G. W., Jeong, Y. E., Lee, W. G., Chang, K. S., Roh, J. Y., Yang, S. C., Park, M. Y.,
770 Park, C., & Shin, E. H. (2015). Detection of Japanese encephalitis virus genotype V in Culex
771 orientalis and Culex pipiens (Diptera: Culicidae) in Korea. *PLoS ONE*, *10*(2), e0116547.
772 https://doi.org/10.1371/journal.pone.0116547
773 Kraemer, M. U. G., Reiner, R. C., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D. M., Yi, D.,
774 Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Bisanzio, D., Perkins, T. A.,
775 Lai, S., Lu, X., Jones, P., Coelho, G. E., Carvalho, R. G., … Golding, N. (2019). Past and future
776 spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. *Nature Microbiology*, *4*(5),
777 854–863. https://doi.org/10.1038/s41564-019-0376-y
778 Kukutla, P., Steritz, M., & Xu, J. (2013). Depletion of ribosomal RNA for mosquito gut metagenomic
779 RNA-seq. *Journal of Visualized Experiments*, *74*, 50093. https://doi.org/10.3791/50093
780 Kumar, N., Creasy, T., Sun, Y., Flowers, M., Tallon, L. J., & Dunning Hotopp, J. C. (2012). Efficient
781 subtraction of insect rRNA prior to transcriptome analysis of Wolbachia-Drosophila lateral gene
782 transfer. *BMC Research Notes*, *5*, 230. https://doi.org/10.1186/1756-0500-5-230
783 Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary
784 genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–
785 1549. https://doi.org/10.1093/molbev/msy096
786 Logue, K., Chan, E. R., Phipps, T., Small, S. T., Reimer, L., Henry-Halldin, C., Sattabongkot, J., Siba,
787 P. M., Zimmerman, P. A., & Serre, D. (2013). Mitochondrial genome sequences reveal deep
788 divergences among Anopheles punctulatus sibling species in Papua New Guinea. *Malaria*
789 *Journal*, *12*(1), 1–11. https://doi.org/10.1186/1475-2875-12-64/FIGURES/3
790 Lorenz, C., Alves, J. M. P., Foster, P. G., Suesdek, L., & Sallum, M. A. M. (2021). Phylogeny and

temporal diversification of mosquitoes (Diptera: Culicidae) with an emphasis on the Neotropical fauna. *Systematic Entomology*, *46*(4), 798–811. https://doi.org/10.1111/SYEN.12489

Lutomiah, J., Bast, J., Clark, J., Richardson, J., Yalwala, S., Oullo, D., Mutisya, J., Mulwa, F., Musila, L., Khamadi, S., Schnabel, D., Wurapa, E., & Sang, R. (2013). Abundance, diversity, and distribution of mosquito vectors in selected ecological regions of Kenya: Public health implications. *Journal of Vector Ecology*, *38*(1), 134–142. https://doi.org/10.1111/j.1948-7134.2013.12019.x

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268

Maquart, P. O., & Boyer, S. (2022). Culex vishnui. *Trends in Parasitology*, *Vector of the Month*, 491–492. https://doi.org/10.1016/J.PT.2022.01.003

Maquart, P. O., Sokha, C., & Boyer, S. (2021). Mosquito diversity (Diptera: Culicidae) and medical importance, in a bird sanctuary inside the flooded forest of Prek Toal, Cambodia. *Journal of Asia-Pacific Entomology*, *24*(4), 1221–1227. https://doi.org/10.1016/j.aspen.2021.08.001

Mitchell, C. J., Forattini, O. P., & Miller, B. R. (1986). Vector competence experiments with Rocio virus and three mosquito species from the epidemic zone in Brazil. *Revista de Saúde Pública*, *20*(3), 171–177. https://doi.org/10.1590/s0034-89101986000300001

Morlan, J. D., Qu, K., & Sinicropi, D. V. (2012). Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS ONE*, *7*(8), e42882. https://doi.org/10.1371/journal.pone.0042882

Mukwaya, L. G., Kayondo, J. K., Crabtree, M. B., Savage, H. M., Biggerstaff, B. J., & Miller, B. R. (2000). Genetic differentiation in the yellow fever virus vector, Aedes simpsoni complex, in Africa: Sequence variation in the ribosomal DNA internal transcribed spacers of anthropophilic and non-anthropophilic populations. *Insect Molecular Biology*, *9*(1), 85–91. doi: 10.1046/j.1365-2583.2000.00161.x

Mwangangi, J. M., Muturi, E. J., Muriu, S. M., Nzovu, J., Midega, J. T., & Mbogo, C. (2013). The role of Anopheles arabiensis and Anopheles coustani in indoor and outdoor malaria transmission in Taveta District, Kenya. *Parasites and Vectors*, *6*, 114. https://doi.org/10.1186/1756-3305-6-114

821 Navarro, J. C., & Weaver, S. C. (2004). Molecular phylogeny of the Vomerifer and Pedroi Groups in
822 the spissipes section of the subgenus Culex (Melanoconion). *Journal of Medical Entomology*,
823 *41*(4), 575–581. https://doi.org/10.1603/0022-2585-41.4.575
824 Nchoutpouen, E., Talipouo, A., Djiappi-Tchamen, B., Djamouko-Djonkam, L., Kopya, E., Ngadjeu, C.
825 S., Doumbe-Belisse, P., Awono-Ambene, P., Kekeunou, S., Wondji, C. S., & Antonio-Nkondjio,
826 C. (2019). Culex species diversity, susceptibility to insecticides and role as potential vector of
827 Lymphatic filariasis in the city of Yaoundé, Cameroon. *PLoS Neglected Tropical Diseases*,
828 *13*(4), e0007229. https://doi.org/10.1371/journal.pntd.0007229
829 Ndiaye, E. H., Fall, G., Gaye, A., Bob, N. S., Talla, C., Diagne, C. T., Diallo, D., Ba, Y., Dia, I., Kohl,
830 A., Sall, A. A., & Diallo, M. (2016). Vector competence of Aedes vexans (Meigen), Culex
831 poicilipes (Theobald) and Cx. quinquefasciatus Say from Senegal for West and East African
832 lineages of Rift Valley fever virus. *Parasites and Vectors*, *9*, 94. https://doi.org/10.1186/s13071-
833 016-1383-y
834 Nepomichene, T. N. J. J., Raharimalala, F. N., Andriamandimby, S. F., Ravalohery, J. P., Failloux, A.
835 B., Heraud, J. M., & Boyer, S. (2018). Vector competence of Culex antennatus and Anopheles
836 coustani mosquitoes for Rift Valley fever virus in Madagascar. *Medical and Veterinary*
837 *Entomology*, *32*(2), 259–262. https://doi.org/10.1111/mve.12291
838 Nikolay, B., Diallo, M., Boye, C. S. B., & Sall, A. A. (2011). Usutu virus in Africa. *Vector-Borne and*
839 *Zoonotic Diseases*, *11*(11), 1417–1423. https://doi.org/10.1089/vbz.2011.0631
840 Phelps, W. A., Carlson, A. E., & Lee, M. T. (2021). Optimized design of antisense oligomers for
841 targeted rRNA depletion. *Nucleic Acids Research*, *49*(1), e5.
842 https://doi.org/10.1093/nar/gkaa1072
843 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O.
844 (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-
845 based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596.
846 https://doi.org/10.1093/nar/gks1219
847 Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System: Barcoding.
848 *Molecular Ecology Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x
849 Ratovonjato, J., Olive, M. M., Tantely, L. M., Andrianaivolambo, L., Tata, E., Razainirina, J.,
850 Jeanmaire, E., Reynes, J. M., & Elissa, N. (2011). Detection, isolation, and genetic

851 characterization of Rift Valley fever virus from anopheles (Anopheles) coustani, anopheles
852 (Anopheles) squamosus, and culex (Culex) antennatus of the haute matsiatra region,
853 Madagascar. *Vector-Borne and Zoonotic Diseases*, *11*(6), 753–759.
854 https://doi.org/10.1089/vbz.2010.0031
855 Rausch, T., Fritz, M. H. Y., Untergasser, A., & Benes, V. (2020). Tracy: Basecalling, alignment,
856 assembly and deconvolution of sanger chromatogram trace files. *BMC Genomics*, *21*(1), 230.
857 https://doi.org/10.1186/s12864-020-6635-8
858 Rausch, T., Hsi-Yang Fritz, M., Korbel, J. O., & Benes, V. (2019). Alfred: Interactive multi-sample
859 BAM alignment statistics, feature counting and feature annotation for long- and short-read
860 sequencing. *Bioinformatics*, *35*(14), 2489–2491. https://doi.org/10.1093/bioinformatics/bty1007
861 Reidenbach, K. R., Cook, S., Bertone, M. A., Harbach, R. E., Wiegmann, B. M., & Besansky, N. J.
862 (2009). Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae)
863 based on nuclear genes and morphology. *BMC Evolutionary Biology*, *9*(1), 1–14.
864 https://doi.org/10.1186/1471-2148-9-298/FIGURES/4
865 Ruzzante, L., Reijnders, M. J. M. F., & Waterhouse, R. M. (2019). Of Genes and Genomes: Mosquito
866 Evolution and Diversity. *Trends in Parasitology*, *35*(1), 32–51.
867 https://doi.org/10.1016/J.PT.2018.10.003/ATTACHMENT/B9BE6BC5-D73A-4FEF-A654-
868 CC1374C59925/MMC1.MP4
869 Sallum, M. A. M., & Forattini, O. P. (1996). Revision of the Spissipes Section of Culex (Melanoconion)
870 (diptera: Culicidae). *Journal of the American Mosquito Control Association*, *12*(3), 517–600.
871 Sirivanakarn, S. (1982). A review of the Systematics and a Proposed Scheme of Internal
872 Classification of the New World Subgenus Melanoconion of Culex (Diptera, Culicidae). *Mosquito*
873 *Systematics*, *14*(4), 265–333.
874 Stevenson, J. C., Simubali, L., Mbambara, S., Musonda, M., Mweetwa, S., Mudenda, T., Pringle, J.
875 C., Jones, C. M., & Norris, D. E. (2016). Detection of plasmodium falciparum infection in
876 anopheles squamosus (diptera: Culicidae) in an area targeted for malaria elimination, Southern
877 Zambia. *Journal of Medical Entomology*, *53*(6), 1482–1487. https://doi.org/10.1093/jme/tjw091
878 Sun, L., Li, T. J., Fu, W. B., Yan, Z. T., Si, F. L., Zhang, Y. J., Mao, Q. M., Demari-Silva, B., & Chen,
879 B. (2019). The complete mt genomes of Lutzia halifaxia, Lt. fuscanus and Culex pallidothorax
880 (Diptera: Culicidae) and comparative analysis of 16 Culex and Lutzia mt genome sequences.

881 *Parasites and Vectors*, *12*, 368. https://doi.org/10.1186/s13071-019-3625-2
882 Takhampunya, R., Kim, H. C., Tippayachai, B., Kengluecha, A., Klein, T. A., Lee, W. J., Grieco, J., &
883 Evans, B. P. (2011). Emergence of Japanese encephalitis virus genotype v in the Republic of
884 Korea. *Virology Journal*, *8*, 449. https://doi.org/10.1186/1743-422X-8-449
885 Talaga, S., Duchemin, J. B., Girod, R., & Dusfour, I. (2021). The Culex Mosquitoes (Diptera:
886 Culicidae) of French Guiana: A Comprehensive Review With the Description of Three New
887 Species. *Journal of Medical Entomology*, *58*(1), 182–221. https://doi.org/10.1093/JME/TJAA205
888 Thongsripong, P., Chandler, J. A., Kittayapong, P., Wilcox, B. A., Kapan, D. D., & Bennett, S. N.
889 (2021). Metagenomic shotgun sequencing reveals host species as an important driver of virome
890 composition in mosquitoes. *Scientific Reports*, *11*(1), 8448. https://doi.org/10.1038/s41598-021-
891 87122-0
892 Torres-Gutierrez, C., Bergo, E. S., Emerson, K. J., de Oliveira, T. M. P., Greni, S., & Sallum, M. A. M.
893 (2016). Mitochondrial COI gene as a tool in the taxonomy of mosquitoes Culex subgenus
894 Melanoconion. *Acta Tropica*, *164*, 137–149. https://doi.org/10.1016/j.actatropica.2016.09.007
895 Torres-Gutierrez, C., De Oliveira, T. M. P., Emerson, K. J., Bergo, E. S., & Sallum, M. A. M. (2018).
896 Molecular phylogeny of Culex subgenus Melanoconion (Diptera: Culicidae) based on nuclear
897 and mitochondrial protein-coding genes. *Royal Society Open Science*, *5*, 171900.
898 https://doi.org/10.1098/rsos.171900
899 Turell, M. J., O'guinn, M. L., Dohm, D., Zyzak, M., Watts, D., Fernandez, R., Calampa, C., Klein, T. A.,
900 & Jones, J. W. (2008). Susceptibility of Peruvian Mosquitoes to Eastern Equine Encephalitis
901 Virus. *Journal of Medical Entomology*, *45*(4), 720–725.
902 https://doi.org/10.1093/JMEDENT/45.4.720
903 Turell, Michael J. (1999). Vector competence of three Venezuelan mosquitoes (Diptera: Culicidae) for
904 an epizootic IC strain of Venezuelan equine encephalitis virus. *Journal of Medical Entomology*,
905 *36*(4), 407–409. https://doi.org/10.1093/jmedent/36.4.407
906 Ughasi, J., Bekard, H. E., Coulibaly, M., Adabie-Gomez, D., Gyapong, J., Appawu, M., Wilson, M. D.,
907 & Boakye, D. A. (2012). Mansonia africana and Mansonia uniformis are Vectors in the
908 transmission of Wuchereria bancrofti lymphatic filariasis in Ghana. *Parasites and Vectors*, *5*(1),
909 1–5. https://doi.org/10.1186/1756-3305-5-89
910 Valentine, M. J., Murdock, C. C., & Kelly, P. J. (2019). Sylvatic cycles of arboviruses in non-human

primates. *Parasites and Vectors*, *12*(1), 1–18. https://doi.org/10.1186/S13071-019-3732-0/TABLES/4

Vasconcelos, P. F. C., Costa, Z. G., Travassos da Rosa, E. S., Luna, E., Rodrigues, S. G., Barros, V. L. R. S., Dias, J. P., Monteiro, H. A. O., Oliva, O. F. P., Vasconcelos, H. B., Oliveira, R. C., Sousa, M. R. S., Barbosa Da Silva, J., Cruz, A. C. R., Martins, E. C., & Travassos Da Rosa, J. F. S. (2001). Epidemic of jungle yellow fever in Brazil, 2000: Implications of climatic alterations in disease spread. *Journal of Medical Virology*, *65*(3), 598–604. https://doi.org/10.1002/jmv.2078.abs

Vázquez González, A., Ruiz, S., Herrero, L., Moreno, J., Molero, F., Magallanes, A., Sánchez-Seco, M. P., Figuerola, J., & Tenorio, A. (2011). West Nile and Usutu viruses in mosquitoes in Spain, 2008-2009. *American Journal of Tropical Medicine and Hygiene*, *85*(1), 178–181. https://doi.org/10.4269/ajtmh.2011.11-0042

Vezenegho, S. B., Issaly, J., Carinci, R., Gaborit, P., Girod, R., Dusfour, I., & Briolant, S. (2022). Discrimination of 15 Amazonian Anopheline Mosquito Species by Polymerase Chain Reaction—Restriction Fragment Length Polymorphism. *Journal of Medical Entomology*, *59*(3), 1060–1064. https://doi.org/10.1093/JME/TJAC008

Weaver, S. C., Ferro, C., Barrera, R., Boshell, J., & Navarro, J. C. (2004). Venezuelan Equine Encephalitis. *Annual Review of Entomology*, *49*, 141–174. https://doi.org/10.1146/annurev.ento.49.061802.123422

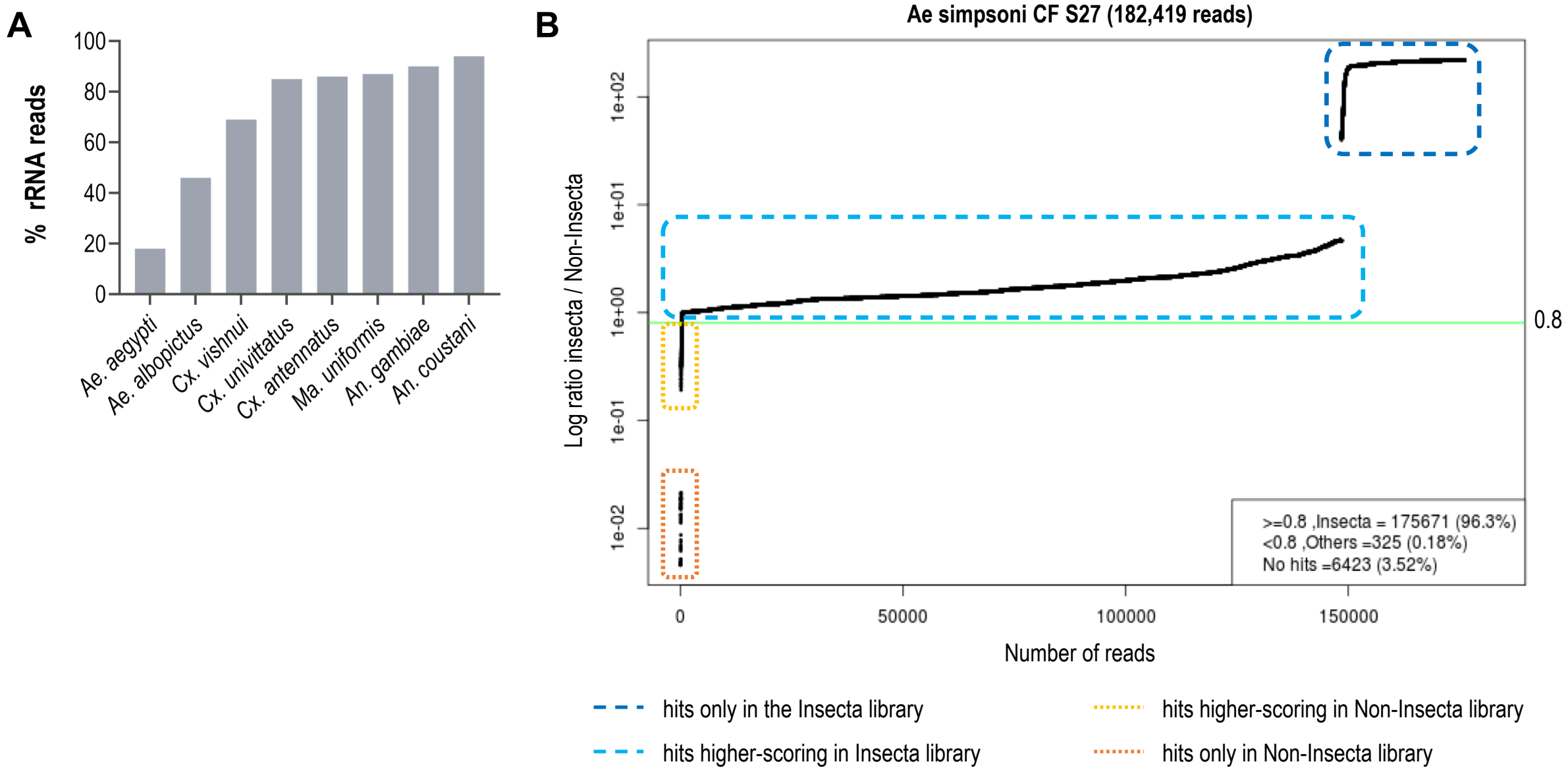
Webster, J. P., Gower, C. M., Knowles, S. C. L., Molyneux, D. H., & Fenton, A. (2016). One health - an ecological and evolutionary framework for tackling Neglected Zoonotic Diseases. *Evolutionary Applications*, *9*(2), 313–333. https://doi.org/10.1111/eva.12341

Weedall, G. D., Irving, H., Hughes, M. A., & Wondji, C. S. (2015). Molecular tools for studying the major malaria vector Anopheles funestus: Improving the utility of the genome using a comparative poly(A) and Ribo-Zero RNAseq analysis. *BMC Genomics*, *16*(1), 931. https://doi.org/10.1186/s12864-015-2114-z

WHO. (2017). Global vector control response 2017–2030. In *World Health Organization*.

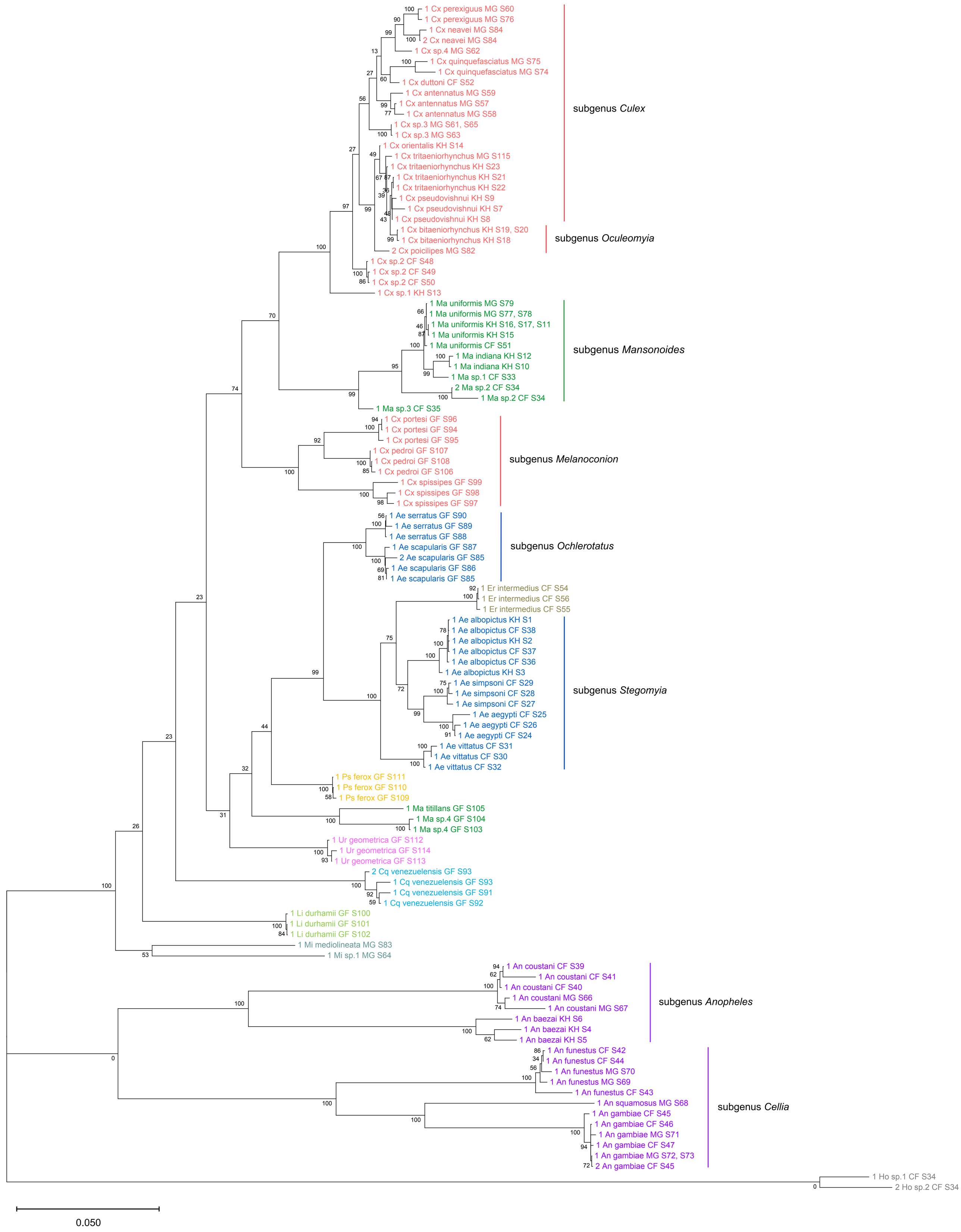
Zakrzewski, M., Rašić, G., Darbro, J., Krause, L., Poo, Y. S., Filipović, I., Parry, R., Asgari, S., Devine, G., & Suhrbier, A. (2018). Mapping the virome in wild-caught Aedes aegypti from Cairns and Bangkok. *Scientific Reports*, *8*(1), 4690. https://doi.org/10.1038/s41598-018-22945-y

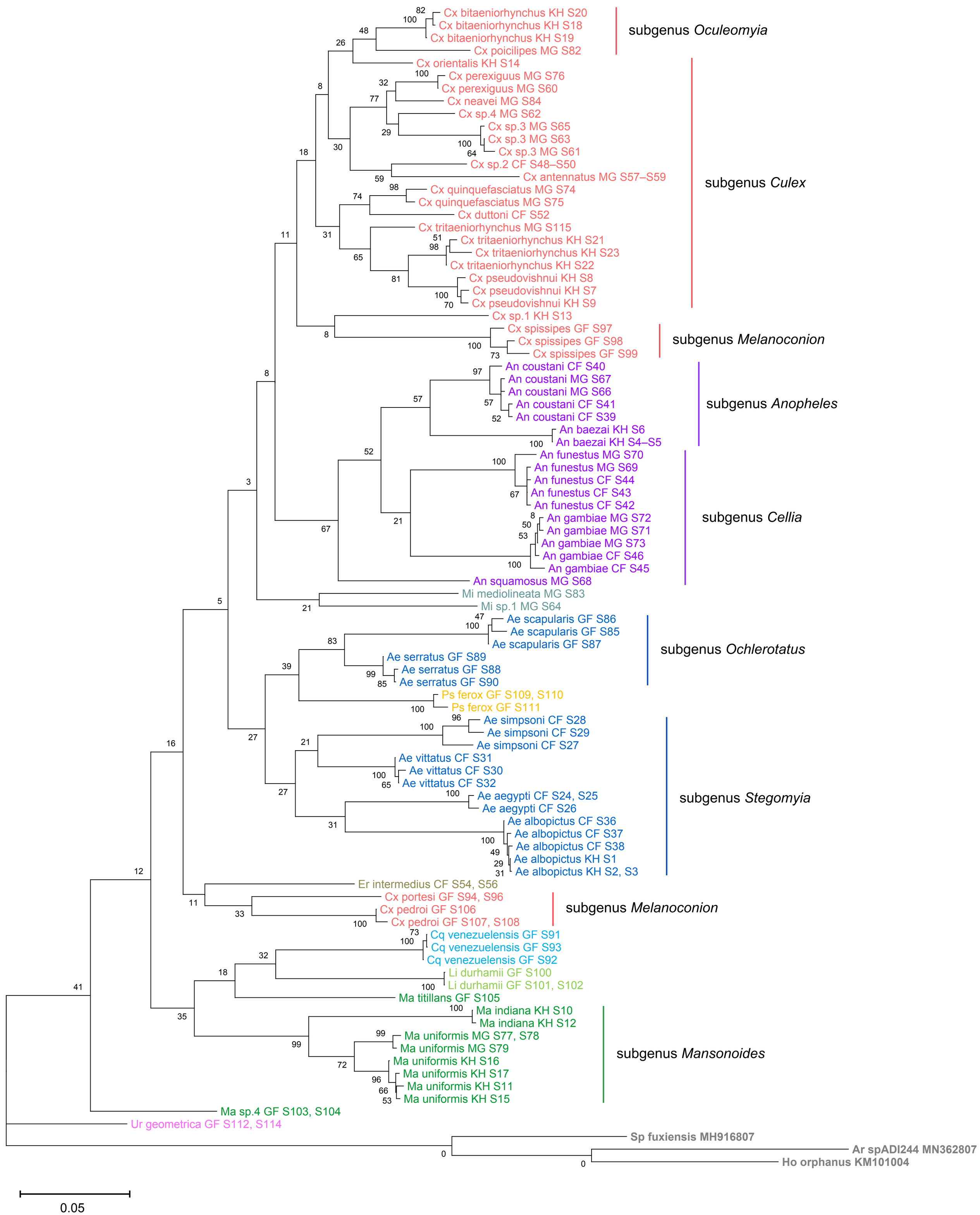
Zittra, C., Flechl, E., Kothmayer, M., Vitecek, S., Rossiter, H., Zechmeister, T., & Fuehrer, H. P. (2016). Ecological characterization and molecular differentiation of Culex pipiens complex taxa and Culex torrentium in eastern Austria. *Parasites and Vectors*, *9*, 197. https://doi.org/10.1186/s13071-016-1495-4


A
% rRNA reads
100
80
60
40
20
0
Ae. aegypti
Ae. albopictus
Cx. vishnui
Cx. univittatus
Cx. antennatus
Ma. uniformis
An. gambiae
An. coustani
B
Ae simpsoni CF S27 (182,419 reads)
Log ratio insecta / Non-Insecta
1e+02
1e+01
1e+00
1e-01
1e-02
0.8
>=0.8 ,Insecta = 175671 (96.3%)
<0.8 ,Others =325 (0.18%)
No hits =6423 (3.52%)
0
50000
100000
150000
Number of reads
hits only in the Insecta library
hits higher-scoring in Insecta library
hits higher-scoring in Non-Insecta library
hits only in Non-Insecta library

1 Cx perexiguus MG S60
1 Cx perexiguus MG S76
Cx perexiguus MT808440.1
1 Cx neavei MG S84
2 Cx neavei MG S84
1 Cx sp.4 MG S62
1 Cx antennatus MG S58
1 Cx antennatus MG S57
1 Cx antennatus MG S59
1 Cx sp.3 MG S61, S65
1 Cx sp.3 MG S63
1 Cx duttoni CF S52
1 Cx quinquefasciatus MG S75
1 Cx quinquefasciatus MG S74
Cx quinquefasciatus XR 005279053.1
Cx quinquefasciatus XR 005279034.1
Cx quinquefasciatus XR 005279033.1
Cx pipiens MT808425.1
Cx pipiens XR 005606618.1
1 Cx orientalis KH S14
1 Cx tritaeniorhynchus MG S115
1 Cx tritaeniorhynchus KH S23
1 Cx pseudovishnui KH S7
1 Cx pseudovishnui KH S8
1 Cx pseudovishnui KH S9
1 Cx tritaeniorhynchus KH S21
1 Cx tritaeniorhynchus KH S22
1 Cx bitaeniorhynchus KH S18
1 Cx bitaeniorhynchus KH S19, 2 S20
1 Cx bitaeniorhynchus KH S20
2 Cx poicilipes MG S82
1 Cx poicilipes MG S82
Cx theileri MT808435.1
Cx modestus MT808439.1
1 Cx sp.2 CF S48
1 Cx sp.2 CF S49
1 Cx sp.2 CF S50
1 Cx sp.1 KH S13
Cx impudicus MT808438.1
1 Cx portesi GF S95
1 Cx portesi GF S94
1 Cx portesi GF S96
1 Cx pedroi GF S107
1 Cx pedroi GF S108
1 Cx pedroi GF S106
1 Cx spissipes GF S98
1 Cx spissipes GF S99
1 Cx spissipes GF S97
Cu longiareolata MT808423.1
Cu annulata MT808429.1
1 Ma uniformis MG S77–S79
1 Ma uniformis KH S11, S15–S17
1 Ma uniformis CF S51
1 Ma indiana KH S10, S12
1 Ma sp.1 CF S33
1 Ma sp.3 CF S35
1 Ma sp.2 CF S34
2 Ma sp.2 CF S34
1 Ma sp.4 GF S103
1 Ma sp.4 GF S104
1 Ma titillans GF S105
Cq richiardii MT672341.1
Cq richiardii MT808427.1
Cq buxtoni MT808430.1
1 Cq venezuelensis GF S92
1 Cq venezuelensis GF S91
1 Cq venezuelensis GF S93
2 Cq venezuelensis GF S93
1 Ae scapularis GF S86
1 Ae scapularis GF S87
1 Ae scapularis GF S85
2 Ae scapularis GF S85
1 Ae serratus GF S88
1 Ae serratus GF S89
1 Ae serratus GF S90
Oc sticticus MW677941.1
Oc sticticus MT808442.1
Oc caspius MT808431.1
Oc detritus MT808422.1
Oc pulcritarsis MT808426.1
Ae geniculatus MT808437.1
Oc rusticus MT808432.1
1 Ae vittatus CF S31
1 Ae vittatus CF S30
1 Ae vittatus CF S32
Ae vexans MT808428.1
1 Ae aegypti CF S24
1 Ae aegypti CF S25
1 Ae aegypti CF S26
1 Ae simpsoni CF S27
1 Ae simpsoni CF S28
1 Ae simpsoni CF S29
1 Ae albopictus KH S3
1 Ae albopictus CF S36
Ae albopictus MT808421.1
Ae albopictus L22060.1
1 Ae albopictus CF S37
1 Ae albopictus CF S38
1 Ae albopictus KH S2
1 Ae albopictus KH S1
1 Er intermedius CF S54
1 Er intermedius CF S55
1 Er intermedius CF S56
1 Li durhamii GF S100–S102
1 Ps ferox GF S109
1 Ps ferox GF S111
1 Ps ferox GF S110
Ur unguiculata MT808444.1
1 Ur geometrica GF S112
1-2 Ur geometrica GF S114
1 Ur geometrica GF S113
1 Mi mediolineata MG S83
1 Mi sp.1 MG S64
1 An coustani CF S39, S41
1 An coustani CF S40
1 An coustani MG S66
1 An coustani MG S67
1 An baezai KH S4
1 An baezai KH S6
1 An baezai KH S5
An sacharovi MT808434.1
An messeae MW677939.1
An melanoon MT808424.1
An algeriensis MT808436.1
An claviger MT808441.1
An albimanus XR 004845806.1
An albimanus XR 004845855.1
An albimanus L78065.1
1 An funestus CF S44
1 An funestus MG S69, S70
1 An funestus CF S42
1 An funestus CF S43
An superpictus MT808443.1
An stephensi XR 004908154.1
An stephensi XR 004907438.1
1 An squamosus MG S68
An merus XR 006006278.1
An arabiensis XR 005740189.1
An arabiensis XR 005740497.1
An arabiensis XR 005740437.1
An merus XR 006005901.1
1 An gambiae CF S45
An merus XR 006005999.1
An coluzzii XR 005752429.1
An arabiensis XR 005740974.1
1 An gambiae CF S47
1 An gambiae MG S71
1 An gambiae CF S46
1 An gambiae MG S72, S73
2 An gambiae CF S45
An coluzzii XR 005752533.1
An coluzzii XR 005752529.1
An coluzzii XR 005752343.1
1 Ho sp.1 CF S34
2 Ho sp.2 CF S34
subgenus Culex
subgenus Oculeomyia
subgenus Melanoconion
subgenus Mansonoides
subgenus Coquillettidia
subgenus Rhynchotaenia
subgenus Ochlerotatus
subgenus Stegomyia
subgenus Anopheles
subgenus Nyssorhynchus
subgenus Cellia
0.05

Cx sp.4 MG S62
Cx poicilipes MG S82
Cx pseudovishnui KH S7–S9; Cx tritaeniorhynchus KH S21–S23
Cx tritaeniorhynchus MG S115
Cx orientalis KH S14
Cx bitaeniorhynchus KH S18–S20
Cx quinquefasciatus MG S75
Cx sp.2 CF S48, S50
Cx sp.2 CF S49
Cx neavei MG S84
Cx sp.3 MG S61, S65
Cx sp.3 MG S63
Cx perexiguus MG S76
Cx duttoni CF S52
Cx perexiguus MG S60
Cx antennatus MG S57, S59
Cx antennatus MG S58
Ma sp.3 CF S35
Cx quinquefasciatus MG S74
Cx sp.1 KH S13
Cx portesi GF S94
Cx portesi GF S95, S96
Cx pedroi GF S106, S108
Cx pedroi GF S107
Cx spissipes GF S97–S99
subgenus Melanoconion
Ps ferox GF S109, S110
Ps ferox GF S111
Ae scapularis GF S85, S86
Ae scapularis GF S87
Ae serratus GF S88, S90
subgenus Ochlerotatus
Ae aegypti CF S25, S26
Ae aegypti CF S24
Ae albopictus KH S1, S3
Ae albopictus CF S36–S38; Ae albopictus KH S2
Ae simpsoni CF S27, S28
Ae simpsoni CF S29
Ae vittatus CF S30
Ae vittatus CF S32
Ae vittatus CF S31
Er intermedius CF S54–S56
subgenus Stegomyia
Ma uniformis CF S51 MG S77, S78
Ma uniformis KH S15
Ma uniformis KH S11, S16, S17
Ma indiana KH S10, S12
Ma uniformis MG S79
Ma sp.1 CF S33
Ma sp.2 CF S34
Ma sp.4 GF S103
Ma sp.4 GF S104
Ma titillans GF S105
subgenus Mansonoides
Ur geometrica GF S112–S114
Mi sp.1 MG S64
Cq venezuelensis GF S91, S92
Cq venezuelensis GF S93
Mi mediolineata MG S83
Li durhamii GF S100
Li durhamii GF S101, S102
An coustani CF S39, S40; An coustani MG S67
An coustani CF S41; An coustani MG S66
An baezai KH S4–S6
subgenus Anopheles
An funestus CF S43
An funestus MG S69
An funestus CF S42
An funestus CF S44; An funestus MG S70
An gambiae CF S45, S46; An gambiae MG S71–S73
An gambiae CF S47
An squamosus MG S68
subgenus Cellia
Ho sp.1 CF S34
Ho sp.2 CF S35
0.05


1 Cx perexiguus MG S60
1 Cx perexiguus MG S76
1 Cx neavei MG S84
2 Cx neavei MG S84
1 Cx sp.4 MG S62
1 Cx quinquefasciatus MG S75
1 Cx quinquefasciatus MG S74
1 Cx duttoni CF S52
1 Cx antennatus MG S59
1 Cx antennatus MG S57
1 Cx antennatus MG S58
1 Cx sp.3 MG S61, S65
1 Cx sp.3 MG S63
1 Cx orientalis KH S14
1 Cx tritaeniorhynchus MG S115
1 Cx tritaeniorhynchus KH S23
1 Cx tritaeniorhynchus KH S21
1 Cx tritaeniorhynchus KH S22
1 Cx pseudovishnui KH S9
1 Cx pseudovishnui KH S7
1 Cx pseudovishnui KH S8
1 Cx bitaeniorhynchus KH S19, S20
1 Cx bitaeniorhynchus KH S18
2 Cx poicilipes MG S82
1 Cx sp.2 CF S48
1 Cx sp.2 CF S49
1 Cx sp.2 CF S50
1 Cx sp.1 KH S13
subgenus Culex
subgenus Oculeomyia
1 Ma uniformis MG S79
1 Ma uniformis MG S77, S78
1 Ma uniformis KH S16, S17, S11
1 Ma uniformis KH S15
1 Ma uniformis CF S51
1 Ma indiana KH S12
1 Ma indiana KH S10
1 Ma sp.1 CF S33
2 Ma sp.2 CF S34
1 Ma sp.2 CF S34
1 Ma sp.3 CF S35
subgenus Mansonoides
1 Cx portesi GF S96
1 Cx portesi GF S94
1 Cx portesi GF S95
1 Cx pedroi GF S107
1 Cx pedroi GF S108
1 Cx pedroi GF S106
1 Cx spissipes GF S99
1 Cx spissipes GF S98
1 Cx spissipes GF S97
subgenus Melanoconion
1 Ae serratus GF S90
1 Ae serratus GF S89
1 Ae serratus GF S88
1 Ae scapularis GF S87
2 Ae scapularis GF S85
1 Ae scapularis GF S86
1 Ae scapularis GF S85
subgenus Ochlerotatus
1 Er intermedius CF S54
1 Er intermedius CF S56
1 Er intermedius CF S55
1 Ae albopictus KH S1
1 Ae albopictus CF S38
1 Ae albopictus KH S2
1 Ae albopictus CF S37
1 Ae albopictus CF S36
1 Ae albopictus KH S3
1 Ae simpsoni CF S29
1 Ae simpsoni CF S28
1 Ae simpsoni CF S27
1 Ae aegypti CF S25
1 Ae aegypti CF S26
1 Ae aegypti CF S24
1 Ae vittatus CF S31
1 Ae vittatus CF S30
1 Ae vittatus CF S32
subgenus Stegomyia
1 Ps ferox GF S111
1 Ps ferox GF S110
1 Ps ferox GF S109
1 Ma titillans GF S105
1 Ma sp.4 GF S104
1 Ma sp.4 GF S103
1 Ur geometrica GF S112
1 Ur geometrica GF S114
1 Ur geometrica GF S113
2 Cq venezuelensis GF S93
1 Cq venezuelensis GF S93
1 Cq venezuelensis GF S91
1 Cq venezuelensis GF S92
1 Li durhamii GF S100
1 Li durhamii GF S101
1 Li durhamii GF S102
1 Mi mediolineata MG S83
1 Mi sp.1 MG S64
1 An coustani CF S39
1 An coustani CF S41
1 An coustani CF S40
1 An coustani MG S66
1 An coustani MG S67
1 An baezai KH S6
1 An baezai KH S4
1 An baezai KH S5
subgenus Anopheles
1 An funestus CF S42
1 An funestus CF S44
1 An funestus MG S70
1 An funestus MG S69
1 An funestus CF S43
1 An squamosus MG S68
1 An gambiae CF S45
1 An gambiae CF S46
1 An gambiae MG S71
1 An gambiae CF S47
1 An gambiae MG S72, S73
2 An gambiae CF S45
subgenus Cellia
1 Ho sp.1 CF S34
2 Ho sp.2 CF S34
0.050


Cx bitaeniorhynchus KH S20
Cx bitaeniorhynchus KH S18
Cx bitaeniorhynchus KH S19
Cx poicilipes MG S82
Cx orientalis KH S14
Cx perexiguus MG S76
Cx perexiguus MG S60
Cx neavei MG S84
Cx sp.4 MG S62
Cx sp.3 MG S65
Cx sp.3 MG S63
Cx sp.3 MG S61
Cx sp.2 CF S48–S50
Cx antennatus MG S57–S59
Cx quinquefasciatus MG S74
Cx quinquefasciatus MG S75
Cx duttoni CF S52
Cx tritaeniorhynchus MG S115
Cx tritaeniorhynchus KH S21
Cx tritaeniorhynchus KH S23
Cx tritaeniorhynchus KH S22
Cx pseudovishnui KH S8
Cx pseudovishnui KH S7
Cx pseudovishnui KH S9
Cx sp.1 KH S13
Cx spissipes GF S97
Cx spissipes GF S98
Cx spissipes GF S99
An coustani CF S40
An coustani MG S67
An coustani MG S66
An coustani CF S41
An coustani CF S39
An baezai KH S6
An baezai KH S4–S5
An funestus MG S70
An funestus MG S69
An funestus CF S44
An funestus CF S43
An funestus CF S42
An gambiae MG S72
An gambiae MG S71
An gambiae MG S73
An gambiae CF S46
An gambiae CF S45
An squamosus MG S68
Mi mediolineata MG S83
Mi sp.1 MG S64
Ae scapularis GF S86
Ae scapularis GF S85
Ae scapularis GF S87
Ae serratus GF S89
Ae serratus GF S88
Ae serratus GF S90
Ps ferox GF S109, S110
Ps ferox GF S111
Ae simpsoni CF S28
Ae simpsoni CF S29
Ae simpsoni CF S27
Ae vittatus CF S31
Ae vittatus CF S30
Ae vittatus CF S32
Ae aegypti CF S24, S25
Ae aegypti CF S26
Ae albopictus CF S36
Ae albopictus CF S37
Ae albopictus CF S38
Ae albopictus KH S1
Ae albopictus KH S2, S3
Er intermedius CF S54, S56
Cx portesi GF S94, S96
Cx pedroi GF S106
Cx pedroi GF S107, S108
Cq venezuelensis GF S91
Cq venezuelensis GF S93
Cq venezuelensis GF S92
Li durhamii GF S100
Li durhamii GF S101, S102
Ma titillans GF S105
Ma indiana KH S10
Ma indiana KH S12
Ma uniformis MG S77, S78
Ma uniformis MG S79
Ma uniformis KH S16
Ma uniformis KH S17
Ma uniformis KH S11
Ma uniformis KH S15
Ma sp.4 GF S103, S104
Ur geometrica GF S112, S114
Sp fuxiensis MH916807
Ar spADI244 MN362807
Ho orphanus KM101004
subgenus Oculeomyia
subgenus Culex
subgenus Melanoconion
subgenus Anopheles
subgenus Cellia
subgenus Ochlerotatus
subgenus Stegomyia
subgenus Melanoconion
subgenus Mansonoides
0.05