bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 **High-throughput Identification of Eukaryotic Parasites and Arboviruses in Mosquitoes**

2 **Running title: High-throughput Surveillance of Parasites and Viruses**

3

4 Matthew V. Cannon[1], Haikel N. Bogale[1], Devika Bhalerao[1], Kalil Keita[2], Denka Camara[2], Yaya Barry[2],

5 Moussa Keita[2], Drissa Coulibaly[3], Abdoulaye K. Kone[3], Ogobara K. Doumbo[3†], Mahamadou A. Thera[3],

6 Christopher V. Plowe[4], Mark A. Travassos[4], Seth R. Irish[5], Joshua Yeroshefsky[6], Jeannine Dorothy[7], Brian

7 Prendergast[7], Brandyce St. Laurent[8], Megan L. Fritz[6], David Serre[1*]

8

9 [1]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA.

10 [2]Programme National de Lutte contre le Paludisme, Guinea.

11 [3]Malaria Research and Training Center, University Science, Techniques and Technologies of Bamako,

12 Mali.

13 [4]Malaria Research Program, Center for Vaccine Development and Global Health, University of Maryland

14 School of Medicine, Baltimore, USA.

15 [5]U.S. President's Malaria Initiative and Entomology Branch, Division of Parasitic Diseases and Malaria,

16 Center for Global Health, Centers for Disease Control and Prevention, Atlanta, USA.

17 [6]Department of Entomology, University of Maryland College Park, College Park, MD, USA.

18 [7]Mosquito Control Program, Maryland Department of Agriculture, Annapolis, MD, USA.

19 [8]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases,

20 Rockville, MD, USA.

21 †Deceased

22

23 * correspondence to dserre@som.umaryland.edu

24 University of Maryland School of Medicine

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

25 Institute for Genome Sciences

26 Department of Microbiology & Immunology

27 HSFIII

28 670 West Baltimore Street, Room 3111

29 Baltimore, MD 21201

30 **Abbreviations:** National Center for Parasitology, Entomology, and Malaria Control (CNM), American
31 Type Culture Collection (ATCC)

32 **Acknowledgements**

33 We wish to thank our collaborators at the National Center for Parasitology, Entomology, and Malaria
34 Control (CNM) in Phnom Penh, Stop Palu and the National Malaria Control Program in Guinea and at the
35 Malaria Research and Training Center, University of Science, Techniques and Technologies, Bamako,
36 Mali for their help with the sample collection and processing, as well as Erin Haser, Rachel Koch, Jessica
37 Moskowitz, Rachel Shuck, and Elizabeth Allan. We also thank Stop Palu (RTI), the U.S. President's
38 Malaria Initiative, and all those participating in collections for support in the sample collection and
39 processing in Guinea. The findings and conclusions in this report are those of the authors and do not
40 necessarily represent the official position of the Centers for Disease Control and Prevention.

41 **Conflict of interest statement:**

42 The authors have stated explicitly that there are no conflicts of interest in connection with this article.

43 **Author contributions**

44 MVC, BSTL, MF and DS designed the study. BSTL, MF, KK, DC, YB, MK, DC, AKK, OKD, MAT, CVP, MT, SI,
45 JD and BP collected and/or provided access to the mosquito samples. MVC, HB, BSTL and DB performed
46 the laboratory experiments. MVC analyzed the data. MVC and DS prepared the manuscript. All authors
47 read and approved the final manuscript.

48

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

49

50 **Abstract**

51 Vector-borne pathogens cause many human infectious diseases and are responsible for high mortality
52 and morbidity throughout the world. They can also cause livestock epidemics with dramatic social and
53 economic consequences. Due to the high costs, vector-borne disease surveillance is often limited to
54 current threats, and the investigation of emerging pathogens typically occur after the reports of clinical
55 cases. Here, we use high-throughput sequencing to detect and identify a wide range of parasites and
56 viruses carried by mosquitoes from Cambodia, Guinea, Mali and Maryland. We apply this approach to
57 individual *Anopheles* mosquitoes as well as pools of mosquitoes captured in traps; and compare the
58 outcomes of this assay when applied to DNA or RNA. We identified known human and animal pathogens
59 and mosquito parasites belonging to a wide range of taxa, insect Flaviviruses, and novel DNA sequences
60 from previously uncharacterized organisms. Our results also revealed that analysis of the content of an
61 entire trap is an efficient approach to monitor and identify potential vector-borne pathogens in large
62 surveillance studies, and that analyses of RNA extracted from mosquitoes is preferable, when possible,
63 over DNA-based analyses. Overall, we describe a flexible and easy-to-customize assay that can provide
64 important information for vector-borne disease surveillance and research studies to efficiently
65 complement current approaches.

66 **Keywords:** vector-borne pathogens, eukaryotic parasites, arbovirus, surveillance, zoonosis, emerging
67 pathogens

68

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

69 **Introduction**

70 Different arthropods can, during a blood feeding, transmit viruses, protists and helminths to humans (1).
71 These organisms cause some of the most prevalent human infectious diseases, including malaria,
72 dengue, schistosomiasis or Chagas disease, and are responsible for more than 700,000 human deaths
73 worldwide every year (2-4). Vector-borne diseases are also responsible for some of the most alarming
74 recent epidemics in the western hemisphere, either due to the emergence of new pathogens (e.g., Zika
75 (5, 6)), the reemergence of historically important pathogens (e.g., Yellow Fever (7)) or the expansion of
76 diseases beyond their historical ranges (e.g., West Nile (8) and Chikungunya (7)). In addition to this
77 burden on human health, many vector-borne diseases affect domesticated animals (e.g., heartworms (9,
78 10)), livestock (e.g., Theileriosis (11, 12)) and wild animals (e.g., avian malaria (13, 14)). Some of these
79 animal diseases have dramatic economic consequences in endemic areas (11, 12), while others are
80 zoonotic diseases, further affecting human populations (15-20).

81 Efficient vector-borne disease surveillance is critical for reducing disease transmission and preventing
82 outbreaks. Past elimination campaigns for vector-borne diseases, usually targeting a specific human
83 pathogen, have often relied on entomological approaches such as widespread insecticide spraying and
84 disruption of larval habitats (21, 22). To be successful, such efforts need to be guided by detailed
85 knowledge of the parasites' and vectors' distributions. Unfortunately, current entomological
86 surveillance approaches are extremely resource-intensive: the collection of samples is time consuming
87 requiring trained personnel, vector species identification is laborious, and the detection of pathogens is
88 expensive since hundreds of mosquitoes typically need to be screened to identify a few infected ones.
89 Consequently, public health officials and vector biologists typically focus on monitoring only a few
90 specific pathogens associated with the most current threats. These constraints are particularly
91 problematic as they hamper the early detection of emerging pathogens and vector surveillance is often
92 implemented in response to reports of clinical cases rather than preventively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

93 We have recently described a sequencing-based method using high-throughput amplicon sequencing to
94 detect known and previously uncharacterized eukaryotic parasites from biological samples in a
95 comprehensive, high-throughput and cost-efficient manner (23). Here, we present the application of this
96 approach to characterize eukaryotic parasites and arboviruses from more than 900 individual *Anopheles*
97 mosquitoes collected in Cambodia, Guinea and Mali, as well as from 25 pools of mosquitoes captured in
98 CDC CO_2-baited light traps in Maryland, USA. We also compare the performance of the assay when
99 screening DNA and RNA from the same samples. Overall, our study demonstrates how this sequencing-
100 based assay could significantly improve monitoring of human and animal vector-borne pathogens.
101

102 **Methods**

103 **Samples**

104 We analyzed a total of 930 individual mosquitoes, as well as 25 pools each containing 50-291
105 mosquitoes (2,589 total) (**Table 1** and **Supplemental Tables 1-3**).
106 First, we analyzed DNA previously extracted from 265 individual *Anopheles* mosquitoes collected in the
107 Cambodian provinces of Pursat, Preah Vihear, and Ratanakiri (24). These mosquitoes were collected
108 using cow- or human-baited tents, human landing collections, CDC light traps and barrier-screen fences
109 and immediately preserved by desiccation upon collection. These 265 *Anopheles* mosquitoes represent
110 22 different species collected between July and August of 2013 (see **Supplemental Table 1** for details).
111 Second, we included DNA samples from 81 individual mosquitoes collected in Bandiagara, Mali. DNA
112 from these samples was extracted using Chelex® 100 (Bio-Rad) after incubation of bisected and
113 homogenized mosquitoes in 1% saponin in PBS.
114 Third, we extracted DNA from 584 individual *Anopheles* mosquitoes collected in six sites in Guinea and
115 preserved in ethanol immediately upon collection. These mosquitoes were collected by human landing
116 catch and pyrethrum spray (**Supplemental Table 2**). Each mosquito was homogenized in 200 µl

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

117 ATL/proteinase K solution using five RNase-free 1 mm zirconium oxide beads in a TissueLyser II for 12
118 minutes at 20 m/s. We centrifuged the solution at 2500 rpm for three minutes and incubated them at
119 55°C for one hour. We performed a second homogenization step for four minutes at 20 m/s followed by
120 a final incubation at 55°C overnight. We then isolated DNA using the Qiagen DNeasy 96 Blood & Tissue
121 Kit according to the manufacturer's instruction and eluted DNA from each sample in 200 µl.
122 Finally, we analyzed 25 pools of mosquitoes collected throughout Prince George's county (Maryland,
123 USA) by the Maryland Department of Agriculture using CO_2-baited light traps (**Supplemental Table 3**).
124 Each pool contains all mosquitoes from one light-trap (~50-291 mosquitoes) and was stored at room
125 temperature for up to 24 hours before long-term storage at -20°C. We homogenized each pool of
126 mosquitoes using a Qiagen TissueLyser II with Teenprep Matrix D 15 ml homogenization tubes (MP
127 Biomedicals) and isolated successively RNA and DNA from each sample using the RNeasy PowerSoil
128 Total RNA kit (Qiagen) with the RNeasy PowerSoil DNA Elution Kit and a final elution volume of 100 µl.

129 **Evaluation of Arbovirus primers**

130 We tested universal flavivirus primers retrieved from the literature (25) on West Nile (n=3), Zika (n=2)
131 and Dengue (n=2) viral RNAs obtained from the American Type Culture Collection (ATCC). We
132 synthesized cDNA from 2 µL of RNA using M-MLV reverse transcriptase (Promega) and random
133 hexamers, and amplified the resulting cDNA with GoTaq® DNA polymerase (Promega) under the
134 following conditions: initial two-minute denaturing step at 95°C followed by 40 cycles of 95°C for 30
135 seconds, 50°C for 30 seconds and 72°C for 40 seconds. A final extension at 72°C for ten minutes was
136 followed by incubation at 4°C. We ran the products on an agarose gel to determine whether each virus
137 RNA was amplifiable.

138 **PCR amplification of pathogen nucleic acids before high-throughput sequencing**

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

139 First, we synthesized cDNA using M-MLV reverse transcriptase (Promega) and random hexamers from
140 either 1) 2 µl of the nucleic acids isolated from the Guinean mosquitoes (i.e., using RNA carried over
141 during the DNA extraction), or 2) 3 µL of RNA extracted from the pools of Maryland mosquitoes.
142 Then, we amplified DNA and cDNA (when available) from each sample, as well as from 176 no-DNA
143 controls, with a total of 11 primer pairs, each targeting a specific taxon known to contain human
144 pathogens (**Table 2**). For each primer pair, we amplified DNA and cDNA using GoTaq® DNA polymerase
145 (Promega) under the following conditions: initial denaturing step at 95°C followed by 40 cycles of 95°C
146 for 30 seconds, 50°C for 30 seconds and 72°C for 30 seconds. A final extension at 72°C for ten minutes
147 was followed by incubation at 4°C. All primers used in these taxon-specific PCRs included 5'-end tails to
148 serve as priming sites for a second PCR. We then pooled all PCR products generated from one sample
149 and performed a second PCR using primers targeting these tails to incorporate, at the end of each
150 amplified molecule, i) a unique oligonucleotide "barcode" specific to each sample and ii) DNA sequences
151 complementary to the Illumina sequencing primers (26, 27) (**Supplemental Figure 1**). Finally, we pooled
152 together the resulting barcoded libraries and sequenced them on an Illumina sequencer to generate an
153 average of 12,703 paired-end reads of 251 or 301 bp per sample.

154 **Bioinformatic analyses**

155 We first separated the reads generated from each sample according to their unique barcodes and
156 merged the overlapping ends of each read pair using PANDAseq (28) to generate consensus DNA
157 sequences and correct sequencing errors (that disproportionally occur at the end of the reads). Note
158 that all primers were designed to amplify DNA sequences shorter than ~450 bp, allowing overlap of at
159 least 50 bp between paired-end reads. All read pairs that did not merge correctly were discarded from
160 further analyses. We identified and trimmed the primer sequences from each read and eliminated all
161 consensus sequences shorter than 100 bp as they likely represent experimental artefacts (e.g., PCR
162 chimeras and primer dimers). Using reads from all samples together, we recorded how many unique

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

163 DNA sequences were obtained and how many reads carried each of these unique DNA sequences.
164 Sequences observed less than ten times in the entire dataset were omitted as they likely resulted from
165 PCR or sequencing errors (26). We then compared each unique DNA sequence to all sequences
166 deposited in the NCBI nt database using BLAST (29) and used custom pipelines
167 (https://github.com/MVesuviusC/2020MosquitoSurveillancePaper) to retrieve the taxonomic
168 information associated with the most similar sequence(s). For each sample, only sequences with at least
169 10 reads and more than 70% identity with an annotated NCBI sequence over the entire sequence length
170 were further considered. This identity cutoff, while low, allows inclusion of results from highly
171 genetically divergent organisms which can then be scrutinized further. This is critical when identifying
172 species without closely related sequences available. If DNA sequences from multiple species were
173 equally similar to one of our sequences, we recorded all corresponding species names. Finally, we
174 summarized, for each mosquito, the parasite species or virus identified, the percentage identity
175 between the reads and the most similar NCBI sequence(s), and the number of reads supporting the
176 identification in this sample.

177 **Phylogenetic analyses**

178 To better characterize specific DNA sequences with ambiguous species identification, we analyzed these
179 sequences together with orthologous sequences from closely related species. Briefly, we used
180 PrimerTree (26) to retrieve NCBI orthologous DNA sequences from all species of the targeted taxon. We
181 aligned these sequences with the DNA sequence(s) amplified from the mosquito(es) using MAFFT (30)
182 and reconstructed neighbor-joining trees using MEGA (31) to estimate the phylogenetic position of the
183 amplified DNA sequences.

184 **Further determination of taxonomical assignments**

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

185 To improve species identification when multiple species had identical DNA sequences, or improve
186 phylogenetic analyses of unknown sequences, we amplified and sequenced specifically chosen DNA loci
187 from pathogens within the same mosquitoes.
188 For differentiating *Theileria* species, we used previously published primers (GGCGGCGTTTATTAGACC,
189 TCAATTCCTTTAAGTTTCAGCC) to amplify an informative portion of the 18S rRNA gene [31] using DNA
190 from 19 samples identified as *Theileria* positive by high-throughput sequencing. Amplification was
191 conducted under the following conditions: initial denaturing step at 95°C for two minutes followed by 40
192 cycles of 95°C for 30 seconds, 50°C for 30 seconds and 72°C for 40 seconds. A final extension at 72°C for
193 five minutes was followed by incubation at 4°C. Since gel electrophoresis revealed off-target
194 amplification (multiple bands), we used a Pasteur pipette to collect a core from the agarose gel,
195 corresponding to the expected 900 bp PCR product, and dissolved it in 100 µl of water at 60°C for 20
196 minutes. We then re-amplified 10 µl of this DNA using 35 PCR cycles with the same conditions. After gel
197 electrophoresis, we treated the PCR reaction with 0.046 µl of Exonuclease I (NEB) and 0.4625 µl of
198 Shrimp alkaline phosphatase (Affymetrix) at 37°C for 30 minutes, with a final five-minute inactivation
199 step at 95°C. We then Sanger sequenced each PCR product in both directions using the forward and
200 reverse primers. We manually trimmed the reads and merged them using Flash (32). We aligned the
201 reads, along with known *Theileria* sequences from the NCBI nucleotide database, using MAFFT (30, 33)
202 and generated a neighbor joining tree with 500 bootstraps and plotted it in MEGA7 (34).
203 To identify the species of the filarial worms detected in two individual mosquitoes, we designed primers
204 to amplify a 3.5 kb portion of the mitochondrial DNA. Briefly, we downloaded all available filarial worm
205 (Filarioidea) mitochondrial sequences from the NCBI nucleotide database, aligned them, generated a
206 consensus sequence and designed primers using primer3 (35). We then used these primers
207 (TTCGTCGTGAGACAGAGCGG, AGGCCATTGACGGATGGTTTGTAC) to amplify DNA from the two positive
208 mosquitoes using the Expand™ Long Range dNTPack kit (Sigma) using the following conditions: initial

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

209 denaturing step at 95°C for two minutes followed by 45 cycles of 92°C for 30 seconds, 55°C for 30
210 seconds and 68°C for five minutes. A final extension at 68°C for ten minutes was followed by incubation
211 at 4°C. We then performed a second PCR to add 10 bp barcodes to the 5' end of both forward and
212 reverse primers to allow differentiating both samples after sequencing. The two barcodes differed by 8
213 and 7 bases for the forward and reverse primers, respectively, with no more than 2 identical bases in a
214 row (**Supplemental Table 4**). For this second PCR, we used the following conditions: initial denaturing
215 step at 95°C for two minutes followed by 10 cycles of 92°C for 30 seconds, 55°C for 30 seconds and 68°C
216 for five minutes. A final extension at 68°C for ten minutes was followed by incubation at 4°C. We
217 purified the amplicons using AMPure XP beads (Beckman Coulter) (2:1 DNA:beads ratio) and then
218 combined equimolar amounts of each barcoded PCR product before circular consensus sequencing on a
219 PacBio Sequel. We then generated a consensus sequence for each sample and aligned these sequences
220 to known nematode mitochondrial sequences using Mafft (30) and generated a neighbor joining tree in
221 MEGA (36).

222 **Assessment of the dynamics of viral and mosquito RNA degradation**

223 To assess the dynamics of viral RNA degradation over time, we analyzed colony *Culex pipiens*
224 mosquitoes known to carry *Culex flavivirus*. The colony was initiated from diapausing adult *Culex pipiens*
225 that were collected from Oak Lawn and Des Plaines, IL, on 2/8/10. These two collections were combined
226 to make one colony, which was determined to be *Culex flavivirus* positive according reverse
227 transcriptase PCR (25). We examined three pools of five mosquitoes for each condition (i.e., stored with
228 no preservative, in ethanol or in RNAlater (Invitrogen)) and at each time point (i.e., fresh, after two-
229 week or after four-week storage at room temperature). After 0, 2 or 4 weeks at room temperature, the
230 mosquitoes were stored at -80°C until RNA isolation. We isolated RNA from each pool of mosquitoes
231 using Qiazol (Qiagen) and eluted into 50 µl. We synthesized cDNA from 7 µl of RNA using m-MLV

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

232 (Promega) with random hexamers for PCRs using *Culex* primers and, separately, on 2 µl of RNA for PCRs
233 using flavivirus primers.
234 For each pool of 5 *Culex* mosquitoes from the *Culex flavivirus*-infected colony, we performed
235 quantitative reverse transcriptase PCR (qRT-PCR) to quantify the amount of mosquito and viral RNAs
236 using the primers Culex_flavivirus_3F (TGCGAARGATCTDGAAGGAG) - Culex_flavivirus_3R
237 (CACGCACAACAAGACGATRA) targeting the virus sequence, and Culicinae_Cox1_379_F
238 (AYCCHCCTCTTTCATCTGGA) - Culicidae_Cox1_670_R (CCTCCTCCAATTGGRTCAAAG) targeting mosquito
239 RNA. We used Perfecta SYBR green PCR mastermix (Quantabio) with the following conditions: initial 15-
240 minute denaturing step at 95°C followed by 40 cycles of 95°C for 30 seconds, 55°C (Culex primers) or
241 50°C (flavivirus primers) for 30 seconds and 72°C for one minute (Culex primers) or 40 seconds (flavivirus
242 primers). We performed standard cycle threshold and melt curve analysis afterwards using default
243 settings.
244
245 **Results**
246 **Amplicon sequencing for high-throughput characterization of microorganisms in mosquitoes**
247 We analyzed 265 *Anopheles* mosquitoes collected in Cambodia, 665 *Anopheles* mosquitoes collected in
248 Guinea and Mali as well as the content of 25 light traps, each containing 50-291 mosquitoes, collected in
249 Maryland, USA. We screened each sample for a wide range of eukaryotic parasites using 10 primer sets
250 designed to amplify DNA from all species of the taxa known to include human pathogens:
251 Apicomplexans, Kinetoplastids, Parabasalids, nematodes, Platyhelminthes and Microsporidians (**Table**
252 **2**). We also screened RNA extracted from the individual African *Anopheles* and from the pools of
253 mosquitoes from Maryland for flaviviruses (see Materials and Methods, **Table 2**). After taxon-specific
254 amplification, we pooled all PCR products generated from the same mosquito together, barcoded them
255 and sequenced all libraries to generate an average of 12,703 paired-end reads per sample

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

256 (**Supplemental Figure 1**). After merging read pairs, stringent quality filters and removal of the products
257 of off-target amplification (e.g., *Anopheles* and bacteria DNA sequences), we obtained 61,177 unique
258 DNA sequences, each represented by ten reads or more, and accounting in total for 6,796,105 reads
259 (**Supplemental Table 5**). These sequences were amplified with all primers and from a total of 185
260 samples: 42 out of 265 Cambodian mosquitoes (16%), 120 out of 665 African mosquitoes (18%), and 23
261 out of the 25 pools (92%) of mosquitoes collected in Maryland were positive for at least one of the taxa
262 tested. On average, each sequence was supported by 1,306 reads per sample (range: 10-43,440). By
263 contrast, out of 176 negative controls, only 12 (7%) yielded any sequence from the targeted taxa and
264 those were represented by 213 reads on average (range: 10-3,539).

265 **Identification of eukaryotic parasites**

266 We retrieved DNA sequences identical to sequences previously amplified from *Theileria* parasites from
267 22 African and 15 Cambodian mosquitoes, as well as from seven of the Maryland traps. *Theileria*
268 sequences were successfully amplified with both the Apicomplexa and Eimeronia primer pairs. All
269 samples positive for *Theileria* with the Eimeronia primers were also positive with the Apicomplexa
270 primers. On the other hand, the Eimeronia primers provided sufficient information to assign each
271 sequence to a single species, while the sequences amplified with Apicomplexa primers were unable to
272 differentiate among the *Theileria* species (see also below). We detected sequences identical to
273 *Plasmodium falciparum* in eight African samples and two Cambodian samples, while sequences most
274 similar (82.0%-99.5% identity) to bird *Plasmodium* species were amplified from 20 of the 25 traps in
275 Maryland (**Table 3**). We also amplified a sequence that was identical to several *Babesia* species (100%
276 identity) in one trap by two different primer pairs. Finally, we detected DNA from a known
277 apicomplexan parasite of mosquitoes, *Ascogregarina barretti* (37), in two of the traps.

278 From all individual mosquitoes, only one Cambodian *Anopheles* yielded a Kinetoplast sequence that was
279 most similar to *Strigomonas culicis* (96.9% identity). By contrast, 22 of the traps were positive for

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

280 Kinetoplasts, yielding sequences similar to sequences from *Angomonas*, *Blastocrithidia, Blechomonas,*
281 *Crithidia, Leptomonas, Paratrypanosoma, Strigomonas, Trypanosoma, Trypanosomatidae, Wallaceina* or
282 *Zelonia* (with 90.3%-100% identity, except for one sequence that matched *Trypanosoma theileri* at
283 82.7% identity) (**Table 3**).

284 Many sequences were amplified using the Microsporidia primers: 72 African mosquitoes were positive
285 with sequences similar or identical to *Culicospora*, *Encephalitozoon*, *Hazarida, Microsporidium* and
286 *Parathelohania* (88.3%-100% identity), while 13 Cambodian samples yielded sequences similar or
287 identical to *Agglomerata*, *Parathelohania* and *Senoma* (91.3%-100% identity) (**Table 3**). Twenty-two
288 traps also yielded Microsporidia sequences closely matching those of *Amblyospora, Andreanna,*
289 *Culicospora, Microsporidium, Parathelohania* and *Takaokaspora* (with 76.0%-100% identity).

290 Regarding parasites from the Parabasalia group, four African mosquitoes were positive for
291 *Tetratrichomona, Trichomonas* or *Tritrichomonas* with high sequence similarity (94.8%-100%) while a
292 single Cambodian mosquito was positive for *Trichomitus* (98.7% identity). No Parabasalia were detected
293 in the Maryland traps.

294 We detected Platyhelminthes sequences in four African mosquitoes, all similar to *Schistosoma mansoni*
295 (92.1%-100% identity). Three Cambodian mosquitoes yielded sequences most similar to those of either
296 *Pleurogenoides or Pleurogenes* (94.9%-95.4% identity). Three traps in Maryland were positive for
297 Platyhelminthes, with sequences most similar to *Haematoloechus* (98.7% identity).

298 The taxonomic resolution of the nematode primers was lower than that of the other taxon-specific
299 primer pairs and the amplified sequences often matched multiple species (or even genera). We
300 amplified nematode sequences from 33 African *Anopheles*, including sequences most similar to
301 *Abursanema*, *Acanthocheilonema*, *Auanema*, *Caenorhabditis*, *Dipetalonema*, *Filarioidea*, *Loa*,
302 *Loxodontofilaria*, *Madathamugadia*, *Onchocerca*, *Pelecitus*, *Setaria* or *Trichuris,*although the sequence
303 similarity (95.3-100%) clearly indicated that, in some cases, the exact identity of the species was

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

304 unknown (see also below). Ten Cambodian mosquitoes were positive for *Setaria digitata* (100% identity)
305 while other mosquitoes yielded sequences that matched *Setaria* and one or more of the following
306 genera: *Aproctella*, *Breinlia*, *Dipetalonema*, *Dirofilaria*, *Loa*, *Loxodontofilaria*, *Madathamugadia*,
307 *Onchocerca*, *Pelecitus*. Nineteen different traps from Maryland produced nematode sequences with
308 particularly high read counts of *Setaria*, *Yatesia* and *Dirofilaria* sequences (98.9% - 100% identity). Other
309 genera detected in the traps included *Acanthocheilonema*, *Aproctella*, *Cercopithifilaria*, *Choriorhabditis*,
310 *Dipetalonema*, *Elaeophora*, *Filarioidea*, *Loa*, *Loxodontofilaria*, *Onchocercidae*.
311 Overall, using this single assay, we screened over 3,500 mosquitoes from three geographic locations and
312 identified DNA sequences from numerous microorganisms encompassing six classes, 12 orders and 23
313 families (**Table 3**).
314 **Identification of Flaviviruses in mosquitoes**
315 To detect and identify flaviviruses, we used a primer pair predicted *in silico* to amplify a wide range of
316 flaviviruses, including all known human pathogens (25), and we validated that these primers successfully
317 amplified cDNA generated from West Nile, Zika and Dengue viruses. Out of 665 individual African
318 mosquitoes, three were positive for viruses most similar to *Anopheles flavivirus variants 1* and *2* (87.2%-
319 99.1% identity) (**Figure 1 and Supplemental Figure 2**) and one was positive for a virus similar to *Culex*
320 *flavivirus* (99.1% identity). Seven Maryland traps (24%) were positive for flaviviruses. These viruses were
321 most similar to the *Calbertado* and *Nienokoue* flaviviruses, although the percent identity was very low
322 (71.1%-74.3%) and they clearly separated from those viruses in phylogenetic analysis (**Figure 1 and**
323 **Supplemental Figure 2**). These sequences likely derive from viruses that have not been sequenced yet
324 but, since that they cluster with other mosquito flaviviruses (**Figure 1 and Supplemental Figure 2**), it is
325 likely that they represent mosquito-infecting viruses rather than new human pathogens.
326 One limitation of our study is that the mosquitoes collected in Maryland, USA were, as typical in many
327 entomological surveys, stored at room temperature upon collection which might have affected RNA

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

328 preservation. To assess the stability of viral and mosquito RNA in samples stored at room temperature,
329 we kept pools of colony mosquitoes known to be infected with *Culex flavivirus* at room temperature for
330 up to four weeks after collection, with and without preservative (ethanol or RNAlater). After RNA
331 extraction and cDNA synthesis, we determined the amount of mosquito and virus RNA amplifiable using
332 real-time PCR (see Material and Methods for details). Without preservative, the mosquito RNA was
333 largely degraded after two weeks (detectable in only one of three replicates) and undetectable after
334 four weeks (**Supplemental Figure 3**). By comparison, under the same conditions, viral RNA was still
335 detectable after four weeks (**Supplemental Figure 3**). As expected, when the mosquitoes were
336 preserved in either ethanol or RNAlater, neither viral nor mosquito RNA showed major change in
337 concentration over four weeks at room temperature.

338 **Follow-up phylogenetic studies**

339 The taxon-specific primers used in the high-throughput sequencing assay were designed to amplify all
340 members of the chosen group while avoiding off-target amplification and providing as much taxonomic
341 information as possible. However, these criteria, combined with the requirement for short sequences
342 (to be sequenceable on a massively parallel sequencer) sometimes limits their resolution.

343 Thus, the Apicomplexa primers amplified multiple *Theileria* sequences but did not distinguish among
344 species. We therefore amplified a longer DNA sequence (900 bp) of the 18S rRNA locus from the
345 *Theileria*-positive African and Cambodian mosquitoes and sequenced them using Sanger sequencing
346 technology. Phylogenetic analysis of these longer sequences, together with known *Theileria* species
347 sequences deposited in NCBI, showed that the parasites amplified from the Cambodian mosquitoes
348 were closely related to *T. sinensis,* while those from African mosquitoes were most closely related to *T.*
349 *velifera* and *T. mutans* (**Figure 2**).

350 We also detected, in several African mosquitoes, filarial worm sequences whose taxonomic assignment
351 was uncertain. One sequence was 100% identical to both *Loa loa* and *Dipetalonema sp. YQ-2006* (also

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

352 known as *Mansonella*) while the sequence obtained from the same mosquitoes using a different primer
353 pair was also most similar to *Dipetalonema* (*Mansonella*) but with 99.2% identity. To clarify the
354 taxonomy of these sequences, we used PacBio long read technology to sequence 3.5kb of filarial worm
355 mitochondrial DNA (amplified by long range PCR from these two mosquitoes). We compared these
356 sequences to known filarial worm mitochondrial DNA sequences and found that these were most similar
357 to, but distinct from, *Mansonella perstans* (94 and 96 nucleotide differences or~97.0% identity), while
358 *Loa loa* was much more distantly related (~83.3% identity) (**Figure 3**). The genetic distance between
359 *Mansonella perstans* and *Loa loa* in this tree was much higher (519 nucleotide differences, 83.5%
360 identity) than using short amplicon data where these two sequences were identical, providing greater
361 confidence in the phylogenetic analysis. We concluded that the filarial worms are most likely either
362 *Mansonella perstans* or a very closely related species.

363 **Analysis of individual vs pooled mosquitoes**

364 We analyzed both individual mosquitoes and pools of 50-291 mosquitoes. For the pools, 23 out of the
365 25 produced sequences demonstrating that the amplification of pools of up to 291 mosquitoes is
366 feasible without significant PCR inhibition. Out of the 930 individual mosquitoes, 162 (17.4%) had at
367 least ten reads from one or more parasites or arboviruses. By comparison, 23 out of the 25 traps (92%)
368 yielded such sequences, meaning that fewer samples needed to be screened to detect pathogens when
369 samples were pooled. Note however that the individual mosquitoes and the CDC traps were not
370 collected from the same geographic locations and this could possibly cofound the results described.

371 **Analysis of DNA vs. RNA extracted from the same mosquito traps**

372 All primers used for detecting parasites and viruses are located within single-exon genes and can amplify
373 either DNA or cDNA with the same efficiency. However, the PCRs target genes that are typically highly
374 expressed (e.g., ribosomal RNA genes) and we would therefore expect many more copies of RNA than
375 DNA per cell, (although this could be diminished by the faster degradation of RNA molecules compared

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

376 to DNA). To evaluate the relative sensitivity of our assay for screening DNA and RNA, we compared the
377 results obtained by analyzing matched DNA and RNA isolated from the same mosquito pools. We found
378 that for Spirurida, Kinetoplast, Microsporidia and *Plasmodium* PCR assays, 62.2% of the sequences
379 identified were detected only in the cDNA sample and not in the corresponding DNA sample from the
380 same trap. For those cases where a sequence was detected in both cDNA and DNA from a given trap,
381 the cDNA yielded more reads in 89 of 119 instances (with, on average, 24.5-fold more reads). Read
382 counts were higher in the DNA for only 29 of 119 cases with an average fold difference of 2.8 (one had
383 equal read counts in cDNA and DNA). Out of these 29 cases, 22 (75.9%) of the sequences were most
384 similar to *Trypanosoma* species despite these sequences representing only 17.4 of all sequences. On
385 average, the cDNA samples produced 258 – 1,169 more reads per hit than the matching DNA samples
386 (**Figure 3**), despite the storage of the samples at room temperature without preservative for more than
387 24 hours.
388
389 **Discussion**
390 Vector-borne disease surveillance is an essential component of infectious disease control as it can
391 enable rapid detection of outbreaks and guide targeted elimination efforts (e.g., through insecticide
392 spraying). However, current approaches are extremely demanding in regards to human and financial
393 resources, both for the sample collection and the identification of potential pathogens. Consequently,
394 public health officials and vector biologists often have to focus on a handful of parasites associated with
395 the most current threats. Current detection approaches also often lead to duplicated efforts, as
396 different agencies interested in specific pathogens perform sample collection independently and have a
397 high risk of failing to detect emerging pathogens until they cause outbreaks. Here, we describe
398 application of a genomic assay that allows identification of a wide range of pathogens that can cause

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

399 human and animal diseases, as well as of parasites of the vector that could potentially be useful as
400 biological controls.
401 The analyses of several hundred mosquitoes collected in Cambodia, Mali, Guinea and Maryland revealed
402 well-known human pathogens including *P. falciparum*, which was the target of the initial study of the
403 Cambodian samples (24). In addition, we detected *Theileria* species and *Setaria digitata*, which cause
404 livestock diseases in Southeast Asia (38-42). While we were initially unable to conclusively determine
405 the exact *Theileria* species with our initial assay, targeted follow-up studies using longer amplicons and
406 Sanger sequencing (**Figure 2**) revealed that the sequences amplified from the African mosquitoes were
407 most closely related to *T. velifera* and *T. mutans,* which are both known to infect African cattle (43),
408 whereas the Cambodian mosquitoes carried sequences most closely related to *T. sinensis,* a species that
409 infects cattle in China (44).
410 *Theileria* parasites are transmitted by ticks, not mosquitoes, and the DNA sequences recovered likely
411 derive from parasites taken up by the mosquitoes during a blood meal but likely not transmissible to
412 another hosts. The *Schistosoma* species detected in mosquitoes from Africa also likely result from
413 parasites present in a bloodmeal. In this regard, it is interesting to note that when one considers the
414 samples collected in Maryland and analyzed with both DNA and RNA, the read counts (a proxy for the
415 abundance of extracted molecules) for transmissible parasites (e.g., *Plasmodium*) or parasites of the
416 mosquitoes (e.g., *Crithidia*, *Strigomonas* and *Takaokaspora*) were typically higher in the RNA samples
417 than in the matched DNA samples while the opposite was true for parasites "sampled" during the blood
418 meal but unlikely to develop in *Anopheles* mosquitoes (e.g., *Theileria*, *Trypanosoma*) (**Figure 4,**
419 **Supplemental Figure 4 and Supplemental Table 6**). We speculate that this difference is due to the
420 difference between developing, live, parasites still synthesizing RNA molecules and dead (possibly
421 digested) parasites for which the RNA is slowly being degraded. Comparison of DNA and RNA from the

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

422 same mosquito could perhaps provide a tool to differentiate transmissible parasites from those sampled
423 by the vector.

424 We also identified, in two African mosquitoes, sequences similar to known filarial worms but identical to
425 multiple sequences present in the database. Using this information, we characterized longer DNA
426 sequences and showed that these two mosquitoes likely carried *Mansonella perstans* parasites. Since
427 the PCR primers are designed to amplify any member of the selected taxa, they can reveal the presence
428 of novel pathogens as long as they are phylogenetically related to known parasites. This feature is a key
429 advantage of our assay for vector-borne disease surveillance as it may enable early detection of
430 emerging pathogens and zoonoses and provide a basis for rapid response.

431 In addition to known human parasites and potential emerging pathogens, this single-stop assay also
432 provides another source of information valuable for vector-borne disease control: 9% of the individual
433 mosquitoes and 62% of pooled mosquito samples screened yielded sequences of microsporidians
434 related to well-characterized arthropod parasites, which could potentially be used to guide the
435 development of targeted biological vector control. This ability to detect multiple parasites at once in a
436 high-throughput manner and across a wide range of taxonomical groups could reduce duplication of
437 collection efforts and costs, as mosquitoes collected for one purpose could be screened for many
438 parasites affecting both humans and animals. In addition, comprehensive characterization of the
439 parasites present in a given mosquito may also improve our understanding of the general factors
440 regulating infection and transmission: several studies have shown that immunity and previous infections
441 can influence the response of mosquitoes to human parasites and their transmission (45-47) and
442 information of current infections of wild-caught mosquitoes could, for example, significantly improve
443 our assessment of their vector capacity.

444 Several of the infectious diseases that have recently caused major public health challenges by spreading
445 outside of their typical range(7, 8) or emerging as novel human infectious diseases(5, 6), are caused by

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

446 viruses transmitted by mosquitoes. We therefore extended our assay to capture, using the same
447 approach as for eukaryotic parasites, both known and novel flaviviruses. Since flaviviruses are RNA
448 viruses and RNA degrades much faster than DNA, we first examined how nucleic acid degradation
449 influenced our ability to detect virus over time. To test RNA preservation, we collected mosquitoes
450 known to carry *Culex flavivirus* and isolated RNA from pools of five mosquitoes, either immediately
451 frozen or kept at room temperature for two or four weeks, with either no preservative, ethanol or
452 RNAlater. The mosquitoes stored in preservatives had minimal loss of viral (and mosquito) RNAs as
453 determined by qRT-PCR (**Supplemental Figure 3**). Even when stored without preservatives, viral RNA
454 were detectable after 4 weeks at room temperature (although with a reduction of, on average, 10.7 PCR
455 cycles), demonstrating a remarkable stability of the RNA, possibly due to protection provided by the
456 viral capsid (by contrast very little mosquito RNA remained amplifiable after two weeks at room
457 temperature, **Supplemental Figure 3**). As a proof-of-principle and to demonstrate the potential of this
458 approach for viral disease surveillance, we screened the Maryland mosquito pools and the individual
459 African mosquitoes for flaviviruses. We identified several viruses, distinct from known viruses (**Figure 1**)
460 and, based on their phylogenic position, likely to infect mosquitoes rather than humans.

461 Based on the results described above, we believe that this single high-throughput assay can provide a
462 wide range of information critical for vector-borne disease researchers and public health officials.
463 However, several limitations need to be noted. One caveat is that, whereas false positive detection of a
464 species is highly unlikely (aside from laboratory cross-contamination), several factors could lead to false
465 negatives. Thus, while the primers were designed to amplify all known sequences of a given taxon as
466 effectively as possible, nucleotide differences at the primer binding sites could prevent efficient
467 amplification of a specific species. This potential problem could be particularly problematic if several
468 related parasites are present in the same sample but are differentially amplified: for example, it could
469 be possible that a *Plasmodium* parasite might be mis-detected if the sequences generated by an

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

470 Apicomplexan primer pair are out competed by *Theileria* sequences. Similarly, poor preservation of the
471 nucleic acids in one sample could also lead to false negatives. False negatives could also occur for
472 stochastic reasons: if only a few parasite cells are present in one sample (*e.g.*, an *Anopheles* mosquito
473 infected by a *Plasmodium* ookinete) it is possible that no DNA will be present in the PCR reaction
474 (especially if the extract gets divided across many reactions). One approach to circumvent this limitation
475 could be to test cDNA instead of DNA (48, 49): our analyses of the Maryland mosquitoes showed that,
476 for many primer sets, amplification of cDNA resulted in higher read counts than amplification of DNA
477 extracted from the same samples, despite the sub-optimal preservation of these samples. Another
478 limitation is the specificity of taxonomic assignment. As discussed above, if the sequenced amplicon
479 does not contain enough information to distinguish similar species, subsequent experiments may be
480 required to confirm pathogen identity for important detection events.
481 Finally, we showed that analyses of fairly large pools of mosquitoes (up to ~300 mosquitoes) were
482 possible with our assay. This feature could be extremely useful in specific situations, such as for
483 efficiently detecting emerging pathogens, monitoring the spread of pathogens into new regions, or for
484 validating the success of elimination control programs.
485

486 **Conclusion**

487 This study demonstrates how our high-throughput, one-stop assay could efficiently complement current
488 toolkits to prevent vector-borne diseases by providing a comprehensive description of known and
489 emerging human viruses and parasites, informing on animal pathogens that could affect a region's
490 economy, and indicating possible biological control candidates that could be used against these disease
491 vectors. One additional feature of this sequencing-based assay is the ease of customizing it to different
492 settings and research questions. Since the assay relies on PCR primers, it is straightforward to add and

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

493 remove primers for specific taxa of interest, or to combine them with additional PCRs to characterize,
494 for example, the source of the blood meal (27).

495 **Acknowledgements**

496 We wish to thank our collaborators at the National Center for Parasitology, Entomology, and Malaria
497 Control (CNM) in Phnom Penh, Stop Palu and the National Malaria Control Program in Guinea and at the
498 Malaria Research and Training Center, University of Science, Techniques and Technologies, Bamako,
499 Mali for their help with the sample collection and processing, as well as Erin Haser, Rachel Koch, Jessica
500 Moskowitz, Rachel Shuck, and Elizabeth Allan. We also thank Stop Palu (RTI), the U.S. President's
501 Malaria Initiative, and all those participating in collections for support in the sample collection and
502 processing in Guinea. The findings and conclusions in this report are those of the authors and do not
503 necessarily represent the official position of the Centers for Disease Control and Prevention.

504 **Conflict of interest statement:**

505 The authors have stated explicitly that there are no conflicts of interest in connection with this article.

506 **Author contributions**

507 MVC, BSTL, MF and DS designed the study. BSTL, MF, KK, DC, YB, MK, DC, AKK, OKD, MAT, CVP, MT, SI,
508 JD and BP collected and/or provided access to the mosquito samples. MVC, HB, BSTL and DB performed
509 the laboratory experiments. MVC analyzed the data. MVC and DS prepared the manuscript. All authors
510 read and approved the final manuscript.

511 **Availability of data and material**

512 The datasets generated for the current study are available in the NCBI SRA repository (Accession
513 numbers SRR12797126 - SRR12797220, SRR12797360 - SRR12797683, SRR12796164 - SRR12796923 and
514 SAMN16182375 - SAMN16183134)

515 **Funding**

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

516 This study was supported by DS and by the Intramural Research Program (NIAID-NIH) to BSTL. SRI was
517 funded by the U.S. President's Malaria Initiative.
518

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

519 **References**

520 1. Ecker, D. J., Sampath, R., Willett, P., Wyatt, J. R., Samant, V., Massire, C., Hall, T. A., Hari, K.,
521 McNeil, J. A., Buchen-Osmond, C., and Budowle, B. (2005) The Microbial Rosetta Stone
522 Database: a compilation of global and emerging infectious microorganisms and bioterrorist
523 threat agents. *BMC microbiology* **5**, 19
524 2. WHO. (2017) Vector-borne diseases. In *Fact sheet* (Organization, W. H., ed)
525 3. (2008) In *Vector-Borne Diseases: Understanding the Environmental, Human Health, and*
526 *Ecological Connections, Workshop Summary*, Washington (DC)
527 4. Collaborators, G. B. D. M. (2018) Global, regional, and national age-sex-specific mortality and life
528 expectancy, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017.
529 *Lancet* **392**, 1684-1735
530 5. Gorshkov, K., Shiryaev, S. A., Fertel, S., Lin, Y. W., Huang, C. T., Pinto, A., Farhy, C., Strongin, A. Y.,
531 Zheng, W., and Terskikh, A. V. (2018) Zika Virus: Origins, Pathological Action, and Treatment
532 Strategies. *Frontiers in microbiology* **9**, 3252
533 6. Liu, Z. Y., Shi, W. F., and Qin, C. F. (2019) The evolution of Zika virus from Asia to the Americas.
534 *Nat Rev Microbiol* **17**, 131-139
535 7. Higuera, A., and Ramirez, J. D. (2019) Molecular epidemiology of dengue, yellow fever, Zika and
536 Chikungunya arboviruses: An update. *Acta Trop.* **190**, 99-111
537 8. Sejvar, J. J. (2016) West Nile Virus Infection. *Microbiol Spectr* **4**
538 9. McCall, J. W., Genchi, C., Kramer, L. H., Guerrero, J., and Venco, L. (2008) Heartworm disease in
539 animals and humans. *Adv. Parasitol.* **66**, 193-285
540 10. Otranto, D., Dantas-Torres, F., Brianti, E., Traversa, D., Petric, D., Genchi, C., and Capelli, G.
541 (2013) Vector-borne helminths of dogs and humans in Europe. *Parasit Vectors* **6**, 16
542 11. Nene, V., Kiara, H., Lacasta, A., Pelle, R., Svitek, N., and Steinaa, L. (2016) The biology of Theileria
543 parva and control of East Coast fever - Current status and future trends. *Ticks Tick Borne Dis* **7**,
544 549-564
545 12. Gachohi, J., Skilton, R., Hansen, F., Ngumi, P., and Kitala, P. (2012) Epidemiology of East Coast
546 fever (Theileria parva infection) in Kenya: past, present and the future. *Parasit Vectors* **5**, 194
547 13. Clark, N. J., Clegg, S. M., and Lima, M. R. (2014) A review of global diversity in avian
548 haemosporidians (Plasmodium and Haemoproteus: Haemosporida): new insights from
549 molecular data. *Int. J. Parasitol.* **44**, 329-338
550 14. Lapointe, D. A., Atkinson, C. T., and Samuel, M. D. (2012) Ecology and conservation biology of
551 avian malaria. *Ann. N. Y. Acad. Sci.* **1249**, 211-226
552 15. Cox-Singh, J., Davis, T. M., Lee, K. S., Shamsul, S. S., Matusop, A., Ratnam, S., Rahman, H. A.,
553 Conway, D. J., and Singh, B. (2008) Plasmodium knowlesi malaria in humans is widely distributed
554 and potentially life threatening. *Clin. Infect. Dis.* **46**, 165-171
555 16. Daneshvar, C., William, T., and Davis, T. M. E. (2018) Clinical features and management of
556 Plasmodium knowlesi infections in humans. *Parasitology* **145**, 18-31
557 17. White, N. J. (2008) Plasmodium knowlesi: the fifth human malaria parasite. *Clin. Infect. Dis.* **46**,
558 172-173
559 18. Brasil, P., Zalis, M. G., de Pina-Costa, A., Siqueira, A. M., Junior, C. B., Silva, S., Areas, A. L. L.,
560 Pelajo-Machado, M., de Alvarenga, D. A. M., da Silva Santelli, A. C. F., Albuquerque, H. G., Cravo,
561 P., Santos de Abreu, F. V., Peterka, C. L., Zanini, G. M., Suarez Mutis, M. C., Pissinatti, A.,
562 Lourenco-de-Oliveira, R., de Brito, C. F. A., de Fatima Ferreira-da-Cruz, M., Culleton, R., and
563 Daniel-Ribeiro, C. T. (2017) Outbreak of human malaria caused by Plasmodium simium in the
564 Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation. *Lancet Glob Health*
565 **5**, e1038-e1046

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

566 19. McCarthy, J., and Moore, T. A. (2000) Emerging helminth zoonoses. *Int. J. Parasitol.* **30**, 1351-
567 1360
568 20. Cutler, S. J., Fooks, A. R., and van der Poel, W. H. (2010) Public health threat of new, reemerging,
569 and neglected zoonoses in the industrialized world. *Emerg. Infect. Dis.* **16**, 1-7
570 21. Benelli, G., and Beier, J. C. (2017) Current vector control challenges in the fight against malaria.
571 *Acta Trop.* **174**, 91-96
572 22. Walker, K., and Lynch, M. (2007) Contributions of Anopheles larval control to malaria
573 suppression in tropical Africa: review of achievements and potential. *Med. Vet. Entomol.* **21**, 2-
574 21
575 23. Cannon, M. V., Bogale, H., Rutt, L., Humphrys, M., Korpe, P., Duggal, P., Ravel, J., and Serre, D.
576 (2018) A high-throughput sequencing assay to comprehensively detect and characterize
577 unicellular eukaryotes and helminths from biological and environmental samples. *Microbiome* **6**,
578 195
579 24. St Laurent, B., Oy, K., Miller, B., Gasteiger, E. B., Lee, E., Sovannaroth, S., Gwadz, R. W.,
580 Anderson, J. M., and Fairhurst, R. M. (2016) Cow-baited tents are highly effective in sampling
581 diverse Anopheles malaria vectors in Cambodia. *Malar J* **15**, 440
582 25. Patel, P., Landt, O., Kaiser, M., Faye, O., Koppe, T., Lass, U., Sall, A. A., and Niedrig, M. (2013)
583 Development of one-step quantitative reverse transcription PCR for the rapid detection of
584 flaviviruses. *Virol J* **10**, 58
585 26. Cannon, M. V., Hester, J., Shalkhauser, A., Chan, E. R., Logue, K., Small, S. T., and Serre, D. (2016)
586 In silico assessment of primers for eDNA studies using PrimerTree and application to
587 characterize the biodiversity surrounding the Cuyahoga River. *Scientific reports* **6**, 22908
588 27. Logue, K., Keven, J. B., Cannon, M. V., Reimer, L., Siba, P., Walker, E. D., Zimmerman, P. A., and
589 Serre, D. (2016) Unbiased Characterization of Anopheles Mosquito Blood Meals by Targeted
590 High-Throughput Sequencing. *PLoS Negl Trop Dis* **10**, e0004512
591 28. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., and Neufeld, J. D. (2012)
592 PANDAseq: paired-end assembler for illumina sequences. *BMC bioinformatics* **13**, 31
593 29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment
594 search tool. *J. Mol. Biol.* **215**, 403-410
595 30. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7:
596 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780
597 31. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular
598 Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729
599 32. Magoc, T., and Salzberg, S. L. (2011) FLASH: fast length adjustment of short reads to improve
600 genome assemblies. *Bioinformatics* **27**, 2957-2963
601 33. Katoh, K., Rozewicki, J., and Yamada, K. D. (2017) MAFFT online service: multiple sequence
602 alignment, interactive sequence choice and visualization. *Brief Bioinform*
603 34. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis
604 Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870-1874
605 35. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G.
606 (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115
607 36. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: Molecular Evolutionary
608 Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547-1549
609 37. Siegel, J. P., Novak, R. J., and Maddox, J. V. (1992) Effects of Ascogregarina barretti
610 (Eugregarinida: Lecudinidae) infection on Aedes triseriatus (Diptera: Culicidae) in Illinois. *J. Med.*
611 *Entomol.* **29**, 968-973

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

612 38. Nakano, H., Tozuka, M., Ikadai, H., Ishida, H., Goto, R., Kudo, N., Katayama, Y., Muranaka, M.,
613 Anzai, T., and Oyamada, T. (2007) Morphological survey of bovine Setaria in the abdominal
614 cavities of cattle in Aomori and Kumamoto Prefectures, Japan. *J. Vet. Med. Sci.* **69**, 413-415
615 39. Weerasooriya, G., Sivakumar, T., Lan, D. T., Long, P. T., Takemae, H., Igarashi, I., Inoue, N., and
616 Yokoyama, N. (2016) Epidemiology of bovine hemoprotozoa parasites in cattle and water
617 buffalo in Vietnam. *J. Vet. Med. Sci.* **78**, 1361-1367
618 40. Bawm, S., Shimizu, K., Hirota, J., Tosa, Y., Htun, L. L., Maw, N. N., Thein, M., Kato, H., Sakurai, T.,
619 and Katakura, K. (2014) Molecular prevalence and genetic diversity of bovine Theileria orientalis
620 in Myanmar. *Parasitol Int* **63**, 640-645
621 41. Liu, G. H., Li, J. Y., and Zhu, X. Q. (2017) Characterization of the complete mitochondrial genome
622 of Setaria digitata (Nematoda: Setariidae) from China. *J. Helminthol.* **91**, 772-776
623 42. Shin, S. S., Cho, K. O., and Wee, S. H. (2002) Ocular infection of cattle with Setaria digitata. *J. Vet.*
624 *Med. Sci.* **64**, 7-10
625 43. Mans, B. J., Pienaar, R., and Latif, A. A. (2015) A review of Theileria diagnostics and
626 epidemiology. *Int J Parasitol Parasites Wildl* **4**, 104-118
627 44. Liu, A., Guan, G., Liu, Z., Liu, J., Leblanc, N., Li, Y., Gao, J., Ma, M., Niu, Q., Ren, Q., Bai, Q., Yin, H.,
628 and Luo, J. (2010) Detecting and differentiating Theileria sergenti and Theileria sinensis in cattle
629 and yaks by PCR based on major piroplasm surface protein (MPSP). *Exp. Parasitol.* **126**, 476-481
630 45. Cirimotich, C. M., Dong, Y., Garver, L. S., Sim, S., and Dimopoulos, G. (2010) Mosquito immune
631 defenses against Plasmodium infection. *Dev. Comp. Immunol.* **34**, 387-395
632 46. Bian, G., Joshi, D., Dong, Y., Lu, P., Zhou, G., Pan, X., Xu, Y., Dimopoulos, G., and Xi, Z. (2013)
633 Wolbachia invades Anopheles stephensi populations and induces refractoriness to Plasmodium
634 infection. *Science* **340**, 748-751
635 47. Meister, S., Kanzok, S. M., Zheng, X. L., Luna, C., Li, T. R., Hoa, N. T., Clayton, J. R., White, K. P.,
636 Kafatos, F. C., Christophides, G. K., and Zheng, L. (2005) Immune signaling pathways regulating
637 bacterial and malaria parasite infection of the mosquito Anopheles gambiae. *Proc. Natl. Acad.*
638 *Sci. U. S. A.* **102**, 11420-11425
639 48. Kamau, E., Tolbert, L. S., Kortepeter, L., Pratt, M., Nyakoe, N., Muringo, L., Ogutu, B., Waitumbi,
640 J. N., and Ockenhouse, C. F. (2011) Development of a highly sensitive genus-specific quantitative
641 reverse transcriptase real-time PCR assay for detection and quantitation of plasmodium by
642 amplifying RNA and DNA of the 18S rRNA genes. *J. Clin. Microbiol.* **49**, 2946-2953
643 49. Adams, M., Joshi, S. N., Mbambo, G., Mu, A. Z., Roemmich, S. M., Shrestha, B., Strauss, K. A.,
644 Johnson, N. E., Oo, K. Z., Hlaing, T. M., Han, Z. Y., Han, K. T., Thura, S., Richards, A. K., Huang, F.,
645 Nyunt, M. M., and Plowe, C. V. (2015) An ultrasensitive reverse transcription polymerase chain
646 reaction assay to detect asymptomatic low-density Plasmodium falciparum and Plasmodium
647 vivax infections in small volume blood samples. *Malar J* **14**, 520
648 50. Habibi, G. (2012) Phylogenetic Analysis of Theileria annulata Infected Cell Line S15 Iran Vaccine
649 Strain. *Iran J Parasitol* **7**, 73-81

650

651

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

652 **Tables**

653 **Table 1. Sample summary.**

Origin	Number	Genera	Nucleic acids analyzed
Cambodia	265 individuals	*Anopheles*	DNA
Guinea and Mali	665 individuals	*Anopheles*	DNA/RNA
Maryland, USA	25 pools	*Aedes, Anopheles, Culex, Culiseta, Ochlerotatus, Psorophora, Uranotaenia*	DNA/RNA

654

655

656 **Table 2. Primers used.**

Taxon targeted	Name	Forward Primer	Reverse Primer	Amplicon Length
Apicomplexa (primers A) (23)	Apicomp18S_365-613	GACCTATCAGCTTTCGACGG	CCCTCCAATTGWTACTCTGGR	244–248
Apicomplexa (primers B) (23)	Plasmo18S_883-1126	TGYGTTTGAATACTAYAGCATGG	TCTGATCGTCTTCACTCCCTT	228–244
Apicomplexa (primers C) (23)	Eimeriorina18S_302-730	TTGGMCTACCGTGGCARTGA	TCAAGGCAAHWGCCTGCTT	420–470
Kinetoplastida (23)	Kineto_18S4	AAATTAAACCGCACGCTCCA	GCAAACGATGACACCCATGA	250–300
Microsporidia (23)	Microsporidia_18S	BCAGGTTGATTCTGCCTGACR	ACCAGWCTTGCCCTCCARTT	370–440
Parabasalia (23)	Parab18S_288-654	TAGGCTATCACGGGTAACGG	GCGTCCTGATTTGTTCACAG	326–364
Platyhelminthes (23)	Platy_18S3	CAATTGGAGGGCAAGTCTGG	TGCTTTCGCWKTAGTTTGTCTG	350–550
Nematode (primers A) (23)	Spirur18S_1435-1858	CACCCGTGAGGATTGACAG	CGATCACGGAGGATTTTCAA	320–335
Nematode (primers B) (23)	Spirurida_18S2	CGTCATTGCTGCGGTTAAAA	CCGTCCTTCGAACCTCTGAC	380–410
Nematode (primers C) (23)	Tricho_18S2	AGTGGAGCATGCGGCTTAAT	TGCAATTCCCTRTCCCAGTC	380–440
Flavivirus (25)	FlaviallS_FlaviallAS2	TACAACATGATGGGGAARAGAGARAA	GTGTCCCAGCCNGCKGTGTCATCWGC	263–266
Theileria (50)	Theileria	GGCGGCGTTTATTAGACC	TCAATTCCTTTAAGTTTCAGCC	894–1045
Filarial (long range PCR)	Filarial	TTCGTCGTGAGACAGAGCGG	AGGCCATTGACGGATGGTTTGTAC	3500

657

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

658 **Table 3. Genera amplified from each group of samples.**

Target	Genus	Maryland Pools	Africa	Cambodia	Identity
Apicomplexa		**21**	**31**	**17**	
	Ascogregarina	2	0	0	100.00%
	Babesia	1	0	0	100.00%
	Cryptosporidium	0	1	0	92.08%
	Hepatocystis	1	0	0	85.45%
	Hepatocystis/Plasmodium	4	0	0	93.50%
	Paraschneideria	1	0	0	91.63%
	Plasmodium	20	8	2	82.02%-100.00%
	Theileria	7	22	15	99.10%-100.00%
Kinetoplastida		**22**	**0**	**1**	
	Angomonas/Crithidia	5	0	0	99.56%-100.00%
	Blastocrithidia	1	0	0	98.27%
	Blastocrithidia/Crithidia/Leptomonas/Wallaceina	4	0	0	100.00%
	Blechomonas	2	0	0	90.27%-91.44%
	Crithidia	15	0	0	99.13%-100.00%
	Crithidia/Trypanosomatidae	6	0	0	100.00%
	Paratrypanosoma	2	0	0	94.86%
	Strigomonas	8	0	1	96.89%-100.00%
	Trypanosoma	22	0	0	82.70%-100.00%
	Zelonia	1	0	0	98.71%
Flaviviridae		**7**	**4**	**0**	
	Aedes flavivirus/Calbertado virus	1	0	0	74.26%
	Anopheles flavivirus	0	3	0	87.20%-99.06%
	Calbertado virus	7	0	0	71.09%-74.26%
	Culex flavivirus	0	1	0	99.06%
	Nienokoue virus	1	0	0	72.99%
Microsporidia		**22**	**72**	**13**	
	Agglomerata	0	0	3	96.23%-96.52%
	Amblyospora	2	0	0	94.41%
	Andreanna	1	0	0	76.01%
	Culicospora	1	6	0	99.70%-100.00%
	Encephalitozoon	0	6	0	100.00%
	Hazardia	0	1	0	97.08%
	Microsporidium	7	34	0	88.29%-97.31%
	Parathelohania	5	28	10	91.27%-100.00%
	Senoma	0	0	3	94.20%
	Takaokaspora	14	0	0	94.61%-94.94%
Nematoda		**19**	**33**	**10**	
	Abursanema	0	1	0	95.34%
	Acanthocheilonema	3	13	0	100.00%
	Aproctella	2	0	0	100.00%
	Aproctella/Setaria	1	0	3	99.64%
	Auanema	0	1	0	98.21%
	Breinlia/Dipetalonema/Dirofilaria/ Onchocerca/Pelecitus/Setaria	0	0	6	99.64%
	Caenorhabditis	0	1	0	100.00%
	Cercopithifilaria	2	0	0	100.00%
	Choriorhabditis	1	0	0	99.71%
	Dipetalonema	4	3	0	98.68%-99.20%
	Dipetalonema/Loa	0	7	0	100.00%

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Dipetalonema/Loa/Loxodontofilaria/Yatesia	1	0	0	99.64%
	Dipetalonema/Loa/Madathamugadia	0	2	0	99.64%
	Dipetalonema/Loa/Onchocerca	0	2	0	99.64%
	Dipetalonema/Loa/Pelecitus	0	1	0	99.64%
	Dipetalonema/Loa/Setaria	2	0	6	99.64%
	Dirofilaria	1	0	0	100.00%
	Dirofilaria/Onchocercidae	1	0	0	99.44%
	Elaeophora/Setaria	1	0	0	99.44%
	Filarioidea	5	3	0	100.00%
	Loxodontofilaria	1	0	0	100.00%
	Loxodontofilaria/Setaria	0	0	5	99.64%
	Madathamugadia/Setaria	0	0	6	99.64%
	Onchocerca/Setaria	0	0	6	99.64%
	Setaria	13	13	10	98.94%-100.00%
	Trichuris	0	1	0	99.77%
	Yatesia	8	0	0	100.00%
Parabasalia		**0**	**4**	**1**	
	Tetratrichomonas	0	3	0	99.69%-100.00%
	Trichomitus	0	0	1	98.68%
	Trichomonas	0	1	0	99.34%
	Tritrichomonas	0	1	0	94.75%
Platyhelminthes		**3**	**4**	**3**	
	Haematoloechus	3	0	0	98.68%
	Pleurogenes/Pleurogenoides	0	0	1	94.93%
	Pleurogenoides	0	0	3	95.16%-95.37%
	Schistosoma	0	4	0	92.05%-100.00%

659

660 **Figure legends**

661 **Figure 1. Phylogenetic analysis of flavivirus sequences amplified from mosquitoes.** The neighbor-
662 joining tree shows the relationships between the flavivirus sequences amplified from mosquito pools
663 from Maryland (red circles) and from individual African mosquitoes (blue triangles). Phylogenetic tree
664 without compressed branch available in **Supplemental Figure 2**.

665 **Figure 2. Phylogenetic analysis of *Theileria* sequences amplified from Cambodian and African**
666 ***Anopheles* mosquitoes.** The neighbor-joining tree shows the relationships between the 18S rRNA
667 *Theileria* sequences amplified from samples positive by high-throughput sequences and those from
668 known *Theileria* species deposited in NCBI. Sequences amplified from Cambodian mosquitoes are
669 indicated in green circles, those amplified from African mosquitoes in red squares.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426319; this version posted January 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

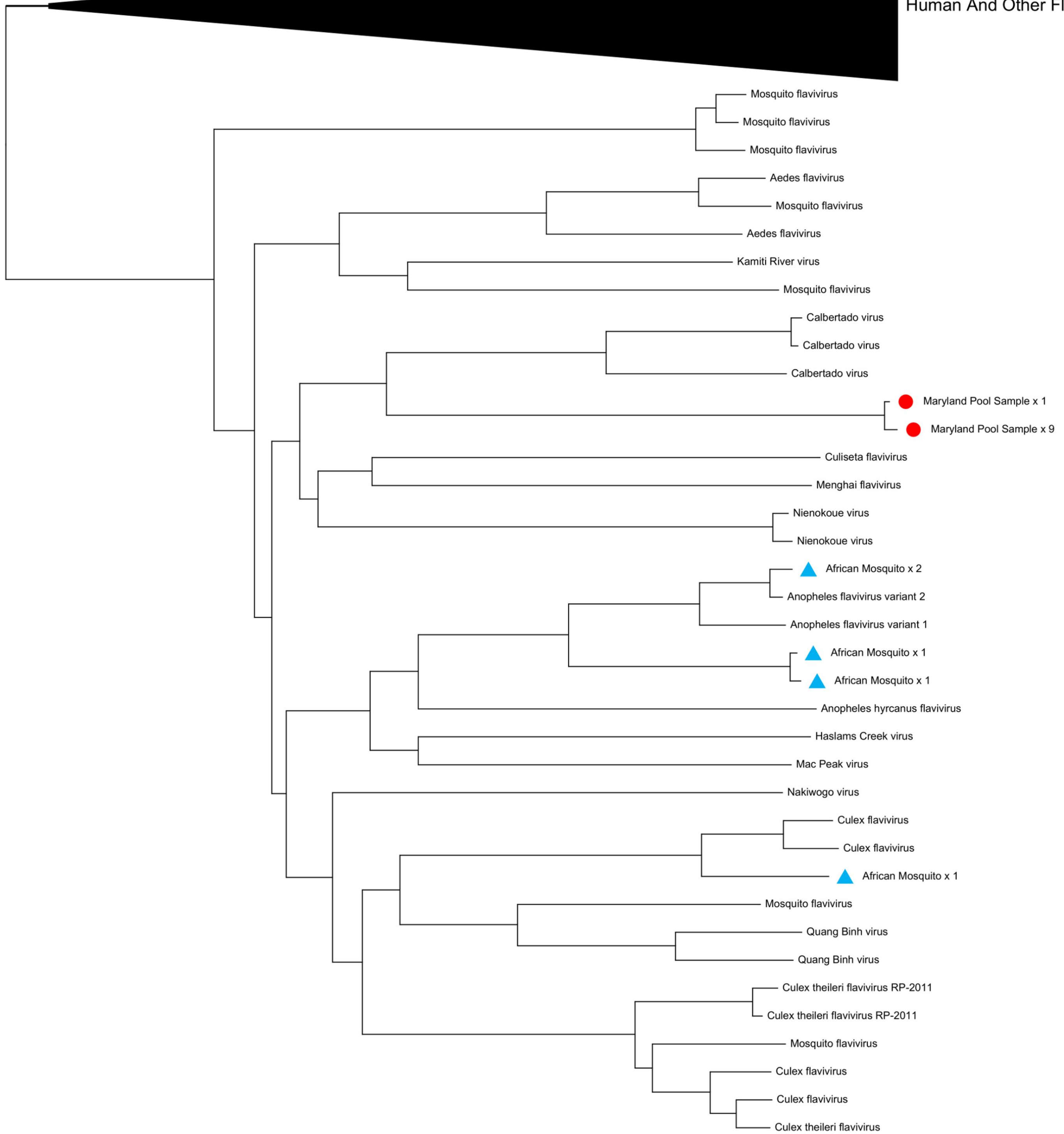
670 **Figure 3. Phylogenetic analysis of unknown filarial worm sequences amplified from Guinean
671 mosquitoes.** The neighbor-joining tree shows the relationships between annotated filarial worm
672 sequences and a 3.5 kb sequence amplified from two African mosquitoes (red squares) positive for
673 filarial worms and sequenced using PacBio chemistry.

674 **Figure 4. Comparison of the number of reads obtained for different taxa from matched DNA and cDNA
675 samples derived from Maryland mosquito pools.**

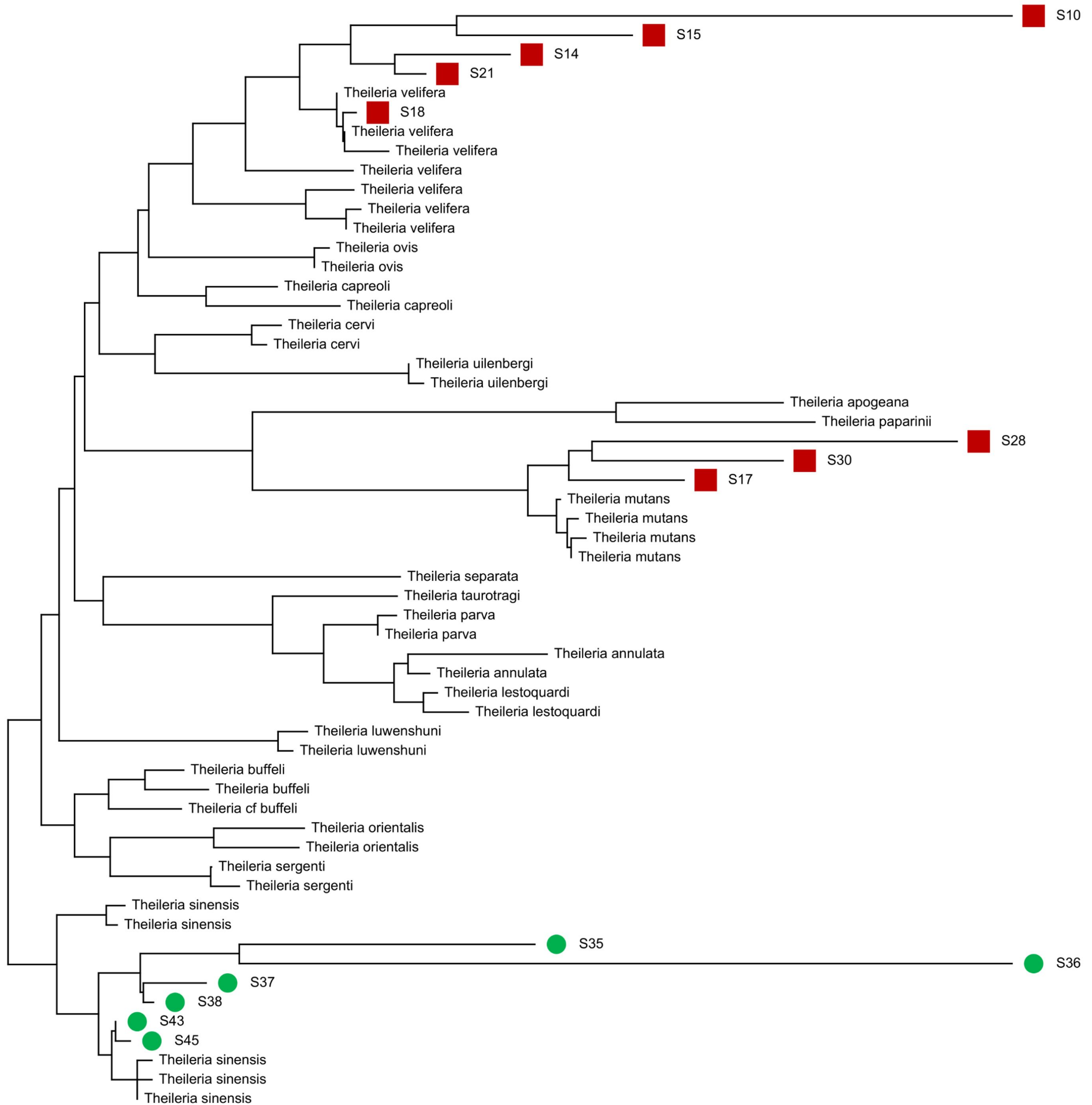
676 Each panel represents results from one primer set and each pair of points connected by a line shows the
677 number of reads matching a single species detected in both the DNA (left) and RNA (right) from the
678 same sample. For five primers (red asterisks), the RNA samples yield significantly more reads than the
679 matching DNA samples ($p < 0.05$, Bonferroni-corrected pairwise t-tests).

680

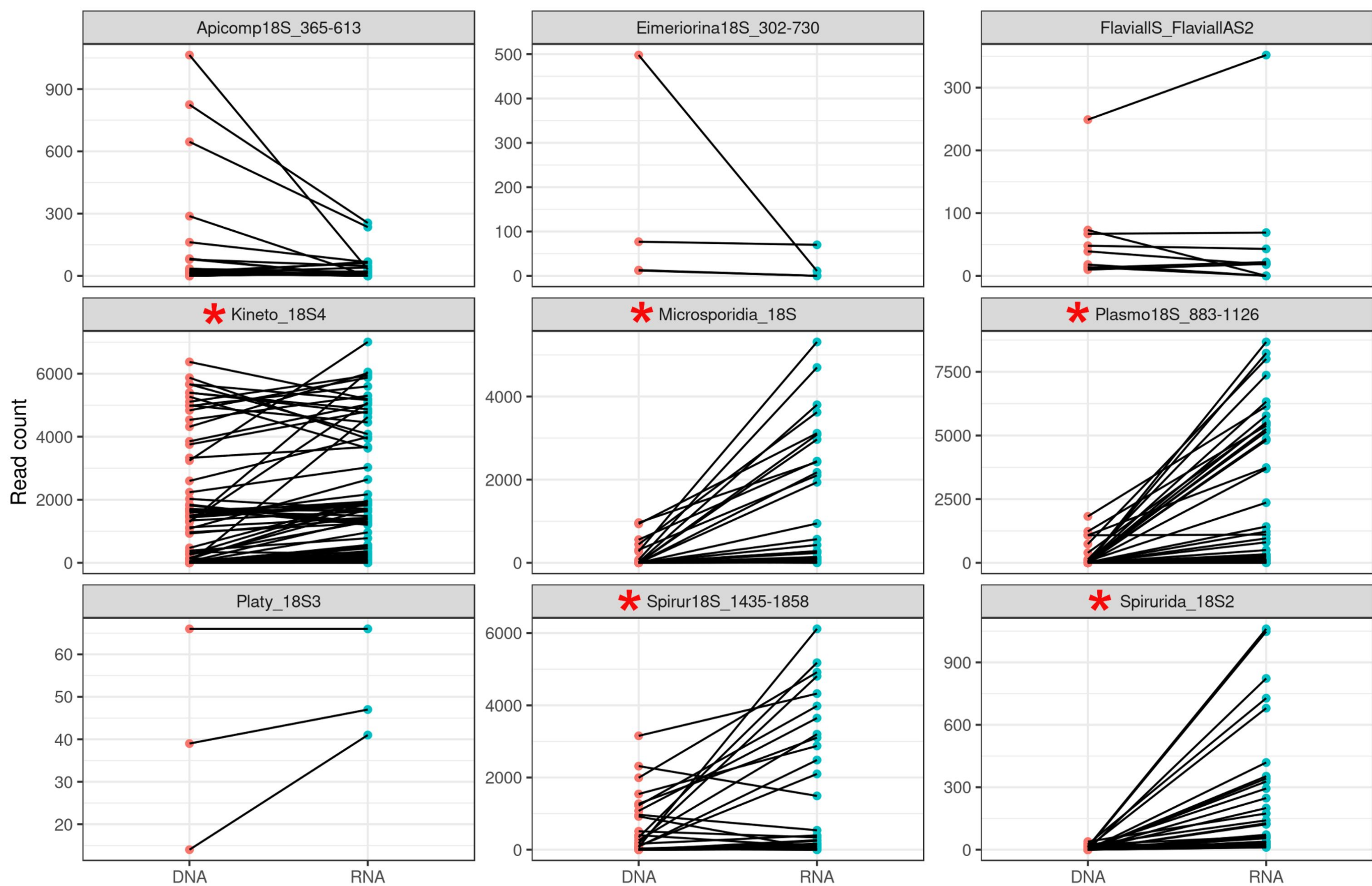
681



Human And Other Flaviviruses
Mosquito flavivirus
Mosquito flavivirus
Mosquito flavivirus
Aedes flavivirus
Mosquito flavivirus
Aedes flavivirus
Kamiti River virus
Mosquito flavivirus
Calbertado virus
Calbertado virus
Calbertado virus
Maryland Pool Sample x 1
Maryland Pool Sample x 9
Culiseta flavivirus
Menghai flavivirus
Nienokoue virus
Nienokoue virus
African Mosquito x 2
Anopheles flavivirus variant 2
Anopheles flavivirus variant 1
African Mosquito x 1
African Mosquito x 1
Anopheles hyrcanus flavivirus
Haslams Creek virus
Mac Peak virus
Nakiwogo virus
Culex flavivirus
Culex flavivirus
African Mosquito x 1
Mosquito flavivirus
Quang Binh virus
Quang Binh virus
Culex theileri flavivirus RP-2011
Culex theileri flavivirus RP-2011
Mosquito flavivirus
Culex flavivirus
Culex flavivirus
Culex theileri flavivirus
10


S10
S15
S14
S21
Theileria velifera
S18
Theileria velifera
Theileria velifera
Theileria velifera
Theileria velifera
Theileria velifera
Theileria velifera
Theileria ovis
Theileria ovis
Theileria capreoli
Theileria capreoli
Theileria cervi
Theileria cervi
Theileria uilenbergi
Theileria uilenbergi
Theileria apogeana
Theileria paparinii
S28
S30
S17
Theileria mutans
Theileria mutans
Theileria mutans
Theileria mutans
Theileria separata
Theileria taurotragi
Theileria parva
Theileria parva
Theileria annulata
Theileria annulata
Theileria lestoquardi
Theileria lestoquardi
Theileria luwenshuni
Theileria luwenshuni
Theileria buffeli
Theileria buffeli
Theileria cf buffeli
Theileria orientalis
Theileria orientalis
Theileria sergenti
Theileria sergenti
Theileria sinensis
Theileria sinensis
S35
S36
S37
S38
S43
S45
Theileria sinensis
Theileria sinensis
Theileria sinensis
10

Onchocerca ochengi
Onchocerca ochengi
Onchocerca volvulus
Onchocerca volvulus
Onchocerca flexuosa
Dirofilaria immitis complete
Dirofilaria repens
Dirofilaria repens
Loa loa
Wuchereria bancrofti
Wuchereria bancrofti
Brugia pahangi
Brugia malayi
Brugia timori
Acanthocheilonema viteae
Litomosoides sigmodontis
Mansonella perstans
Mosquito 1
Mosquito 2
Chandlerella quiscali
Setaria labiatopapillosa
Setaria digitata
Setaria digitata
Thelazia callipaeda
Thelazia callipaeda
Spirocerca lupi
Gongylonema pulchrum
Gongylonema pulchrum
Heliconema longissimum



Apicomp18S_365-613
Eimeriorina18S_302-730
FlaviallS_FlaviallAS2
Kineto_18S4
Microsporidia_18S
Plasmo18S_883-1126
Platy_18S3
Spirur18S_1435-1858
Spirurida_18S2
Read count
DNA
RNA