Investigation of the seasonal microbiome of *Anopheles coluzzii* in Mali

Benjamin J Krajacich [Corresp., 1] , Diana L Huestis [1] , Adama Dao [2] , Alpha S Yaro [2] , Moussa Diallo [2] , Asha Krishna [1] , Jiannong Xu [3] , Tovi Lehmann [Corresp. 1]

[1] Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland, United States of America

[2] Malaria Research and Training Center, Faculty of Medicine, Pharmacy, and Odontostomatology, University of Bamako, Bamako, Mali

[3] Biology Department, New Mexico State University, Las Cruces, New Mexico, United States of America

Corresponding Authors: Benjamin J Krajacich, Tovi Lehmann
Email address: bkrajacich@gmail.com, tlehmann@niaid.nih.gov

The poorly understood mechanisms of the seasonal maintenance of *Anopheles* spp. mosquitoes through the dry season in Africa remain a critical gap in our knowledge of *Plasmodium* disease transmission. While it is thought that adult mosquitoes remain in a dormant state throughout this seven-month dry season, the nature of this state remains unknown and has largely not been recapitulated in laboratory settings. To elucidate possible life history traits allowing for this phenotype, the spatiotemporal change in the microbiome of mosquitoes in the dry and wet seasons in Mali was analyzed by sequencing the 16S ribosome bacterial region in whole bodies of adult mosquitoes collected from two locations with varying water availability. These locations were a village near the Niger River with year-round water sources (N'Gabakoro, "Riparian"), and an area closer to the Sahara with highly seasonal breeding sites (Thierola Area, "Sahelian"). The 16S bacterial data consisted of 2057 unique sequence variants in 426 genera across 184 families. With these, we found several compositional differences that were seasonally and spatially linked. Counter to our initial hypothesis, there was a more pronounced seasonal difference in the bacterial microbiome in the Riparian than Sahelian area. These major seasonal shifts were in *Ralstonia, Sphingorhabdus,* and *Duganella* spp. bacteria that are usually soil and water-associated, indicating that these changes may be from bacteria acquired in the larval environment, rather than during adulthood. In the Sahelian dry season mosquitoes, there was a unique intracellular bacteria, *Anaplasma*, which likely was acquired through non-human blood feeding. Coupled with this finding, cytochrome B analysis showed a greater heterogeneity in host choice of *An. coluzzii* independent of season in the Thierola area compared to N'Gabakoro (77.5% vs. 94.6% human-origin blood meal, respectively), which may indicate a relaxation of anthropophily in some locations. This study highlights the diversity present in the bacterial composition of individual mosquitoes, characterizes

the spatial and seasonal differences in this composition, and indicates some possible qualitative biomarkers in areas of intense seasonal change.

1   Title: Interrogation of the Seasonal Microbiome of *Anopheles coluzzii* in Mali

2   Authors: Benjamin J. Krajacich*[1], Diana L. Huestis[1], Adama Dao[2], Alpha S. Yaro[2], Moussa

3   Diallo[2], Asha Krishna[1], Jiannong Xu[3], Tovi Lehmann*[1]

4   *Corresponding authors

5   [1]Laboratory of Malaria and Vector Research, NIAID, NIH, Rockville, Maryland, USA

6   [2]Malaria Research and Training Center (MRTC)/ Faculty of Medicine, Pharmacy and Odonto-

7   stomatology, Bamako, Mali

8   [3]Biology Department, New Mexico State University, Las Cruces, New Mexico, USA

9   **Abstract:**

10   The poorly understood mechanisms of the seasonal maintenance of *Anopheles* spp.
11   mosquitoes through the dry season in Africa remain a critical gap in our knowledge of
12   *Plasmodium* disease transmission. While it is thought that adult mosquitoes remain in a dormant
13   state throughout this seven-month dry season, the nature of this state remains unknown and has
14   largely not been recapitulated in laboratory settings. To elucidate possible life history traits
15   allowing for this phenotype, the spatiotemporal change in the microbiome of mosquitoes in the
16   dry and wet seasons in Mali was analyzed by sequencing the 16S ribosomal RNA gene tag in
17   whole bodies of adult mosquitoes collected from two locations with varying water availability.
18   These locations were a village near the Niger River with year-round water sources (N'Gabakoro,
19   "Riparian"), and an area closer to the Sahara with highly seasonal breeding sites (Thierola Area,
20   "Sahelian"). The 16S bacterial data consisted of 2057 unique sequence variants in 426 genera
21   across 184 families. With these, we found several compositional differences that were seasonally
22   and spatially linked. Counter to our initial hypothesis, there was a more pronounced seasonal
23   difference in the bacterial microbiome in the Riparian than Sahelian area. These major seasonal
24   shifts were in *Ralstonia, Sphingorhabdus,* and *Duganella* spp. bacteria that are usually soil and
25   water-associated, indicating that these changes may be from bacteria acquired in the larval
26   environment, rather than during adulthood. In the Sahelian dry season mosquitoes, there was a
27   unique intracellular bacteria, *Anaplasma*, which likely was acquired through non-human blood
28   feeding. Coupled with this finding, cytochrome B analysis showed a greater heterogeneity in
29   host choice of *An. coluzzii* independent of season in the Thierola area compared to N'Gabakoro
30   (77.5% vs. 94.6% human-origin blood meal, respectively), which may indicate a relaxation of
31   anthropophily in some locations. This study highlights the diversity present in the bacterial
32   composition of individual mosquitoes, characterizes the spatial and seasonal differences in this
33   composition, and indicates some possible qualitative biomarkers in areas of intense seasonal
34   change.
35
36
37

Introduction:

The mosquito microbiome still is a largely unknown variable in a range of processes that may contribute greatly to the ability of mosquitoes to spread disease. It has been noted that the microbiome can effect vector competence for arboviruses and parasites (Cirimotich et al., 2011; Cirimotich, Ramirez & Dimopoulos, 2011; Gendrin & Christophides, 2013; Bahia et al., 2014; Dennison, Jupatanakul & Dimopoulos, 2014; Jupatanakul, Sim & Dimopoulos, 2014; van Tol & Dimopoulos, 2016), and that this microbiome contains signatures which may link mosquito populations geographically (Buck et al., 2016). Recent work has shown there is some seasonal variation in primarily *An. gambiae* s.s. mosquito microbiota from the forest-savannah regions that maintain some breeding sites during the dry season (Akorli et al., 2016). However, there has been no work to date that has characterized the variability present in the whole-body mosquito microbiome in areas with more distinct seasonality such as the Sahel where no larval habitat could be found during the dry season, and there have been no studies of mosquito fungal or sub-species level bacterial microbiome composition in these areas. This is pertinent as in these highly seasonal areas mosquitoes of the *Anopheles gambiae* s.l. complex have evolved cryptic strategies to cope with the harshness of the dry season, with *An. coluzzii* (M form) mosquitoes favoring aestivation (remaining with limited activity) and *An. gambiae* s.s. (S form) migrating to areas where larval sites are available year round (Dao et al., 2014).

This exploratory study utilizes quantitative and qualitative measures to evaluate how the microbiome differs in an area in West Africa with highly seasonal water availability and mosquito abundance. We compare seasonal changes within and between locations, initially hypothesizing that we would find the greatest difference in mosquito microbiota found in Sahelian dry season due to this location having the clearest demarcation between physiological

61   states of reproductive vs. aestivating mosquitoes in the wet and dry seasons, respectively.

62   Additionally, we compared how laboratory mosquitoes compare to these field-caught specimens

63   in their microbial composition.

64   **Methods:**

65   *Mosquito collection and field sites:*

66   *An. coluzzii* species mosquitoes were collected via indoor aspiration in three locations,

67   Zanga (Latitude 13.688050°, Longitude -7.221029°), M'Piabougou (13.599830°, -7.192859°),

68   and N'Gabakoro (12.683870°, -7.840419°) in the Koulikoro region of western Mali from

69   September 2009 to August 2010 (Fig. 1). Mosquitoes from each location were noted for blood

70   fed status, had their thoraces punctured, and were added to 50 µl RNA*later* stabilization solution

71   (ThermoFisher Scientific, Waltham, MA, USA) for preservation. Fourteen additional laboratory

72   reared *An. coluzzii* were also preserved in groups of young (3 days old, Samples 55-57) and old

73   (14 days old, Samples 58-60) post emergence to compare microbiome development over time.

74   Mosquitoes were speciated to molecular form based on a direct PCR performed on two

75   legs with a standard protocol (Fanello, Santolamazza & della Torre, 2002). DNA was extracted

76   from mosquitoes with a DNeasy kit according to manufacturer instructions (Qiagen, Valencia,

77   CA, USA), and eluted in 50 µl of the provided elution buffer.

78   *Normalized Difference Vegetation Index (NDVI) analysis:*

79   To further quantify the differences in seasonality between our Sahelian and riparian

80   areas, we calculated the NDVI around the center point of each of the 3 field sampling locations

81   with a ~5 km square on the NDVI output from the Metop-AVHRR S10 ("ENDVI10", 1 km

82   resolution, daily measurements with best composite image per dekad available) global satellite

83   database (Accessed 5/1/2017, from: http://www.vito-

84   eodata.be/PDF/portal/Application.html#Home, Borgne, Legendre & Marsouin, 2007). The 35

85   measurements spanning the sampling collection period were downloaded for the African

86   subcontinent in Erdas Imagine (IMG) format and were analyzed using the 'raster' library in R

87   with a spatial grid cell size of 0.0025° defined with the 'sp' package (Pebesma & Bivand, 2005;

88   Bivand, Pebesma & Gomez-Rubio, 2013; Hijmans, 2016). NDVI data was plotted for four

89   representative time points using the 'rasterVis' package (Fig. 1A) and for each of the available

90   NDVI images during the sampling period using the 'ggplot2' and 'scales' packages (Fig. 1B)

91   (Wickham, 2009, 2016; Perpiñán & Hijmans, 2016). Shape files for Mali and the sampling

92   regions are acquired from the Database of Global Administrative Areas (GADM) database with

93   'raster'.

94   *454 Sequencing:*

95   Next-generation sequencing of the V1-V3 regions of 16S ribosomal RNA gene was

96   performed by MR DNA (Shallowater, TX, USA) on a Roche 454 sequencer using the forward

97   primer "27FMod"- 5'-AGRGTTTGATCMTGGCTCAG-3' and the reverse primer

98   "519Rmodbio" – 5'-GWATTACCGCGGCKGCTG -3'. Amplification and sequencing

99   conditions used are described in full elsewhere (Dowd et al., 2008; Pirmohamed et al., 2010).

100   Mosquitoes were sequenced either as individuals from 15ng of DNA, or from a pool of three

101   mosquitoes (5ng from each) with concentrations being determined by NanoDrop

102   (ThermoFisher).

103   *Processing Pipeline:*

104   The .fasta and .qual files with primers and barcodes removed are split using the

105   split.libraries.py command in QIIME v1.9.0 using the –d flag to record quality scores (Caporaso

106   et al., 2010). The output was then demultiplexed into individual fastq files in QIIME. FastQ files

107 are then imported into R version 3.4.0 using the RStudio IDE version 1.0.44 using the 'dada2'

108 package (RStudio Team, 2015; Callahan et al., 2016a; R Core Team, 2016). Processing using

109 this pipeline largely follows the Bioconductor workflow from Callahan *et al.* (Callahan et al.,

110 2016b). Briefly, data are trimmed/filtered with settings truncLen=325, truncQ=2, maxN=0, and

111 maxEE=2. Sequencing error rates are learned on a random sample of *n*=25, with suggested

112 parameters for 454 data of HOMOPOLYMER_GAP_PENALTY=-1 and BAND_SIZE=32.

113 Chimeras are removed, and taxa is assigned with a naïve Bayesian classifier algorithm over 100

114 bootstraps with a minimum bootstrap level of 50 (Wang et al., 2007), and taxonomy is assigned

115 using the 'silva_nr_v123_train_set' for 16S rRNA or the UNITE database for ITS formatted for

116 use in dada2 (Silva training set: https://zenodo.org/record/158958, UNITE database (General

117 FASTA Release): https://unite.ut.ee/repository.php) (Quast et al., 2013; Callahan, 2016). For 16S

118 this is based on a 325nt length amplicon and the ITS has a variable length amplicon with a

119 minimum of 50 nt length.

120   The sequences are aligned with the 'DECIPHER' package in R, output with the

121 'phangorn' package version 2.1.1 into FastTree 2 software for generation of a generalized time-

122 reversible (GTR) maximum-likelihood phylogenetic tree with rescaling of branch lengths and

123 computation of Gamma20-based likelihood (Price, Dehal & Arkin, 2010; Schliep, 2011; Wright,

124 2016). This tree is read back into R using the 'ape' package (Paradis, Claude & Strimmer, 2004),

125 and then all the data are combined with the 'phyloseq' package for data manipulation and

126 visualization (McMurdie & Holmes, 2013). After contaminant reads from eukaryotic sources

127 were removed the Shannon index of richness is calculated for each mosquito. Phylogenetic tree

128 visualization is performed using the package 'ggtree' (Yu et al., 2017).

129 *Differential abundance testing and clustering analysis:*

130   Seasonal differences in microbial abundance were analyzed with the 'DESeq2' package

131   on genus agglomerated data (Love, Anders & Hu-, 2016). This method creates a negative

132   binomial generalized linear model (GLM) to estimate maximum likelihood for each genus's

133   count based on $\log_2$ fold changes between two conditions (i.e. Dry vs. Wet at that location),

134   while accounting for size factors (differences in sequence depth between samples) and dispersion

135   (between sample variance) using Bayesian shrinkage (Love, Huber & Anders, 2014; Weiss et al.,

136   2017). We compared samples seasonally per location, and also all field vs. all laboratory

137   samples. The *p*-value from the Wald's test is then adjusted for a false discovery rate with the

138   Benjamini-Hochberg approach and local smoothed dispersion fit (Benjamini & Hochberg, 1995).

139   Shrunken (adjusted if they have high dispersion or low sampling coverage) $\log_2$ fold changes

140   between seasons at each location are generated by contrasting "PermDry" to "PermWet";

141   "SahelDry" to "SahelWet"; and "Field" to "Lab" for the samples from the riparian, Sahelian, and

142   field sites, respectively.

143   Hierarchical multiple testing, a procedure by which you incorporate the innate structure

144   in the data (in this case the phylogenetic hierarchy) to adjust your false discovery rate and

145   improve power in comparing count data (Yekutieli, 2008), was implemented in R with the

146   'structssi' package (Sankaran & Holmes, 2014). We defer the majority of details of this approach

147   to the parent literature (Yekutieli, 2008), and to its implementation literature for microbiome data

148   (Sankaran & Holmes, 2014; Callahan et al., 2016b). Briefly, we control the false discovery rate

149   by organizing the hypotheses based on the phylogenetic tree of the data, and only testing for

150   differential abundance if the parent hypothesis is found to be significant between groups at a

151   coarse level of discrimination (a 'hFDR' rate of 0.75) (Callahan et al., 2016b). This analysis is

152   performed on 16S sequence variant count data that has been variance stabilizing transformed by

153   'DESeq2' and shifted so all values are positive. Genera with greater than 1.5 shrunken-log$_2$fold

154   change that are found to be significant at an adjusted *p*-value of less than 0.05 are presented.

155   Ordination plots were generated from prevalence filtered (must not be sample singletons)

156   and rarefied data (rarefied to 500 reads) using 'phyloseq' and 'vegan' with principal coordinates

157   analysis on Bray-Curtis similarity distances (Beals, 1984; Oksanen et al., 2017). Analysis of

158   centroid (center point) differences between sample groups (Season and Location) were compared

159   using a pairwise permutation multivariate analysis of variance test (PERMANOVA) with 9999

160   permutations with Benjamini-Hochberg false discovery rate correction using the

161   'RVAideMemoire' package (Hervé, 2017).

162   *Presence/absence analysis for seasonally indicative genera:*

163   In addition to analysis of the variance in bacterial taxon abundance described above, we

164   performed two analyses based on the presence/absence of bacterial taxa in mosquitoes,

165   regardless of the abundance. In the first, we performed random forest supervised classification

166   analysis on genus-level data using the 'caret' package in R (Breiman, 2001; Liaw & Wiener,

167   2002; Kuhn et al., 2016). The proximity metric calculated during the random forest model

168   generation was utilized to generate a multi-dimensional scaling plot (MDS) to evaluate the

169   degree of difference and clustering between and within sample groups via PERMANOVA.

170   Prevalence for each of the sample groups for the top 10 most important genera in the random

171   forest classification model are presented.

172   Additionally, for field samples we enumerated taxa unique to the Sahelian dry season

173   sample, found their relative frequency (in terms of mosquito hosts), and evaluated how unique

174   taxa are to that sample group ("private" - i.e. found only in that sample group) in SAS 9.4 (SAS

175   Institute Inc., Cary, NC, USA). Using "informative" taxa (excluding rare and ubiquitous taxa),

176   we sequentially tested if the frequency (prevalence) of any informative taxa differed between the

177   dry and wet seasons in the Sahel, followed by testing that the Sahelian dry season differs from

178   riparian dry season, and that it also differed from the riparian wet season in a consistent

179   direction. These tests employed the sequential Fisher's exact tests (with the last two being one-

180   way tests), only passing genera significant at $p < 0.1$ forwards to the next test. The final results

181   only report those genera whose prevalence difference is in the same direction (i.e. higher in

182   Sahel Dry compared to all other conditions).

183   *Cytochrome B bloodmeal analysis:*

184   The mammalian host source of each mosquito and pool per location was analyzed via a

185   size-discriminate multiplexed cytochrome B PCR developed by Kent and Norris (Kent & Norris,

186   2005). This PCR allows for the identification of bloodmeals from pigs, humans, goat, dog and

187   cows.

188   **Results:**

189   Unlike the Sahelian villages, the proximity of N'Gabakoro to the Niger allows mosquito

190   breeding year-round. To further assess the seasonal differential in aridity between the Sahelian

191   and riparian villages, we analyzed the normalized difference vegetation index (NDVI, Fig. 1.).

192   Due to its proximity to the Niger river, we found that there is a desiccation lag-period in which

193   the vegetation index does not drop at the same rate in N'Gabakoro as it does in the two villages

194   that have only rainfall as a water source (Fig. 1A top right, Fig. 1B). Additionally, while the

195   vegetation minima are reached roughly at the same time in the dry season between sites (late

196   March to early April), the overall NDVI remains higher in N'Gabakoro during the transition

197   periods to and from this dry season low point.

198   *454 Sequencing:*

199   Of the 58 samples sent for 454 sequencing, 53 returned sequence for 16S and 26 returned

200   sequence for ITS. For 16S sequencing 2643 sequence variants (SVs) were returned though a

201   range of these variants aligned to likely contaminating eukaryotic reads in the SILVA database

202   and were filtered, leaving 2057 sequence variants with 426 genera across 184 families. Many of

203   these reads when searched against the 'nr' database using BLAST had hits to *An. gambiae* strain

204   PEST sequence (Altschul et al., 1990). This could indicate some cross-reactivity of the primers

205   with other ribosomal sequence, or mis-annotated reference sequence. The fungal internal

206   transcribed spacer (ITS) region sequencing had a more limited success in amplification in terms

207   of both average read counts per successful sample (755 vs. 4554 for ITS and 16S, respectively),

208   and in the amount of coverage across locations/seasons. ITS failed to amplify from any samples

209   in N'Gabakoro or the laboratory samples, and 5 samples only had reads that align in BLAST nr

210   to *Anopheles* spp. ITS region sequence with no representation in the fungal UNITE database.

211   Due to this, no comparisons were made between season or location on the fungal microbiome.

212   All *Anopheles* ITS reads (63332/79195 total) were filtered. After filtering, the most abundant

213   fungal genera were *Aspergillus* (20.7% of total reads without *Anopheles*) with presence in 12/21

214   samples with fungal ITS reads followed by *Malassezia, Cladosporium,* and *Phoma* (15/21,

215   13/21, 9/21 of samples, respectively, Fig. S1).

216   *Characteristics of bacterial communities between sampling locations:*

217   The most abundant genera across field-caught mosquitoes was *Ralstonia* (25.1% of all

218   field reads, in 42/47 samples), and the most prevalent genera was *Propionibacterium* (46/47

219   samples) (Fig. 2). Laboratory samples were largely dominated by reads to *Asaia* (62.9% of all

220   laboratory reads), with this genus being the most prevalent taxa in 5/6 samples. There was no

221   significant difference in the Shannon diversity between groups of field samples via Kruskal-

222   Wallis chi-squared with Dunn's multiple comparisons test, though the mean diversity of the

223   laboratory samples was significantly lower than all field groups (Fig. S2A-C). While overall the

224   dry season had the highest diversity, this difference was not significantly higher than the wet

225   season overall (dry: 2.54, wet: 2.03, *p*=0.087, Fig. S2B). The mean number of sequence variants

226   per sample varied from 42.5-65.13 in the field to 20 in the laboratory, with dry season samples

227   from both field locations being significantly different than laboratory samples (Fig. S3A,

228   *p*=0.0117, 0.0066 for Thierola Area and N'Gabakoro, respectively). The mean number of genera

229   were also higher in the field than in the laboratory (27.36-42.13 in field, 14.83 in lab, Fig. S3B),

230   again with the dry seasons being significantly higher than the laboratory samples (*p*= 0.0215,

231   0.0119 for Thierola and N'Gabakoro, respectively). The Pearson correlation between the overall

232   genus abundance and frequency (across mosquitoes) was moderate (r=0.42 and r=0.70, P<0.001,

233   n=424 for native and log-transformed values), indicating that the genera more frequent across

234   mosquitoes were more abundant (sequence reads). However, the majority of sequence variants

235   and genera were sample singletons (91.9% and 43.7%, respectively), however the majority of

236   reads were found in non-singletons (68.0% of sequence variant reads, 98.8% of genera-level

237   reads) (Figs. S4A and S4B). This would indicate there remains a high degree of heterogeneity

238   between samples, though dominant species have some conservation between groups. Finally,

239   there were higher mean amounts of group specific (private) sequence variants within the Sahel

240   dry season (53.6 compared to 35.1, 47.5, 31.2, and 12.7 for Sahel Dry, Sahel Wet, Riparian Dry,

241   Riparian Wet, and laboratory samples, respectively), though this difference was only

242   significantly different between the laboratory samples and the two dry seasons (*p* = 0.0173 and

243   0.0048 for Riparian dry and Sahel dry via Dunn's test, respectively).

244   *Differential abundance testing of seasonality:*

245   We tested differences in genera abundance between seasons at each location, and

246   between field and laboratory samples. In Sahelian locations near Thierola, five bacterial species

247   were found to be abundant at an un-adjusted p-value of < 0.05 with *Anaplasma*, *Bacillus*,

248   *Arthrobacter* spp. being more abundant in the dry season individual samples (Fig. S5A, $\log_2$ fold-

249   change: 4.79, 3.89, and 3.01), and *Acinetobacter* and *Brevundimonas* spp. being less prevalent in

250   the dry season samples (Fig. S5A, $\log_2$ fold-change: -3.48 and -3.11). In riparian site

251   (N'Gabakoro), we found that reads of *Ralstonia* spp. was significantly reduced in mosquitoes

252   (Fig. S5B, $\log_2$ fold-change: -5.10, Benjamini-Hochberg adjusted p-value: $3.33e^{-04}$). As with the

253   Sahelian group, the Riparian dry season had some bacterial species that did not pass the false

254   discovery rate adjusted p-value cutoff (*Ideonella* and *Enterobacter*). Comparing all field to all

255   laboratory samples found six genera to be differentially expressed at p-adj < 0.05, with the

256   laboratory samples being biased towards *Asaia* genera, and having the presence of

257   *Gluconobacter* bacteria that were absent from all field samples (Fig. S5C).

258   In effort to improve discriminative power, additional analysis was performed using

259   hierarchical multiple testing on sequence variants to determine which SVs were differentially

260   abundant between seasons and locations. For the Thierola area, only one *Anaplasma* sequence

261   variants was found to be differentially expressed in the dry season compared to wet ($\log_2$ fold-

262   change: 3.50, p-adj: 0.027, Fig. 3A). Seven sequence variants were found to be differentially

263   expressed between seasons in the permanent water location, with one more abundant in the wet

264   season (Fig. 3B, *Ralstonia*, $\log_2$ fold-change: -4.91, p-adj: $1.4e^{-03}$), and six more abundant in the

265   dry (*Sphingorhabdus,* two *Duganella* SVs, *Janthinobacterium, Xenophilus, Pseudomonas,* and a

266   *Cyanobacteria*; $\log_2$ fold-change 1.63-3.94, p-adj: $3.19e^{-02}$ to $3.91e^{-05}$). In the laboratory versus

267   field comparison, six sequence variants were found to be significant (Fig. 3C, *Asaia,*

268   *Elizabethkingia,* two *Gluconobacter* SVs, *Alcaligenes,* and *Ralstonia*, $\log_2$ fold-change 2.97-

269   4.11, *p*-adj: $1.4e^{-03}$ to $5.61e^{-15}$).

270   *Seasonal and temporal discrimination using ordination:*

271   After rarefaction to a depth of 500 reads and filtering of sequence variants present in only

272   a single sample, 164 sequence variants and 48 samples were used for ordination analysis.

273   Ordination of the samples via principal coordinates analysis (PCoA) on the Bray-Curtis

274   dissimilarity distance showed clustering of the five seasonal and location based groupings, with

275   differing levels of between-group spread (Fig. 4A). Testing of the group centroid locations via a

276   pairwise PERMANOVA found that the group composition of laboratory and 'permdry' samples

277   differed from all other groups, while 'saheldry' differed from 'permwet' and 'sahelwet' at an

278   adjusted *p*-value < 0.10 (Fig. 4B).

279   *Qualitative presence/absence analysis:*

280   As abundance testing found limited seasonal differences for each location, we also

281   investigated whether supervised learning approaches could discriminate dry vs. wet season

282   samples based on the differences in presence/absence of bacteria between groups. We again

283   found that the seasonal samples from the permanent water clustered separately more strongly

284   than the Sahelian locations into two populations (Fig. S6A), though all groups other than the wet

285   season locations were found to have significantly different center points via PERMANOVA

286   (Fig. S6C).

287   To further compare the nature of the uniqueness present in the microbiome of possibly

288   aestivating mosquitoes in the Sahelian dry season, we employed sequential testing between

289   groups. Firstly, we tested for genera that showed significant differences between the dry and wet

290   seasons in the Sahel. Secondly, we tested for differences between the Sahelian and Riparian dry

291   season, within the subset of genera identified earlier. Thirdly, we tested for differences between

292   the Sahelian dry and Riparian wet season within the subset that passed the first tests. In each

293   step, we employed exact tests that accommodated sample size at the individual test level (the

294   second and third tests were one-sided, as dictated by the direction of the difference in the first

295   test. see methods). A total of 14 genera exhibited difference between the Sahelian dry and wet

296   season at P<0.1 (with highest significance for *Bacillus, Anaplasma*, and *Microvirga* (P<0.00004,

297   <0.0026, 0.007, respectively). In the subsequent tests, only *Anaplasma* which was exclusively

298   present in the Sahelian dry season (54% vs. 0%, *p*=0.009), *Bacillus* (90.9% vs. 7-38%, *p*=0.020),

299   *Intestinibacter* (45.4% vs. 0-7%, *p* = 0.022) and *Microvirga* (63.6% vs. 7-14%, *p*=0.029) were

300   considered putatively "characteristic" of the Sahelian dry season.

301   *Comparison of Sequence Variants (SV) against Operational Taxonomic Units (OTU):*

302   With the DADA2 denoising algorithm, sequencing errors can be parsed from true

303   sequence variants in the 16S rRNA sequenced region. For interpretability in display, the 16S SVs

304   and OTUs for the top 19 most prevalent genera were placed on a phylogenetic tree to

305   demonstrate the loss in information from clustering sequences with >97% nucleic acid identity

306   compared to maintaining variants with as little as one nucleotide difference (Figs. S7A and S7B).

307   There were 35.8% more SVs than OTUs across all genera (2057 and 1515, respectively), with

308   79.7% more SVs than OTUs in the top 19 most abundant genera.

309   *Cytochrome B analysis of blood feeding host preference:*

310   The *Anaplasma* spp. reads found to be more prevalent during the dry season were blasted

311   against the nr database were found to align to *Anaplasma ovis* (99-100% identity based on

312   strain)*, a pathogen of goats, sheep, and wild ruminants (de la Fuente et al., 2007). To define the

313   source of these reads we performed polymerase chain reaction to determine bloodmeal origin

314   and the rate of anthropophily seasonally and by location. We found that areas of seasonal water

315   availability had a slight decrease in the degree of anthropophily overall (Fig. 5 left side, 77.5% to

316   94.6%, comparing "Human" to all others, p=0.0490 with Two-tailed Fisher's exact test).

317   Additionally, the Thierola area had a statistically significantly higher proportion of zoophily in

318   the dry season compared to the dry season in N'Gabakoro (28.6% mixed or non-human blood-

319   feeding compared to 0%, p=0.0318).

320   **Discussion:**

321   The mosquito microbiome varies considerably throughout its development (Wang et al.,

322   2011; Coon et al., 2014), and has been previously shown to have distinct bacterial characteristics

323   due to its location and seasonal environment (Akorli et al., 2016; Buck et al., 2016). In this

324   study, we analyzed how the microbiome changes in *An. coluzzii* collected from areas with

325   differing degrees of seasonality in Mali. This variation in seasonality in our sampling locations

326   was due to several factors. The first is that due to being lower in latitude, the areas around

327   N'Gabakoro do not dry out as rapidly as the more northerly areas neared to the Sahel.

328   Additionally, the presence of the Niger River provides the possibility for year round

329   development of larvae in the Southern area that is likely not possible in the North. Our main

330   hypothesis was that large seasonal differences between Sahelian dry and wet season will be

331   present due to the unique physiological state of Sahelian *An. coluzzii* mosquitoes during

332   aestivation (Huestis et al., 2012; Yaro et al., 2012; Dao et al., 2014; Huestis & Lehmann, 2014).

333   If confirmed, aestivation-specific microbiome taxa might be used as a predictor of this state in

334   populations where both aestivators and reproductive adults coexist.

335   Counter to our initial hypothesis that dry season Sahelian mosquitoes would have the

336   most distinct microbiome composition, we found that dry season mosquitoes from areas with

337 year-round breeding (N'Gabakoro) were, overall, the most different among our mosquito groups

338 (below). The Sahelian dry season did exhibit some unique characteristics including high

339 microbial diversity with highest number of private sequence variants, and had slight differences

340 in overall microbial composition (Figs. 3), though these differences were not as pronounced as

341 those between seasons in permanent water locations. Additionally, several bacterial genera were

342 significantly elevated in the Sahelian DS including *Anaplasma* (exclusively present in the

343 Sahelian DS), *Bacillus* (90.9% in Sahel dry vs. 7-37.5% in others), *Microvirga* (63.6% vs. 7-

344 14%), and *Intestinibacter* (45.4% vs. 0-7%). We also note that three of these genera (*Anaplasma,*

345 *Bacillus*, and *Microvirga*) were also found to be important to the random forest classification

346 model (Figs. S6B and S6D). Only *Anaplasma* was found to be significantly different based on

347 read count abundance ($\log_2$ fold-change: 3.50, *p*-adj: 0.027, Fig. 3). The putative route of

348 acquisition of *Anaplasma* is described in depth below, but the most likely route of acquisition of

349 *Bacillus, Microvirga*, and *Intestinibacter* would be through the larval environment or possibly

350 plant feeding due to their soil/water association (Table S1). As no private genera were present in

351 all Sahelian dry season mosquitoes, if bacterial biomarkers of aestivation exist, it may indicate

352 not all mosquitoes in this period are in this state. Independent studies are necessary to test

353 whether these putative differences are indeed "characteristic" of the Sahelian dry season and has

354 relevance to aestivation.

355 The riparian dry season sample was characterized by a reduction of *Ralstonia* species

356 reads, with an increase in *Duganella, Janthinobacterium,* and *Sphingomonas spp.* reads as

357 determined by hierarchical multiple testing, DESeq2, and ordination (Figs 3, 4, and Fig. S5B).

358 The most abundant genera was *Ralstonia* that was present in most wild-caught samples (present

359 in 40/47, dominant taxa in 24/47), and had little representation in laboratory mosquitoes (1/6

360   samples with 10 reads). This genus has been seen previously in wild-caught *Anopheles* from

361   Cameroon (Boissière et al., 2012) and *Aedes aegypti* (Apte-Deshpande et al., 2012), and is

362   believed to be largely soil and water associated (Gilligan et al., 2003; Ryan, Pembroke & Adley,

363   2006). Additionally, we found that the majority of 16S sequence variants were unique to an

364   individual sample (91.9%, Fig. S4), indicating that there is a high degree of heterogeneity

365   between mosquitoes. However these sequence variants were relatively low in abundance, only

366   accounting for 32.0% of reads. These sequence variants also demonstrate the additional sub-

367   species level heterogeneity that the more commonly used operational taxonomic units (OTUs)

368   may miss (Fig. S6).

369   As it has been previously reported in *Aedes aegypti* that most of the adult gut flora is

370   acquired trans-stadially from larval stages to adults (Coon et al., 2014), this may help explain

371   why there were limited seasonal differences in mosquitoes in the northern Sahel. Due to the

372   prolonged dry season, there are no known larval sites available during this period (Dao et al.,

373   2014). Thus, the mosquitoes present would have been larvae in the same rainfall-linked transient

374   water sources that the wet season mosquitoes would have, and any changes present in the

375   microbiome would be due to incorporation of new flora from environmental conditions, or

376   blood/sugar sources. This homogeneity in larval environments is contrasted in the available

377   larval habitat near N'Gabakoro that changes broadly throughout the year, from fresh rain puddles

378   to ground and river water in rock and other pools near the Niger River that develop as flooded

379   areas recede (Edillo et al., 2002; Coulibaly et al., 2007), or from standing water present in the

380   more urban area of Bamako nearby (though these pools are likely unsuitable for *Anopheles* spp.

381   growth). This hypothesis may explain why the dry season Sahelian samples had the most

382   pronounced difference in genera that are likely acquired through blood feeding, and the Riparian

areas had the largest differences in bacteria that are predominantly soil and water associated (Table S1). This may also explain why the previous analysis of seasonality in a wetter climate of Ghana showed limited seasonal differences in *Anopheles coluzzii* (Akorli et al., 2016). Another difference between the studies is related to the Ghanaian study use of day 1 post emergence mosquitoes that were collected as larvae in differing seasonal water sources and were non-bloodfed (Akorli et al., 2016). This may then limit effects of the adult environment upon the microbiome. Future studies should compare water samples from collection locations to further refine what is the main route of acquisition of these genera, and whether this shift is consistent between years.

The overall uniqueness of the laboratory samples was also found to be pronounced in our samples contrary to other studies that find high degree of similarity between field and laboratory samples (Wang et al., 2011; David et al., 2016). We found all bacterial community compositions from each field location/season different from the laboratory samples via pairwise PERMANOVA (Figs 4, S6). At the genus level, the *Asaia* spp. that dominated in all laboratory samples (highest abundance in 5/6 samples, found in 6/6) was found sparingly in samples from all field locations (5/47) and was the dominant genera in none (Fig. 2). This genus has been seen in previous studies of mosquito microbiota from several mosquito genera (Osei-Poku et al., 2012; Minard et al., 2013). Its absence in most field mosquitoes may indicate it is not a ubiquitous part of the *Anopheles* flora throughout Africa, and that the less-diverse, *Asaia* dominated mosquito flora in colony mosquitoes may poorly recapitulate the wild state of mosquitoes in all natural areas. Additionally, as it was postulated that *Asaia's* colonization limits wide-spread *Wolbachia* presence (Hughes et al., 2014; Rossi et al., 2015), it may explain why *Wolbachia* has been found in mosquitoes in Burkina Faso with limited prevalence of *Asaia*

406   (Favia et al., 2007; Baldini et al., 2014; Buck et al., 2016). However, we found no *Wolbachia*

407   reads in any samples, though as these reads are difficult to amplify it may be below the limit of

408   detection (Baldini et al., 2014). Additionally, the presence of *Gluconobacter* spp. bacteria in

409   laboratory mosquitoes (but no field mosquitoes) may point to other ways in which the artificial

410   nature of the laboratory environment (i.e. sugar feeding on Karo syrup) could impact the

411   microbial composition, and possibly additional metrics such as vector competence.

412   A caveat to our microbiome composition data is that the sampling and next-generation

413   sequencing in this study was performed prior to the knowledge that there can be contaminant

414   "kitome" DNA present from DNA extraction kits, molecular water, and cross-over

415   contamination on machines (Salter et al., 2014). Due to this, we do not have control samples

416   from each of these items, and due to the time since processing, these kits are no longer available

417   to be sequenced. The issues of contamination have been reported to be more severe with samples

418   of low-bacterial abundance (Salter et al., 2014). We hope that in our samples this contamination

419   is limited due to whole-body sampling which would increase bacterial reads, and through finding

420   that there were no significant negative correlations between read count and presence of known

421   kitome genera (two-tailed Pearson's r, Table S2). This lack of correlation should not imply there

422   is no contamination possibility in our samples, but is the best assessment we can provide

423   retrospectively that presence of these genera is not linked to low-biomass in our samples.

424   Additionally, no reads were broadly found in all samples, and only the laboratory samples were

425   collected in different time periods than the others. Future studies should include these controls,

426   and assessment of aquatic bacterial species from the larval sites to limit these possible

427   confounders. Additional use of newer versions of next-generation sequencing approaches could

428   also help to improve sequencing coverage, allowing for more efficient filtering of these low

429   prevalence reads.

430   The presence of *Anaplasma ovis* bacteria in the Sahelian dry season samples was

431   unexpected, but has some precedence in the literature. *A. ovis* 16S sequence has been found

432   previously in *Anopheles gambiae* and *An. funestus* collected in western Kenya (Lindh et al.,

433   2005), and it remains unknown if this finding has any relevance to the disease's transmission or

434   is simply related to it being an intracellular pathogen of the blood that could be inadvertently

435   picked up during feeding. *Anaplasma phagocytophilum* has been shown to modulate *Ixodes* tick

436   microbiota, though the effects of *Anaplasma* species on mosquito microbiota has not been

437   studied to our knowledge (Abraham et al., 2017). This pathogen is also in the same order as the

438   *Wolbachia* genus bacteria that have achieved considerable attention as an *Aedine* symbiont and

439   possible vectorial-competence modifying species (Cirimotich, Ramirez & Dimopoulos, 2011;

440   Hughes et al., 2014; Minard et al., 2014).

441   To expand upon this finding of a zoonotic pathogen reads present in these classically

442   anthropophilic mosquitoes (Besansky, Hill & Costantini, 2004), the analysis of host-choice PCR

443   with cytochrome B allowed us to refine the biting characteristics of each population across

444   seasons. We found that the relaxation of strict anthropophily in the seasonal areas (Fig. 5) which

445   may follow what has been reported previously in areas of low host-availability (Lefèvre et al.,

446   2009). Due to the climactic severity during the dry season in the Sahel, the acquisition of blood

447   from the nearest source is likely less taxing than feeding on the preferred host. Though as there is

448   not an increase in zoophily in the N'Gabakoro samples in March-April in the dry season (when

449   conditions are similar between locations) and there is some zoophily in the wet season, there may

450   be an innate degree of zoophily in all *An. coluzzii* even in favorable conditions.

Conclusions:

451

452   This study of the seasonal microbiome of *An. coluzzii* indicates some unique aspects that

453   illuminates some behavioral characteristics during this cryptic and poorly understood time

454   period. It is largely characterized by mosquitoes in areas of higher seasonality exhibiting a

455   reduction in anthropophily, likely due to the unique environmental pressures seen in these areas,

456   and a higher overall diversity during the dry season regardless of the presence of permanent

457   water. Furthermore, it appears that the mosquito larval environment may be performing a

458   seeding function for the adult microbiome, and suggests minor evidence for the occurrence of

459   aestivation. Future analysis of the seasonality of host-choice could help to refine these findings,

460   and to show possible impacts towards transmission.

461

466

467

References:

469   Abraham NM., Liu L., Jutras BL., Yadav AK., Narasimhan S., Gopalakrishnan V., Ansari JM.,
470   Jefferson KK., Cava F., Jacobs-Wagner C., Fikrig E. 2017. Pathogen-mediated
471   manipulation of arthropod microbiota to promote infection. *Proceedings of the National*
472   *Academy of Sciences*:201613422. DOI: 10.1073/pnas.1613422114.
473   Akorli J., Gendrin M., Pels NAP., Yeboah-Manu D., Christophides GK., Wilson MD. 2016.
474   Seasonality and locality affect the diversity of Anopheles gambiae and Anopheles coluzzii
475   midgut microbiota from Ghana. *PLoS ONE* 11. DOI: 10.1371/journal.pone.0157529.
476   Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. 1990. Basic local alignment search
477   tool. *Journal of molecular biology* 215:403–10. DOI: 10.1016/S0022-2836(05)80360-2.
478   Apte-Deshpande A., Paingankar M., Gokhale MD., Deobagkar DN. 2012. Serratia odorifera a

479   midgut inhabitant of aedes aegypti mosquito enhances its susceptibility to dengue-2 virus.
480   *PLoS ONE* 7. DOI: 10.1371/journal.pone.0040401.

481 Bahia AC., Dong Y., Blumberg BJ., Mlambo G., Tripathi A., Benmarzouk-Hidalgo OJ., Chandra
482   R., Dimopoulos G. 2014. Exploring Anopheles gut bacteria for Plasmodium blocking
483   activity. *Environmental Microbiology* 16:2980–2994. DOI: 10.1111/1462-2920.12381.

484 Baldini F., Segata N., Pompon J., Marcenac P., Robert Shaw W., Dabiré RK., Diabaté A.,
485   Levashina EA., Catteruccia F. 2014. Evidence of natural Wolbachia infections in field
486   populations of Anopheles gambiae. *Nature Communications* 5:1–7. DOI:
487   10.1038/ncomms4985.

488 Beals EW. 1984. Bray-curtis ordination: An effective strategy for analysis of multivariate
489   ecological data. *Advances in Ecological Research* 14:1–55. DOI: 10.1016/S0065-
490   2504(08)60168-3.

491 Benjamini Y., Hochberg Y. 1995. Controlling the False Discovery Rate : A Practical and
492   Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*
493   57:289–300.

494 Besansky NJ., Hill CA., Costantini C. 2004. No accounting for taste: Host preference in malaria
495   vectors. *Trends in Parasitology* 20:249–251. DOI: 10.1016/j.pt.2004.03.007.

496 Bivand RS., Pebesma E., Gomez-Rubio V. 2013. *Applied spatial data analysis with {R}, Second*
497   *edition*. Springer, NY.

498 Boissière A., Tchioffo MT., Bachar D., Abate L., Marie A., Nsango SE., Shahbazkia HR.,
499   Awono-Ambene PH., Levashina EA., Christen R., Morlais I. 2012. Midgut microbiota of
500   the malaria mosquito vector Anopheles gambiae and interactions with Plasmodium
501   falciparum infection. *PLoS Pathogens* 8:1–12. DOI: 10.1371/journal.ppat.1002742.

502 Borgne P Le., Legendre G., Marsouin A. 2007. Operational Sst Retrieval From MetOp /
503   AVHRR. In: *Proc. 2007 EUMETSAT Conf.* 1–10.

504 Breiman L. 2001. Random forests. *Machine Learning* 45:5–32. DOI:
505   10.1023/A:1010933404324.

506 Buck M., Nilsson LKJ., Brunius C., Dabire RK., Hopkins R., Terenius O. 2016. Bacterial
507   associations reveal spatial population dynamics in Anopheles gambiae mosquitoes.
508   *Scientific Reports* 6:22806:1–9. DOI: 10.1038/srep22806.

509 Callahan B. 2016. The Silva taxonomic training set formatted for DADA2. DOI:
510   10.5281/zenodo.158958.

511 Callahan BJ., Mcmurdie PJ., Rosen MJ., Han AW., Johnson AJA., Holmes SP. 2016a. dada2:
512   high-resolution sample inference from illumina amplicon data. *Nature Methods* 13:581–
513   583. DOI: 10.1038/nMeth.3869.

514 Callahan BJ., Sankaran K., Fukuyama JA., McMurdie PJ., Holmes SP. 2016b. Bioconductor
515   Workflow for Microbiome Data Analysis: from raw reads to community analyses [version
516   2; referees: 3 approved]. *F1000Research* 5:1492–21492.

517 Caporaso JG., Kuczynski J., Stombaugh J., Bittinger K., Bushman FD., Costello EK., Fierer N.,
518   Peña AG., Goodrich JK., Gordon JI., Huttley G a., Kelley ST., Knights D., Koenig JE., Ley
519   RE., Lozupone C a., Mcdonald D., Muegge BD., Pirrung M., Reeder J., Sevinsky JR.,
520   Turnbaugh PJ., Walters W a., Widmann J., Yatsunenko T., Zaneveld J., Knight R. 2010.
521   QIIME allows analysis of high- throughput community sequencing data. *Nature Publishing*
522   *Group* 7:335–336. DOI: 10.1038/nmeth0510-335.

523 Cirimotich CM., Dong Y., Clayton AM., Sandiford SL., Souza-Neto JA., Mulenga M.,
524   Dimopoulos G. 2011. Natural Microbe-Mediated Refractoriness to Plasmodium Infection in

525   Anopheles gambiae. *Science* 332:855–858. DOI: 10.1126/science.1201618.
526 Cirimotich CM., Ramirez JL., Dimopoulos G. 2011. Native microbiota shape insect vector
527   competence for human pathogens. *Cell Host and Microbe* 10:307–310. DOI:
528   10.1016/j.chom.2011.09.006.
529 Coon KL., Vogel KJ., Brown MR., Strand MR. 2014. Mosquitoes rely on their gut microbiota
530   for development. *Molecular Ecology* 23:2727–2739. DOI: 10.1111/mec.12771.
531 Coulibaly MB., Lobo NF., Fitzpatrick MC., Kern M., Grushko O., Thaner D V., Traoré SF.,
532   Collins FH., Besansky NJ. 2007. Segmental duplication implicated in the genesis of
533   inversion 2Rj of anopheles gambiae. *PLoS ONE* 2:1–9. DOI:
534   10.1371/journal.pone.0000849.
535 Dao A., Yaro AS., Diallo M., Timbiné S., Huestis DL., Kassogué Y., Traoré a I., Sanogo ZL.,
536   Samaké D., Lehmann T. 2014. Signatures of aestivation and migration in Sahelian malaria
537   mosquito populations. *Nature* 516:387–90. DOI: 10.1038/nature13987.
538 David MR., Santos LMB Dos., Vicente ACP., Maciel-de-Freitas R. 2016. Effects of
539   environment, dietary regime and ageing on the dengue vector microbiota: Evidence of a
540   core microbiota throughout Aedes aegypti lifespan. *Memorias do Instituto Oswaldo Cruz*
541   111:577–587. DOI: 10.1590/0074-02760160238.
542 Dennison NJ., Jupatanakul N., Dimopoulos G. 2014. The mosquito microbiota influences vector
543   competence for human pathogens. *Current Opinion in Insect Science* 3:6–13. DOI:
544   10.1016/j.cois.2014.07.004.
545 Dowd SE., Wolcott RD., Sun Y., McKeehan T., Smith E., Rhoads D. 2008. Polymicrobial nature
546   of chronic diabetic foot ulcer biofilm infections determined using bacterial tag encoded
547   FLX amplicon pyrosequencing (bTEFAP). *PLoS ONE* 3. DOI:
548   10.1371/journal.pone.0003326.
549 Edillo FE., Touré YT., Lanzaro GC., Dolo G., Taylor CE. 2002. Spatial and habitat distribution
550   of Anopheles gambiae and Anopheles arabiensis (Diptera: Culicidae) in Banambani village,
551   Mali. *Journal of medical entomology* 39:70–77. DOI: 10.1603/0022-2585-39.1.70.
552 Fanello C., Santolamazza F., della Torre A. 2002. Simultaneous identification of species and
553   molecular forms of the Anopheles gambiae complex by PCR-RFLP. *Med Vet Entomol*
554   16:461–464. DOI: 10.1046/j.1365-.
555 Favia G., Ricci I., Damiani C., Raddadi N., Crotti E., Marzorati M., Rizzi A., Urso R., Brusetti
556   L., Borin S., Mora D., Scuppa P., Pasqualini L., Clementi E., Genchi M., Corona S., Negri
557   I., Grandi G., Alma A., Kramer L., Esposito F., Bandi C., Sacchi L., Daffonchio D. 2007.
558   Bacteria of the genus Asaia stably associate with Anopheles stephensi, an Asian malarial
559   mosquito vector. *Proceedings of the National Academy of Sciences* 104:9047–9051. DOI:
560   10.1073/pnas.0610451104.
561 Gendrin M., Christophides GK. 2013. The Anopheles Mosquito Microbiota and Their Impact on
562   Pathogen Transmission. In: *Anopheles mosquitoes - New insights into malaria vectors*. 525–
563   548. DOI: 10.5772/3392.
564 Gilligan PH., Lum G., Vandamme P., Whittier S. 2003. Burkholderia, Stenotrophomonas,
565   Ralstonia, Brevundimonas, Comamonas, Delftia, Pandoraea, and Acidovorax. In: Murray
566   PR, Baron EJ, Jorgenson JH, Pfaller MA, Yolken RH eds. *Manual of clinical microbiology*.
567   ASM Press, 729–748.
568 Hervé M. 2017. RVAideMemoire: Diverse Basic Statistical and Graphical Functions.
569 Hijmans RJ. 2016. raster: Geographic Data Analysis and Modeling.
570 Huestis DL., Lehmann T. 2014. Ecophysiology of Anopheles gambiae s.l.: Persistence in the

571   Sahel. *Infection, Genetics and Evolution* 28:648–661. DOI: 10.1016/j.meegid.2014.05.027.

572 Huestis DL., Yaro AS., Traoré AI., Dieter KL., Nwagbara JI., Bowie AC., Adamou A.,
573   Kassogué Y., Diallo M., Timbiné S., Dao A., Lehmann T. 2012. Seasonal variation in
574   metabolic rate, flight activity and body size of Anopheles gambiae in the Sahel. *The Journal*
575   *of experimental biology* 215:2013–21. DOI: 10.1242/jeb.069468.

576 Hughes GL., Dodson BL., Johnson RM., Murdock CC., Tsujimoto H., Suzuki Y., Patt AA., Cui
577   L., Nossa CW., Barry RM., Sakamoto JM., Hornett EA., Rasgon JL. 2014. Native
578   microbiome impedes vertical transmission of Wolbachia in Anopheles mosquitoes.
579   *Proceedings of the National Academy of Sciences* 111:12498–12503. DOI:
580   10.1073/pnas.1408888111.

581 Jupatanakul N., Sim S., Dimopoulos G. 2014. The insect microbiome modulates vector
582   competence for arboviruses. *Viruses* 6:4294–4313. DOI: 10.3390/v6114294.

583 Kent RJ., Norris DE. 2005. Identification of mammalian blood meals in mosquitoes by a
584   multiplexed polymerase chain reaction targeting cytochrome B. *The American journal of*
585   *tropical medicine and hygiene* 73:336–42.

586 Kuhn M., Contributions:, Weston S., Williams A., Keefer C., Engelhardt A., Cooper T., Mayer
587   Z., Kenkel B., Team the RC., Benesty M., Lescarbeau R., Ziem A., Scrucca L., Tang Y.,
588   Candan C., Hunt. T. 2016. caret: Classification and Regression Training.

589 de la Fuente J., Atkinson MW., Naranjo V., Fernández de Mera IG., Mangold AJ., Keating KA.,
590   Kocan KM. 2007. Sequence analysis of the msp4 gene of Anaplasma ovis strains.
591   *Veterinary Microbiology* 119:375–381. DOI: 10.1016/j.vetmic.2006.09.011.

592 Lefèvre T., Gouagna LC., Dabiré KR., Elguero E., Fontenille D., Renaud F., Costantini C.,
593   Thomas F. 2009. Beyond nature and nurture: Phenotypic plasticity in blood-feeding
594   behavior of Anopheles gambiae s.s. when humans are not readily accessible. *American*
595   *Journal of Tropical Medicine and Hygiene* 81:1023–1029. DOI: 10.4269/ajtmh.2009.09-
596   0124.

597 Liaw A., Wiener M. 2002. Classification and Regression by randomForest. *R News* 2:18–22.

598 Lindh JM., Lindh JM., Terenius O., Terenius O., Faye I., Faye I. 2005. 16S rRNA Gene-Based
599   Identification of Midgut Bacteria from Field-Caught *Anopheles gambiae* Sensu Lato and *A.*
600   *funestus* Mosquitoes Reveals New Species Related to Known Insect Symbionts. *Applied*
601   *and Environmental Microbiology* 71:7217–7223. DOI: 10.1128/AEM.71.11.7217.

602 Love MI., Anders S., Hu- W. 2016. *Differential analysis of count data – the DESeq2 package*.

603 Love MI., Huber W., Anders S. 2014. Moderated estimation of fold change and dispersion for
604   RNA-seq data with DESeq2. *Genome Biology* 15:550. DOI: 10.1186/s13059-014-0550-8.

605 McMurdie PJ., Holmes S. 2013. phyloseq: An R package for reproducible interactive analysis
606   and graphics of microbiome census data. *PLoS ONE* 8:e61217.

607 Minard G., Tran F-H., Dubost A., Tran-Van V., Mavingui P., Moro CV. 2014. Pyrosequencing
608   16S rRNA genes of bacteria associated with wild tiger mosquito Aedes albopictus: a pilot
609   study. *Frontiers in cellular and infection microbiology* 4:59. DOI:
610   10.3389/fcimb.2014.00059.

611 Minard G., Tran FH., Raharimalala FN., Hellard E., Ravelonandro P., Mavingui P., Valiente
612   Moro C. 2013. Prevalence, genomic and metabolic profiles of Acinetobacter and Asaia
613   associated with field-caught Aedes albopictus from Madagascar. *FEMS Microbiology*
614   *Ecology* 83:63–73. DOI: 10.1111/j.1574-6941.2012.01455.x.

615 Oksanen J., Blanchet FG., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin PR.,
616   O'Hara RB., Simpson GL., Solymos P., Stevens MHH., Szoecs E., Wagner H. 2017. vegan:

617   Community Ecology Package.

618   Osei-Poku J., Mbogo CM., Palmer WJ., Jiggins FM. 2012. Deep sequencing reveals extensive

619   variation in the gut microbiota of wild mosquitoes from Kenya. *Molecular Ecology*

620   21:5138–5150. DOI: 10.1111/j.1365-294X.2012.05759.x.

621   Paradis E., Claude J., Strimmer K. 2004. A{PE}: analyses of phylogenetics and evolution in {R}

622   language. *Bioinformatics* 20:289–290.

623   Pebesma EJ., Bivand RS. 2005. Classes and methods for spatial data in {R}. *R News* 5:9–13.

624   Perpiñán O., Hijmans R. 2016. rasterVis.

625   Pirmohamed T., Dowding JM., Singh S., Wasserman B., Heckert E., Karakoti AS., King JES.,

626   Seal S., Self WT. 2010. Nanoceria exhibit redox state-dependent catalase mimetic activity.

627   *Chemical communications (Cambridge, England)* 46:2736–8. DOI: 10.1039/b922024k.

628   Price MN., Dehal PS., Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees

629   for large alignments. *PLoS ONE* 5. DOI: 10.1371/journal.pone.0009490.

630   Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., Glöckner FO. 2013.

631   The SILVA ribosomal RNA gene database project: Improved data processing and web-

632   based tools. *Nucleic Acids Research* 41:590–596. DOI: 10.1093/nar/gks1219.

633   R Core Team. 2016. R: A Language and Environment for Statistical Computing.

634   Rossi P., Ricci I., Cappelli A., Damiani C., Ulissi U., Mancini MV., Valzano M., Capone A.,

635   Epis S., Crotti E., Chouaia B., Scuppa P., Joshi D., Xi Z., Mandrioli M., Sacchi L., O'Neill

636   SL., Favia G. 2015. Mutual exclusion of Asaia and Wolbachia in the reproductive organs of

637   mosquito vectors. *Parasites & Vectors* 8:278. DOI: 10.1186/s13071-015-0888-0.

638   RStudio Team. 2015. RStudio: Integrated Development Environment for R.

639   Ryan M., Pembroke J., Adley C. 2006. Ralstonia pickettii: a persistent Gram-negative

640   nosocomial infectious organism. *Journal of Hospital Infection* 62:278–284. DOI:

641   10.1016/j.jhin.2005.08.015.

642   Salter SJ., Cox MJ., Turek EM., Calus ST., Cookson WO., Moffatt MF., Turner P., Parkhill J.,

643   Loman NJ., Walker AW. 2014. Reagent and laboratory contamination can critically impact

644   sequence-based microbiome analyses. *BMC Biology* 12:87. DOI: 10.1186/s12915-014-

645   0087-z.

646   Sankaran K., Holmes S. 2014. structSSI: Simultaneous and Selective Inference for Grouped or

647   Hierarchically Structured Data. *Journal of statistical software* 59:1–21. DOI:

648   10.18637/jss.v059.i13.

649   Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27:592–593.

650   van Tol S., Dimopoulos G. 2016. *Influences of the Mosquito Microbiota on Vector Competence*.

651   Elsevier Ltd. DOI: 10.1016/bs.aiip.2016.04.006.

652   Wang Q., Garrity GM., Tiedje JM., Cole JR. 2007. Naïve Bayesian classifier for rapid

653   assignment of rRNA sequences into the new bacterial taxonomy. *Applied and*

654   *Environmental Microbiology* 73:5261–5267. DOI: 10.1128/AEM.00062-07.

655   Wang Y., Gilbreath TM., Kukutla P., Yan G., Xu J. 2011. Dynamic gut microbiome across life

656   history of the malaria mosquito anopheles gambiae in Kenya. *PLoS ONE* 6:1–9. DOI:

657   10.1371/journal.pone.0024767.

658   Weiss S., Xu ZZ., Peddada S., Amir A., Bittinger K., Gonzalez A., Lozupone C., Zaneveld JR.,

659   Vázquez-Baeza Y., Birmingham A., Hyde ER., Knight R. 2017. Normalization and

660   microbial differential abundance strategies depend upon data characteristics. *Microbiome*

661   5:27. DOI: 10.1186/s40168-017-0237-y.

662   Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

663 Wickham H. 2016. scales: Scale Functions for Visualization.

664 Wright ES. 2016. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R*

665 *Journal* 8:352–359.

666 Yaro AS., Traoré AI., Huestis DL., Adamou A., Timbiné S., Kassogué Y., Diallo M., Dao A.,

667 Traoré SF., Lehmann T. 2012. Dry season reproductive depression of Anopheles gambiae in

668 the Sahel. *Journal of Insect Physiology* 58:1050–1059. DOI:

669 10.1016/j.jinsphys.2012.04.002.

670 Yekutieli D. 2008. Hierarchical False Discovery Rate–Controlling Methodology. *Journal of the*

671 *American Statistical Association* 103:309–316. DOI: 10.1198/016214507000001373.

672 Yu G., Smith D., Zhu H., Guan Y., Lam TT-Y. 2017. ggtree: an R package for visualization and

673 annotation of phylogenetic trees with their covariates and other associated data. *Methods in*

674 *Ecology and Evolution* 8:28–36. DOI: 10.1111/2041-210X.12628.

675

676

677

678

679

Figure 1

Sampling locations in the Koulikoro District of Mali overlaid over seasonal Normalized Difference Vegetation Index data (panel A) and at each location throughout the sampling period (panel B).

The villages of M'Piabougou and Zanga in the area of Thierola with higher seasonality and lower water availability are compared to N'Gabakoro, a village near Bamako and the Niger River. Examples of wet season (top left, bottom right of panel A), transitional (top right of panel A), and dry season (bottom left panel A) are presented. Microbial sampling dates per village are marked with vertical dashed lines (panel B).
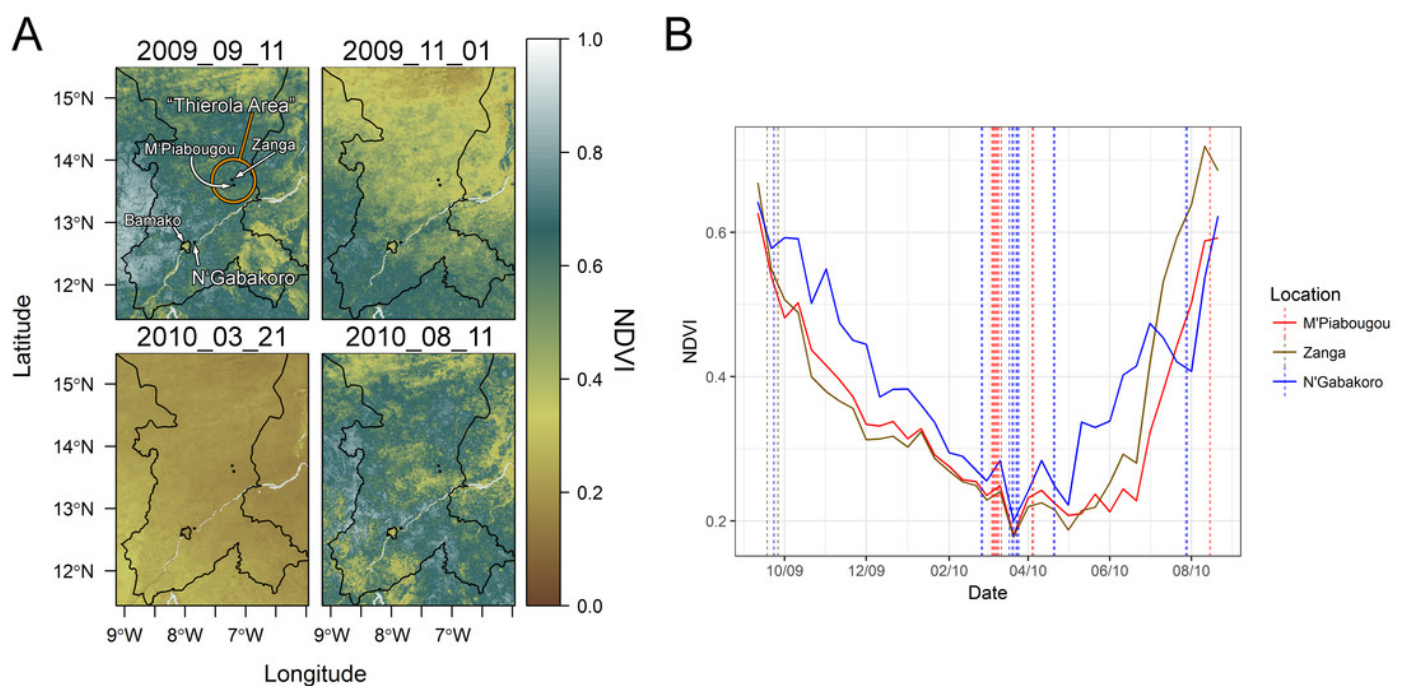
Figure 2

Stacked bar plot (top) and within-sample Shannon diversity (bottom) comparing the 19 most abundant agglomerated microbial genera between dry and wet seasons.

All other taxa are grouped in the "other" category. Sample numbers and whether it is a pooled sample ("P") are marked above and below bar plots, respectively. Only the top 19 genera are shown here for ease of interpretation, all analyses between groups are performed with all sequence variants or taxa.
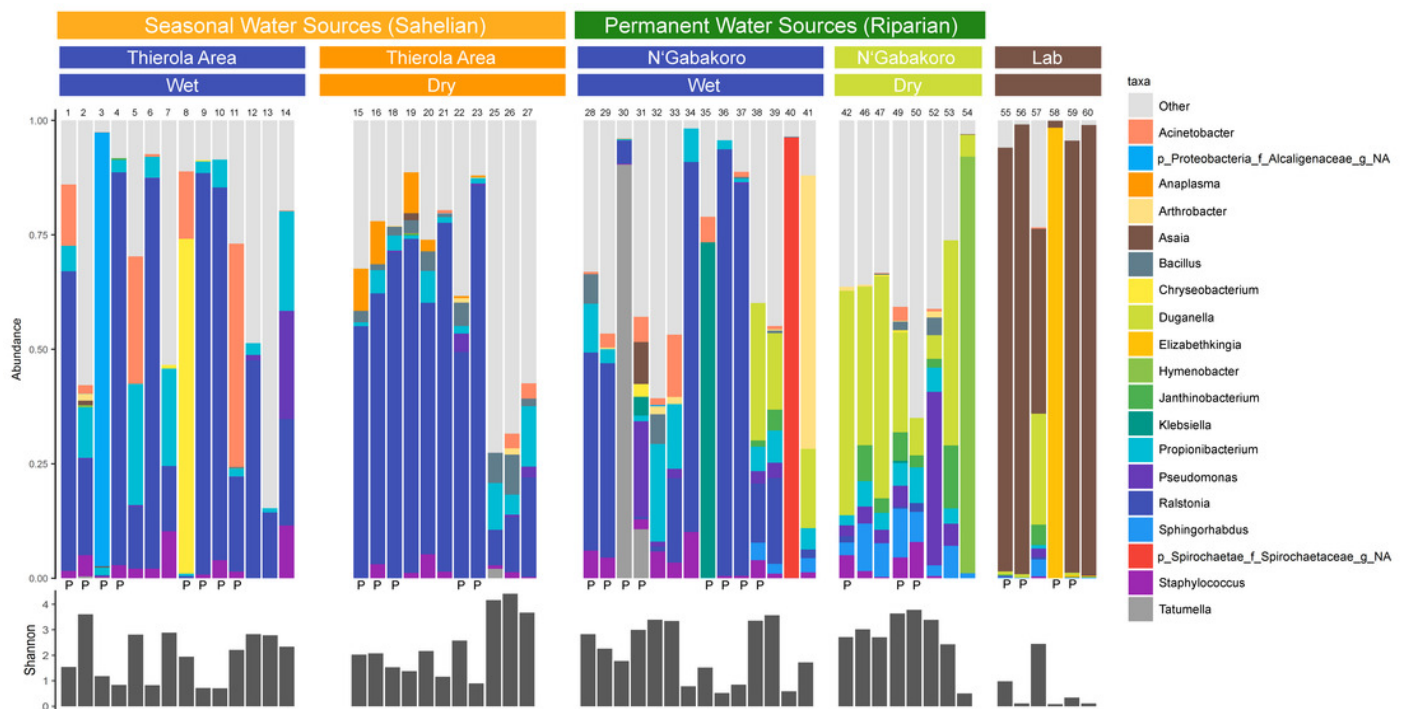
Figure 3

Hierarchical Multiple Testing (HMT) of differentially abundant 16s rRNA sequence variants.

Log_2 fold changes calculated via DESeq2 greater than 1.5-fold that are significant at an adjusted *p*-value < 0.05 are presented for each location. HMT is a false discovery rate adjusting methodology that arranges tested hypotheses via their phylogeny, testing sub-hypotheses only if their parent hypothesis is significant.

A

Thierola Wet to Dry Season

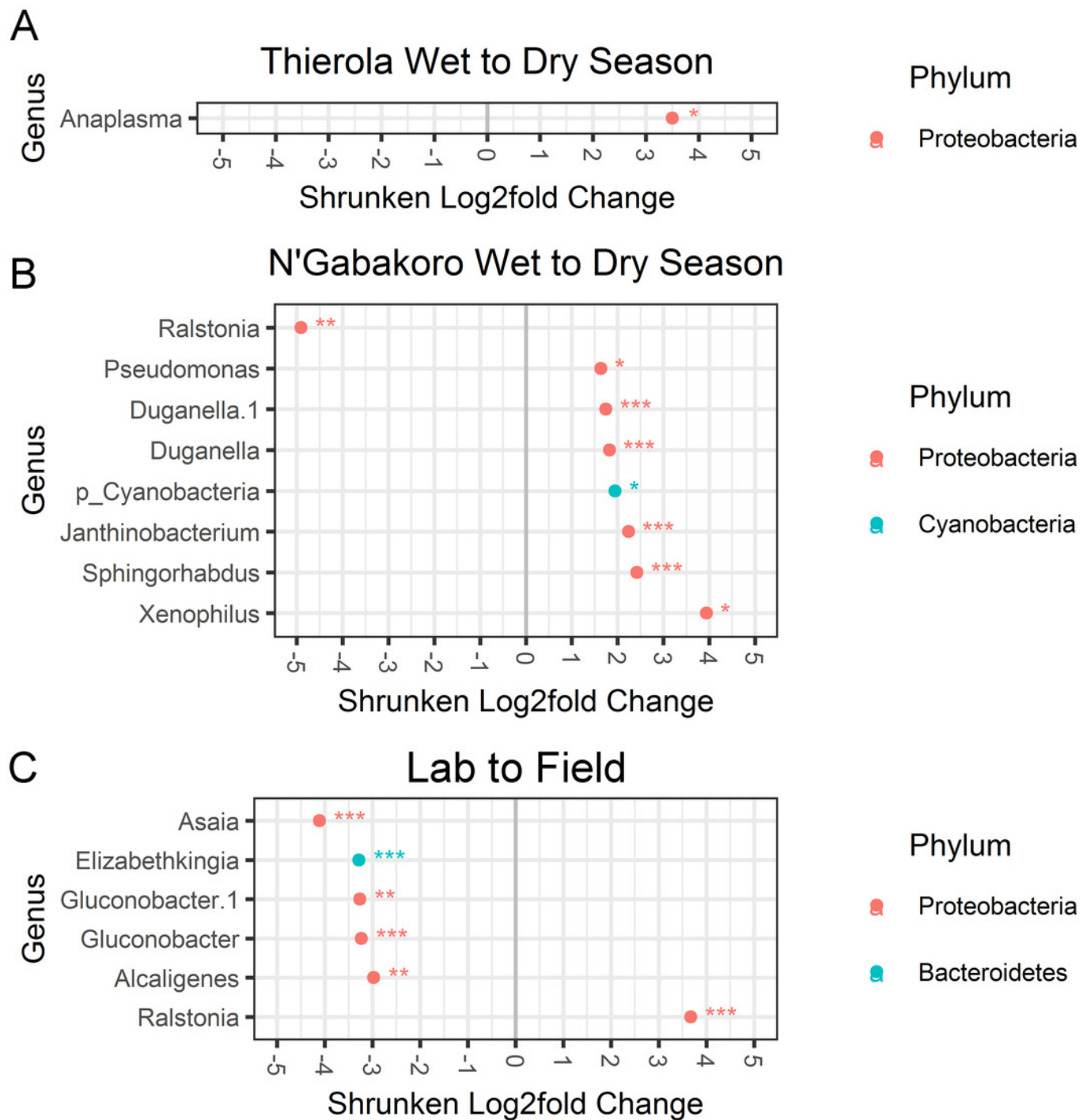## N'Gabakoro Wet to Dry Season

Lab to Field

Figure 4

Principal Coordinates Analysis of Bray-Curtis dissimilarity.

(A) 16s sequence variants that are filtered to only those present in > 1 sample, and that have been rarefied to an even depth of 500 reads are compared via two clustering methodologies. Significance of sample group centroids (center points of each Season/Location grouping) are compared via a pairwise permutation-based multivariate ANOVA (pairwise.PERMANOVA) (Panel B). All *p*-values listed have been adjusted via Benjamini-Hochberg for false discovery rate control.

A
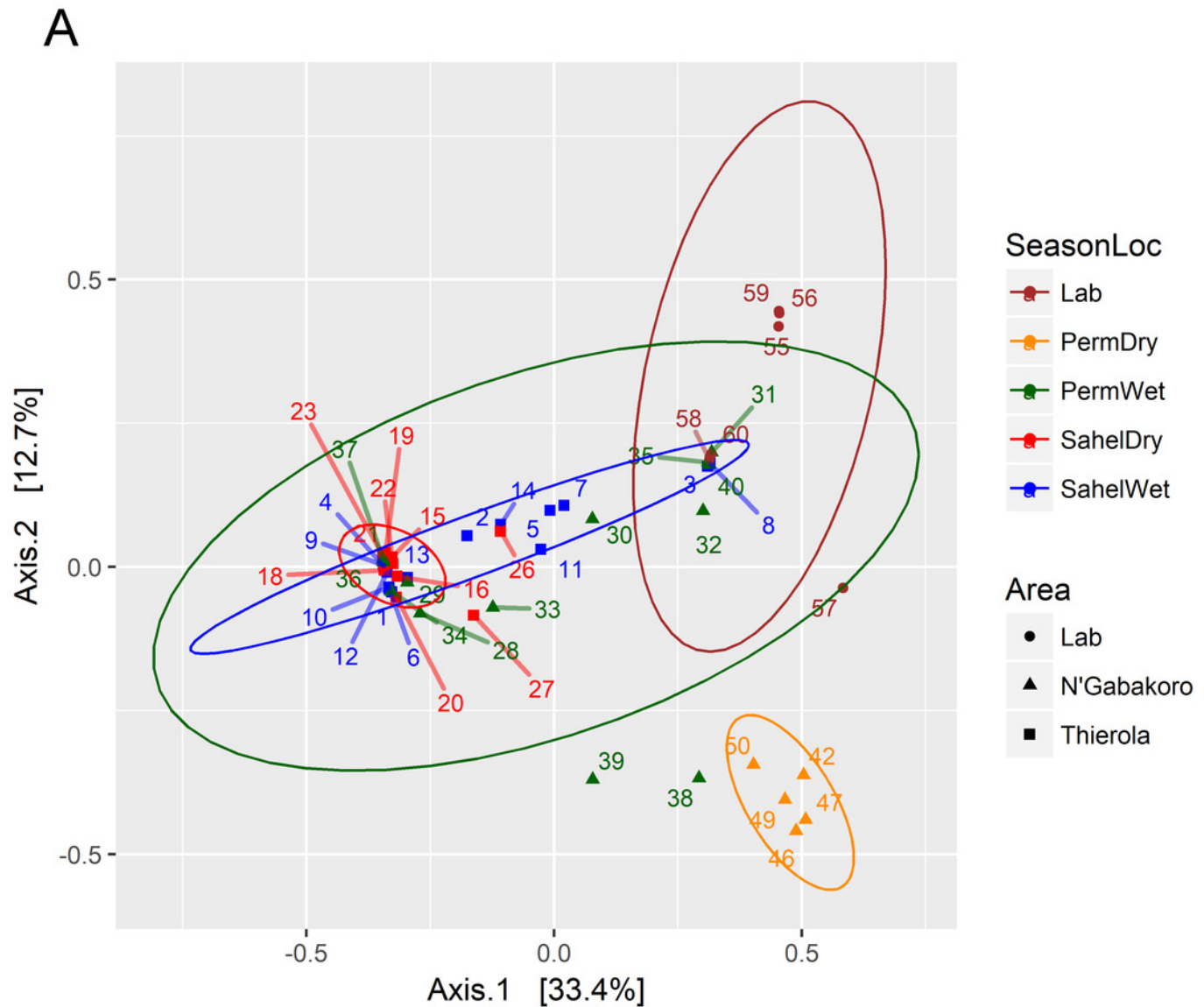
B

	LAB	PERMDRY	PERMWET	SAHELDRY
PERMDRY	3.43E-03			
PERMWET	2.50E-04	6.67E-04		
SAHELDRY	2.50E-04	6.00E-04	0.019	
SAHELWET	2.50E-04	2.50E-04	0.296	0.092

Figure 5

Host-choice preference determined via Cytochrome B PCR.

Significant difference between "human" and "other" bloodmeals for each location and season determined via a contingency table and two-tailed Fisher's exact test.