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

Authors: Benjamin J. Krajacich*¹, Diana L. Huestis¹, Adama Dao², Alpha S. Yaro², Moussa Diallo², Asha Krishna¹, Jiannong Xu³, Tovi Lehmann*¹

*Corresponding authors

¹Laboratory of Malaria and Vector Research, NIAID, NIH, Rockville, Maryland, USA
²Malaria Research and Training Center (MRTC)/ Faculty of Medicine, Pharmacy and Odonto-stomatology, Bamako, Mali
³Biology Department, New Mexico State University, Las Cruces, New Mexico, USA

Abstract:

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 ribosomal RNA gene tag 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.
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
states of reproductive vs. aestivating mosquitoes in the wet and dry seasons, respectively.

Additionally, we compared how laboratory mosquitoes compare to these field-caught specimens in their microbial composition.

**Methods:**

*Mosquito collection and field sites:*

*An. coluzzii* species mosquitoes were collected via indoor aspiration in three locations, Zanga (Latitude 13.688050°, Longitude -7.221029°), M’Piabougou (13.599830°, -7.192859°), and N’Gabakoro (12.683870°, -7.840419°) in the Koulikoro region of western Mali from September 2009 to August 2010 (Fig. 1). Mosquitoes from each location were noted for blood fed status, had their thoraces punctured, and were added to 50 µl RNA*later* stabilization solution (ThermoFisher Scientific, Waltham, MA, USA) for preservation. Fourteen additional laboratory reared *An. coluzzii* were also preserved in groups of young (3 days old, Samples 55-57) and old (14 days old, Samples 58-60) post emergence to compare microbiome development over time.

Mosquitoes were speciated to molecular form based on a direct PCR performed on two legs with a standard protocol (Fanello, Santolamazza & della Torre, 2002). DNA was extracted from mosquitoes with a DNeasy kit according to manufacturer instructions (Qiagen, Valencia, CA, USA), and eluted in 50 µl of the provided elution buffer.

*Normalized Difference Vegetation Index (NDVI) analysis:*

To further quantify the differences in seasonality between our Sahelian and riparian areas, we calculated the NDVI around the center point of each of the 3 field sampling locations with a ~5 km square on the NDVI output from the Metop-AVHRR S10 (“ENDVI10”, 1 km resolution, daily measurements with best composite image per dekad available) global satellite database (Accessed 5/1/2017, from: http://www.vito-
measurements spanning the sampling collection period were downloaded for the African subcontinent in Erdas Imagine (IMG) format and were analyzed using the ‘raster’ library in R with a spatial grid cell size of 0.0025° defined with the ‘sp’ package (Pebesma & Bivand, 2005; Bivand, Pebesma & Gomez-Rubio, 2013; Hijmans, 2016). NDVI data was plotted for four representative time points using the ‘rasterVis’ package (Fig. 1A) and for each of the available NDVI images during the sampling period using the ‘ggplot2’ and ‘scales’ packages (Fig. 1B) (Wickham, 2009, 2016; Perpiñán & Hijmans, 2016). Shape files for Mali and the sampling regions are acquired from the Database of Global Administrative Areas (GADM) database with ‘raster’.

**454 Sequencing:**

Next-generation sequencing of the V1-V3 regions of 16S ribosomal RNA gene was performed by MR DNA (Shallowater, TX, USA) on a Roche 454 sequencer using the forward primer “27FMod”- 5’-AGRGTTTGATCMTGGCTCAG-3’ and the reverse primer “519Rmodbio” – 5’-GWATTACCGCGGCKGCTG -3’. Amplification and sequencing conditions used are described in full elsewhere (Dowd et al., 2008; Pirmohamed et al., 2010).

Mosquitoes were sequenced either as individuals from 15ng of DNA, or from a pool of three mosquitoes (5ng from each) with concentrations being determined by NanoDrop (ThermoFisher).

**Processing Pipeline:**

The .fasta and .qual files with primers and barcodes removed are split using the split.libraries.py command in QIIME v1.9.0 using the –d flag to record quality scores (Caporaso et al., 2010). The output was then demultiplexed into individual fastq files in QIIME. FastQ files
are then imported into R version 3.4.0 using the RStudio IDE version 1.0.44 using the ‘dada2’ package (RStudio Team, 2015; Callahan et al., 2016a; R Core Team, 2016). Processing using this pipeline largely follows the Bioconductor workflow from Callahan et al. (Callahan et al., 2016b). Briefly, data are trimmed/filtered with settings truncLen=325, truncQ=2, maxN=0, and maxEE=2. Sequencing error rates are learned on a random sample of \( n = 25 \), with suggested parameters for 454 data of HOMOPOLYMER_GAP_PENALTY=-1 and BAND_SIZE=32. Chimeras are removed, and taxa is assigned with a naïve Bayesian classifier algorithm over 100 bootstraps with a minimum bootstrap level of 50 (Wang et al., 2007), and taxonomy is assigned using the ‘silva_nr_v123_train_set’ for 16S rRNA or the UNITE database for ITS formatted for use in dada2 (Silva training set: https://zenodo.org/record/158958, UNITE database (General FASTA Release): https://unite.ut.ee/repository.php) (Quast et al., 2013; Callahan, 2016). For 16S this is based on a 325nt length amplicon and the ITS has a variable length amplicon with a minimum of 50 nt length.

The sequences are aligned with the ‘DECIPHER’ package in R, output with the ‘phangorn’ package version 2.1.1 into FastTree 2 software for generation of a generalized time-reversible (GTR) maximum-likelihood phylogenetic tree with rescaling of branch lengths and computation of Gamma20-based likelihood (Price, Dehal & Arkin, 2010; Schliep, 2011; Wright, 2016). This tree is read back into R using the ‘ape’ package (Paradis, Claude & Strimmer, 2004), and then all the data are combined with the ‘phyloseq’ package for data manipulation and visualization (McMurdie & Holmes, 2013). After contaminant reads from eukaryotic sources were removed the Shannon index of richness is calculated for each mosquito. Phylogenetic tree visualization is performed using the package ‘ggtree’ (Yu et al., 2017).

Differential abundance testing and clustering analysis:
Seasonal differences in microbial abundance were analyzed with the ‘DESeq2’ package on genus agglomerated data (Love, Anders & Huber, 2016). This method creates a negative binomial generalized linear model (GLM) to estimate maximum likelihood for each genus’s count based on log₂ fold changes between two conditions (i.e. Dry vs. Wet at that location), while accounting for size factors (differences in sequence depth between samples) and dispersion (between sample variance) using Bayesian shrinkage (Love, Huber & Anders, 2014; Weiss et al., 2017). We compared samples seasonally per location, and also all field vs. all laboratory samples. The p-value from the Wald’s test is then adjusted for a false discovery rate with the Benjamini-Hochberg approach and local smoothed dispersion fit (Benjamini & Hochberg, 1995). Shrunken (adjusted if they have high dispersion or low sampling coverage) log₂ fold changes between seasons at each location are generated by contrasting “PermDry” to “PermWet”; “SahelDry” to “SahelWet”; and “Field” to “Lab” for the samples from the riparian, Sahelian, and field sites, respectively.

Hierarchical multiple testing, a procedure by which you incorporate the innate structure in the data (in this case the phylogenetic hierarchy) to adjust your false discovery rate and improve power in comparing count data (Yekutieli, 2008), was implemented in R with the ‘structssi’ package (Sankaran & Holmes, 2014). We defer the majority of details of this approach to the parent literature (Yekutieli, 2008), and to its implementation literature for microbiome data (Sankaran & Holmes, 2014; Callahan et al., 2016b). Briefly, we control the false discovery rate by organizing the hypotheses based on the phylogenetic tree of the data, and only testing for differential abundance if the parent hypothesis is found to be significant between groups at a coarse level of discrimination (a ‘hFDR’ rate of 0.75) (Callahan et al., 2016b). This analysis is performed on 16S sequence variant count data that has been variance stabilizing transformed by...
‘DESeq2’ and shifted so all values are positive. Genera with greater than 1.5 shrunken-log₂-fold change that are found to be significant at an adjusted p-value of less than 0.05 are presented.

Ordination plots were generated from prevalence filtered (must not be sample singletons) and rarefied data (rarefied to 500 reads) using ‘phyloseq’ and ‘vegan’ with principal coordinates analysis on Bray-Curtis similarity distances (Beals, 1984; Oksanen et al., 2017). Analysis of centroid (center point) differences between sample groups (Season and Location) were compared using a pairwise permutation multivariate analysis of variance test (PERMANOVA) with 9999 permutations with Benjamini-Hochberg false discovery rate correction using the ‘RVAideMemoire’ package (Hervé, 2017).

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

In addition to analysis of the variance in bacterial taxon abundance described above, we performed two analyses based on the presence/absence of bacterial taxa in mosquitoes, regardless of the abundance. In the first, we performed random forest supervised classification analysis on genus-level data using the ‘caret’ package in R (Breiman, 2001; Liaw & Wiener, 2002; Kuhn et al., 2016). The proximity metric calculated during the random forest model generation was utilized to generate a multi-dimensional scaling plot (MDS) to evaluate the degree of difference and clustering between and within sample groups via PERMANOVA.

Prevalence for each of the sample groups for the top 10 most important genera in the random forest classification model are presented.

Additionally, for field samples we enumerated taxa unique to the Sahelian dry season sample, found their relative frequency (in terms of mosquito hosts), and evaluated how unique taxa are to that sample group (‘private’ - i.e. found only in that sample group) in SAS 9.4 (SAS Institute Inc., Cary, NC, USA). Using “informative” taxa (excluding rare and ubiquitous taxa),
we sequentially tested if the frequency (prevalence) of any informative taxa differed between the
dry and wet seasons in the Sahel, followed by testing that the Sahelian dry season differs from
riparian dry season, and that it also differed from the riparian wet season in a consistent
direction. These tests employed the sequential Fisher’s exact tests (with the last two being one-
way tests), only passing genera significant at $p < 0.1$ forwards to the next test. The final results
only report those genera whose prevalence difference is in the same direction (i.e. higher in
Sahel Dry compared to all other conditions).

**Cytochrome B bloodmeal analysis:**

The mammalian host source of each mosquito and pool per location was analyzed via a
size-discriminate multiplexed cytochrome B PCR developed by Kent and Norris (Kent & Norris,
2005). This PCR allows for the identification of bloodmeals from pigs, humans, goat, dog and
cows.

**Results:**

Unlike the Sahelian villages, the proximity of N’Gabakoro to the Niger allows mosquito
breeding year-round. To further assess the seasonal differential in aridity between the Sahelian
and riparian villages, we analyzed the normalized difference vegetation index (NDVI, Fig. 1.).
Due to its proximity to the Niger river, we found that there is a desiccation lag-period in which
the vegetation index does not drop at the same rate in N’Gabakoro as it does in the two villages
that have only rainfall as a water source (Fig. 1A top right, Fig. 1B). Additionally, while the
vegetation minima are reached roughly at the same time in the dry season between sites (late
March to early April), the overall NDVI remains higher in N’Gabakoro during the transition
periods to and from this dry season low point.

**454 Sequencing:**
Of the 58 samples sent for 454 sequencing, 53 returned sequence for 16S and 26 returned sequence for ITS. For 16S sequencing 2643 sequence variants (SVs) were returned though a range of these variants aligned to likely contaminating eukaryotic reads in the SILVA database and were filtered, leaving 2057 sequence variants with 426 genera across 184 families. Many of these reads when searched against the ‘nr’ database using BLAST had hits to *An. gambiae* strain PEST sequence (Altschul et al., 1990). This could indicate some cross-reactivity of the primers with other ribosomal sequence, or mis-annotated reference sequence. The fungal internal transcribed spacer (ITS) region sequencing had a more limited success in amplification in terms of both average read counts per successful sample (755 vs. 4554 for ITS and 16S, respectively), and in the amount of coverage across locations/seasons. ITS failed to amplify from any samples in N’Gabakoro or the laboratory samples, and 5 samples only had reads that align in BLAST nr to *Anopheles* spp. ITS region sequence with no representation in the fungal UNITE database. Due to this, no comparisons were made between season or location on the fungal microbiome. All *Anopheles* ITS reads (63332/79195 total) were filtered. After filtering, the most abundant fungal genera were *Aspergillus* (20.7% of total reads without *Anopheles*) with presence in 12/21 samples with fungal ITS reads followed by *Malassezia*, *Cladosporium*, and *Phoma* (15/21, 13/21, 9/21 of samples, respectively, Fig. S1).

**Characteristics of bacterial communities between sampling locations:**

The most abundant genera across field-caught mosquitoes was *Ralstonia* (25.1% of all field reads, in 42/47 samples), and the most prevalent genera was *Propionibacterium* (46/47 samples) (Fig. 2). Laboratory samples were largely dominated by reads to *Asaia* (62.9% of all laboratory reads), with this genus being the most prevalent taxa in 5/6 samples. There was no significant difference in the Shannon diversity between groups of field samples via Kruskal-
Wallis chi-squared with Dunn’s multiple comparisons test, though the mean diversity of the laboratory samples was significantly lower than all field groups (Fig. S2A-C). While overall the dry season had the highest diversity, this difference was not significantly higher than the wet season overall (dry: 2.54, wet: 2.03, $p=0.087$, Fig. S2B). The mean number of sequence variants per sample varied from 42.5-65.13 in the field to 20 in the laboratory, with dry season samples from both field locations being significantly different than laboratory samples ($p=0.0117, 0.0066$ for Thierola Area and N’Gabakoro, respectively). The mean number of genera were also higher in the field than in the laboratory (27.36-42.13 in field, 14.83 in lab, Fig. S3B), again with the dry seasons being significantly higher than the laboratory samples ($p=0.0215, 0.0119$ for Thierola and N’Gabakoro, respectively). The Pearson correlation between the overall genus abundance and frequency (across mosquitoes) was moderate ($r=0.42$ and $r=0.70$, $P<0.001$, $n=424$ for native and log-transformed values), indicating that the genera more frequent across mosquitoes were more abundant (sequence reads). However, the majority of sequence variants and genera were sample singletons (91.9% and 43.7%, respectively), however the majority of reads were found in non-singletons (68.0% of sequence variant reads, 98.8% of genera-level reads) (Figs. S4A and S4B). This would indicate there remains a high degree of heterogeneity between samples, though dominant species have some conservation between groups. Finally, there were higher mean amounts of group specific (private) sequence variants within the Sahel dry season (53.6 compared to 35.1, 47.5, 31.2, and 12.7 for Sahel Dry, Sahel Wet, Riparian Dry, Riparian Wet, and laboratory samples, respectively), though this difference was only significantly different between the laboratory samples and the two dry seasons ($p = 0.0173$ and 0.0048 for Riparian dry and Sahel dry via Dunn’s test, respectively).

**Differential abundance testing of seasonality:**
We tested differences in genera abundance between seasons at each location, and between field and laboratory samples. In Sahelian locations near Thierola, five bacterial species were found to be abundant at an un-adjusted $p$-value of $< 0.05$ with *Anaplasma*, *Bacillus*, *Arthrobacter* spp. being more abundant in the dry season individual samples (Fig. S5A, log$_2$ fold-change: 4.79, 3.89, and 3.01), and *Acinetobacter* and *Brevundimonas* spp. being less prevalent in the dry season samples (Fig. S5A, log$_2$ fold-change: -3.48 and -3.11). In riparian site (N’Gabakoro), we found that reads of *Ralstonia* spp. was significantly reduced in mosquitoes (Fig. S5B, log$_2$ fold-change: -5.10, Benjamini-Hochberg adjusted $p$-value: $3.33 \times 10^{-4}$). As with the Sahelian group, the Riparian dry season had some bacterial species that did not pass the false discovery rate adjusted $p$-value cutoff (*Ideonella* and *Enterobacter*). Comparing all field to all laboratory samples found six genera to be differentially expressed at $p$-adj $< 0.05$, with the laboratory samples being biased towards *Asaia* genera, and having the presence of *Gluconobacter* bacteria that were absent from all field samples (Fig. S5C).

In effort to improve discriminative power, additional analysis was performed using hierarchical multiple testing on sequence variants to determine which SVs were differentially abundant between seasons and locations. For the Thierola area, only one *Anaplasma* sequence variants was found to be differentially expressed in the dry season compared to wet (log$_2$ fold-change: 3.50, $p$-adj: 0.027, Fig. 3A). Seven sequence variants were found to be differentially expressed between seasons in the permanent water location, with one more abundant in the wet season (Fig. 3B, *Ralstonia*, log$_2$ fold-change: -4.91, $p$-adj: $1.4 \times 10^{-3}$), and six more abundant in the dry (*Sphingorhabdus*, two *Duganella* SVs, *Janthinobacterium*, *Xenophilus*, *Pseudomonas*, and a *Cyanobacteria*; log$_2$ fold-change 1.63-3.94, $p$-adj: $3.19 \times 10^{-2}$ to $3.91 \times 10^{-5}$). In the laboratory versus field comparison, six sequence variants were found to be significant (Fig. 3C, *Asaia*,...
Elizabethkingia, two Gluconobacter SVs, Alcaligenes, and Ralstonia, log₂ fold-change 2.97-4.11, p-adj: 1.4e⁻³ to 5.61e⁻¹⁵).

Seasonal and temporal discrimination using ordination:

After rarefaction to a depth of 500 reads and filtering of sequence variants present in only a single sample, 164 sequence variants and 48 samples were used for ordination analysis. Ordination of the samples via principal coordinates analysis (PCoA) on the Bray-Curtis dissimilarity distance showed clustering of the five seasonal and location based groupings, with differing levels of between-group spread (Fig. 4A). Testing of the group centroid locations via a pairwise PERMANOVA found that the group composition of laboratory and ‘permdry’ samples differed from all other groups, while ‘saheldry’ differed from ‘permwet’ and ‘sahelwet’ at an adjusted p-value < 0.10 (Fig. 4B).

Qualitative presence/absence analysis:

As abundance testing found limited seasonal differences for each location, we also investigated whether supervised learning approaches could discriminate dry vs. wet season samples based on the differences in presence/absence of bacteria between groups. We again found that the seasonal samples from the permanent water clustered separately more strongly than the Sahelian locations into two populations (Fig. S6A), though all groups other than the wet season locations were found to have significantly different center points via PERMANOVA (Fig. S6C).

To further compare the nature of the uniqueness present in the microbiome of possibly aestivating mosquitoes in the Sahelian dry season, we employed sequential testing between groups. Firstly, we tested for genera that showed significant differences between the dry and wet seasons in the Sahel. Secondly, we tested for differences between the Sahelian and Riparian dry
season, within the subset of genera identified earlier. Thirdly, we tested for differences between
the Sahelian dry and Riparian wet season within the subset that passed the first tests. In each
step, we employed exact tests that accommodated sample size at the individual test level (the
second and third tests were one-sided, as dictated by the direction of the difference in the first
test. see methods). A total of 14 genera exhibited difference between the Sahelian dry and wet
season at P<0.1 (with highest significance for Bacillus, Anaplasma, and Microvirga (P<0.00004,
<0.0026, 0.007, respectively). In the subsequent tests, only Anaplasma which was exclusively
present in the Sahelian dry season (54% vs. 0%, p=0.009), Bacillus (90.9% vs. 7-38%, p=0.020),
Intestinibacter (45.4% vs. 0-7%, p = 0.022) and Microvirga (63.6% vs. 7-14%, p=0.029) were
considered putatively “characteristic” of the Sahelian dry season.

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

With the DADA2 denoising algorithm, sequencing errors can be parsed from true
sequence variants in the 16S rRNA sequenced region. For interpretability in display, the 16S SVs
and OTUs for the top 19 most prevalent genera were placed on a phylogenetic tree to
demonstrate the loss in information from clustering sequences with >97% nucleic acid identity
compared to maintaining variants with as little as one nucleotide difference (Figs. S7A and S7B).
There were 35.8% more SVs than OTUs across all genera (2057 and 1515, respectively), with
79.7% more SVs than OTUs in the top 19 most abundant genera.

Cytochrome B analysis of blood feeding host preference:

The Anaplasma spp. reads found to be more prevalent during the dry season were blasted
against the nr database were found to align to Anaplasma ovis (99-100% identity based on
strain), a pathogen of goats, sheep, and wild ruminants (de la Fuente et al., 2007). To define the
source of these reads we performed polymerase chain reaction to determine bloodmeal origin
and the rate of anthropophily seasonally and by location. We found that areas of seasonal water
availability had a slight decrease in the degree of anthropophily overall (Fig. 5 left side, 77.5% to
94.6%, comparing “Human” to all others, \( p=0.0490 \) with Two-tailed Fisher’s exact test).
Additionally, the Thierola area had a statistically significantly higher proportion of zoophily in
the dry season compared to the dry season in N’Gabakoro (28.6% mixed or non-human blood-
feeding compared to 0%, \( p=0.0318 \)).

Discussion:
The mosquito microbiome varies considerably throughout its development (Wang et al.,
2011; Coon et al., 2014), and has been previously shown to have distinct bacterial characteristics
due to its location and seasonal environment (Akorli et al., 2016; Buck et al., 2016). In this
study, we analyzed how the microbiome changes in \textit{An. coluzzii} collected from areas with
differing degrees of seasonality in Mali. This variation in seasonality in our sampling locations
was due to several factors. The first is that due to being lower in latitude, the areas around
N’Gabakoro do not dry out as rapidly as the more northerly areas neared to the Sahel.
Additionally, the presence of the Niger River provides the possibility for year round
development of larvae in the Southern area that is likely not possible in the North. Our main
hypothesis was that large seasonal differences between Sahelian dry and wet season will be
present due to the unique physiological state of Sahelian \textit{An. coluzzii} mosquitoes during
aestivation (Huestis et al., 2012; Yaro et al., 2012; Dao et al., 2014; Huestis & Lehmann, 2014).
If confirmed, aestivation-specific microbiome taxa might be used as a predictor of this state in
populations where both aestivators and reproductive adults coexist.

Counter to our initial hypothesis that dry season Sahelian mosquitoes would have the
most distinct microbiome composition, we found that dry season mosquitoes from areas with
year-round breeding (N’Gabakoro) were, overall, the most different among our mosquito groups (below). The Sahelian dry season did exhibit some unique characteristics including high microbial diversity with highest number of private sequence variants, and had slight differences in overall microbial composition (Figs. 3), though these differences were not as pronounced as those between seasons in permanent water locations. Additionally, several bacterial genera were significantly elevated in the Sahelian DS including *Anaplasma* (exclusively present in the Sahelian DS), *Bacillus* (90.9% in Sahel dry vs. 7-37.5% in others), *Microvirga* (63.6% vs. 7-14%), and *Intestinibacter* (45.4% vs. 0-7%). We also note that three of these genera (*Anaplasma, Bacillus, and Microvirga*) were also found to be important to the random forest classification model (Figs. S6B and S6D). Only *Anaplasma* was found to be significantly different based on read count abundance (log$_2$ fold-change: 3.50, p-adj: 0.027, Fig. 3). The putative route of acquisition of *Anaplasma* is described in depth below, but the most likely route of acquisition of *Bacillus, Microvirga, and Intestinibacter* would be through the larval environment or possibly plant feeding due to their soil/water association (Table S1). As no private genera were present in all Sahelian dry season mosquitoes, if bacterial biomarkers of aestivation exist, it may indicate not all mosquitoes in this period are in this state. Independent studies are necessary to test whether these putative differences are indeed “characteristic” of the Sahelian dry season and has relevance to aestivation.

The riparian dry season sample was characterized by a reduction of *Ralstonia* species reads, with an increase in *Duganella, Janthinobacterium, and Sphingomonas spp.* reads as determined by hierarchical multiple testing, DESeq2, and ordination (Figs 3, 4, and Fig. S5B). The most abundant genera was *Ralstonia* that was present in most wild-caught samples (present in 40/47, dominant taxa in 24/47), and had little representation in laboratory mosquitoes (1/6...
samples with 10 reads). This genus has been seen previously in wild-caught *Anopheles* from Cameroon (Boissière et al., 2012) and *Aedes aegypti* (Apte-Deshpande et al., 2012), and is believed to be largely soil and water associated (Gilligan et al., 2003; Ryan, Pembroke & Adley, 2006). Additionally, we found that the majority of 16S sequence variants were unique to an individual sample (91.9%, Fig. S4), indicating that there is a high degree of heterogeneity between mosquitoes. However these sequence variants were relatively low in abundance, only accounting for 32.0% of reads. These sequence variants also demonstrate the additional subspecies level heterogeneity that the more commonly used operational taxonomic units (OTUs) may miss (Fig. S6).

As it has been previously reported in *Aedes aegypti* that most of the adult gut flora is acquired trans-stadially from larval stages to adults (Coon et al., 2014), this may help explain why there were limited seasonal differences in mosquitoes in the northern Sahel. Due to the prolonged dry season, there are no known larval sites available during this period (Dao et al., 2014). Thus, the mosquitoes present would have been larvae in the same rainfall-linked transient water sources that the wet season mosquitoes would have, and any changes present in the microbiome would be due to incorporation of new flora from environmental conditions, or blood/sugar sources. This homogeneity in larval environments is contrasted in the available larval habitat near N’Gabakoro that changes broadly throughout the year, from fresh rain puddles to ground and river water in rock and other pools near the Niger River that develop as flooded areas recede (Edillo et al., 2002; Coulibaly et al., 2007), or from standing water present in the more urban area of Bamako nearby (though these pools are likely unsuitable for *Anopheles* spp. growth). This hypothesis may explain why the dry season Sahelian samples had the most 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*
(Favia et al., 2007; Baldini et al., 2014; Buck et al., 2016). However, we found no Wolbachia
reads in any samples, though as these reads are difficult to amplify it may be below the limit of
detection (Baldini et al., 2014). Additionally, the presence of Gluconobacter spp. bacteria in
laboratory mosquitoes (but no field mosquitoes) may point to other ways in which the artificial
nature of the laboratory environment (i.e. sugar feeding on Karo syrup) could impact the
microbial composition, and possibly additional metrics such as vector competence.

A caveat to our microbiome composition data is that the sampling and next-generation
sequencing in this study was performed prior to the knowledge that there can be contaminant
“kitome” DNA present from DNA extraction kits, molecular water, and cross-over
contamination on machines (Salter et al., 2014). Due to this, we do not have control samples
from each of these items, and due to the time since processing, these kits are no longer available
to be sequenced. The issues of contamination have been reported to be more severe with samples
of low-bacterial abundance (Salter et al., 2014). We hope that in our samples this contamination
is limited due to whole-body sampling which would increase bacterial reads, and through finding
that there were no significant negative correlations between read count and presence of known
kitome genera (two-tailed Pearson’s r, Table S2). This lack of correlation should not imply there
is no contamination possibility in our samples, but is the best assessment we can provide
retrospectively that presence of these genera is not linked to low-biomass in our samples.

Additionally, no reads were broadly found in all samples, and only the laboratory samples were
collected in different time periods than the others. Future studies should include these controls,
and assessment of aquatic bacterial species from the larval sites to limit these possible
confounders. Additional use of newer versions of next-generation sequencing approaches could
also help to improve sequencing coverage, allowing for more efficient filtering of these low prevalence reads.

The presence of *Anaplasma ovis* bacteria in the Sahelian dry season samples was unexpected, but has some precedence in the literature. *A. ovis* 16S sequence has been found previously in *Anopheles gambiae* and *An. funestus* collected in western Kenya (Lindh et al., 2005), and it remains unknown if this finding has any relevance to the disease’s transmission or is simply related to it being an intracellular pathogen of the blood that could be inadvertently picked up during feeding. *Anaplasma phagocytophilum* has been shown to modulate *Ixodes* tick microbiota, though the effects of *Anaplasma* species on mosquito microbiota has not been studied to our knowledge (Abraham et al., 2017). This pathogen is also in the same order as the *Wolbachia* genus bacteria that have achieved considerable attention as an *Aedine* symbiont and possible vectorial-competence modifying species (Cirimotich, Ramirez & Dimopoulos, 2011; Hughes et al., 2014; Minard et al., 2014).

To expand upon this finding of a zoonotic pathogen reads present in these classically anthropophilic mosquitoes (Besansky, Hill & Costantini, 2004), the analysis of host-choice PCR with cytochrome B allowed us to refine the biting characteristics of each population across seasons. We found that the relaxation of strict anthropophily in the seasonal areas (Fig. 5) which may follow what has been reported previously in areas of low host-availability (Lefèvre et al., 2009). Due to the climactic severity during the dry season in the Sahel, the acquisition of blood from the nearest source is likely less taxing than feeding on the preferred host. Though as there is not an increase in zoophily in the N’Gabakoro samples in March-April in the dry season (when conditions are similar between locations) and there is some zoophily in the wet season, there may be an innate degree of zoophily in all *An. coluzzii* even in favorable conditions.
Conclusions:

This study of the seasonal microbiome of *An. coluzzii* indicates some unique aspects that illuminate some behavioral characteristics during this cryptic and poorly understood time period. It is largely characterized by mosquitoes in areas of higher seasonality exhibiting a reduction in anthropophily, likely due to the unique environmental pressures seen in these areas, and a higher overall diversity during the dry season regardless of the presence of permanent water. Furthermore, it appears that the mosquito larval environment may be performing a seeding function for the adult microbiome, and suggests minor evidence for the occurrence of aestivation. Future analysis of the seasonality of host-choice could help to refine these findings, and to show possible impacts towards transmission.

Acknowledgements:

We thank the residents of M’Piabougou, Zanga, and N’Gabakoro for their assistance and permission to collect samples in/around their houses. We thank Dr. Shifra Goldenberg for her provision of code for NDVI analysis.

References:


Apte-Deshpande A., Paingankar M., Gokhale MD., Deobagkar DN. 2012. Serratia odorifera a
midgut inhabitant of aedes aegypti mosquito enhances its susceptibility to dengue-2 virus.

PLoS ONE 7. DOI: 10.1371/journal.pone.0040401.


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


Hervé M. 2017. RVAideMemoire: Diverse Basic Statistical and Graphical Functions.

Hijmans RJ. 2016. raster: Geographic Data Analysis and Modeling.

Huestis DL., Lehmann T. 2014. Ecophysiology of Anopheles gambiae s.l.: Persistence in the
Huestis DL., Yaro AS., Traoré AI., Dieter KL., Nwagbara JI., Bowie AC., Adamou A.,
Kassogué Y., Diallo M., Timbiné S., Dao A., Lehmann T. 2012. Seasonal variation in
metabolic rate, flight activity and body size of Anopheles gambiae in the Sahel. *The Journal

Hughes GL., Dodson BL., Johnson RM., Murdock CC., Tsujimoto H., Suzuki Y., Patt AA., Cui
microbiome impedes vertical transmission of Wolbachia in Anopheles mosquitoes.
*Proceedings of the National Academy of Sciences* 111:12498–12503. DOI:
10.1073/pnas.1408888111.

Jupatanakul N., Sim S., Dimopoulos G. 2014. The insect microbiome modulates vector

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

Kuhn M., Contributions:, Weston S., Williams A., Keefer C., Engelhardt A., Cooper T., Mayer
Z., Kenkel B., Team the RC., Benesty M., Lescarbeau R., Ziem A., Scrucca L., Tang Y.,

de la Fuente J., Atkinson MW., Naranjo V., Fernández de Mera IG., Mangold AJ., Keating KA.,
*Veterinary Microbiology* 119:375–381. DOI: 10.1016/j.vetmic.2006.09.011.

Lefèvre T., Gouagna LC., Dabiré KR., Elguero E., Fontenille D., Renaud F., Costantini C.,
behavior of Anopheles gambiae s.s. when humans are not readily accessible. *American
Journal of Tropical Medicine and Hygiene* 81:1023–1029. DOI: 10.4269/ajtmh.2009.09-
0124.


Lindh JM., Lindh JM., Terenius O., Terenius O., Faye I., Faye I. 2005. 16S rRNA Gene-Based
Identification of Midgut Bacteria from Field-Caught *Anopheles gambiae* Sensu Lato and *A.
funestus* Mosquitoes Reveals New Species Related to Known Insect Symbionts. *Applied


Love MI., Huber W., Anders S. 2014. Moderated estimation of fold change and dispersion for

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

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

Minard G., Tran FH., Raharimalala FN., Hellard E., Ravelonandro P., Mavingui P., Valiente
Moro C. 2013. Prevalence, genomic and metabolic profiles of Acinetobacter and Asaia
associated with field-caught *Aedes albopictus* from Madagascar. *FEMS Microbiology

Oksanen J., Blanchet FG., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin PR.,
O’Hara RB., Simpson GL., Solymos P., Stevens MHH., Szoebs E., Wagner H. 2017. vegan:
Community Ecology Package.


Perpiñán O., Hijmans R. 2016. rasterVis.


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


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


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


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.
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**

- **Genus**
  - Anaplasma

- **Phylum**
  - Proteobacteria

B. **N’Gabakoro Wet to Dry Season**

- **Genus**
  - Ralstonia
  - Pseudomonas
  - Duganella.1
  - Duganella
  - p_Cyanobacteria
  - Janthinobacterium
  - Sphingorhabdus
  - Xenophilus

- **Phylum**
  - Proteobacteria
  - Cyanobacteria

C. **Lab to Field**

- **Genus**
  - Asaia
  - Elizabethkingia
  - Gluconobacter.1
  - Gluconobacter
  - Alcaligenes
  - Ralstonia

- **Phylum**
  - Proteobacteria
  - Bacteroidetes
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

A plot showing the distribution of samples across different seasons and locations. The plot is labeled with the following categories:

- **SeasonLoc**
  - Lab
  - PermDry
  - PermWet
  - SahelDry
  - SahelWet

- **Area**
  - Lab
  - N’Gabakoro
  - Thierola

### B

<table>
<thead>
<tr>
<th></th>
<th>LAB</th>
<th>PERMDRY</th>
<th>PERMWET</th>
<th>SAHELDRY</th>
<th>SAHELWET</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERMDRY</td>
<td>3.43E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERMWET</td>
<td>2.50E-04</td>
<td>6.67E-04</td>
<td></td>
<td>0.019</td>
<td>0.296</td>
</tr>
<tr>
<td>SAHELDRY</td>
<td>2.50E-04</td>
<td>6.00E-04</td>
<td>0.296</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>SAHELWET</td>
<td>2.50E-04</td>
<td>2.50E-04</td>
<td>0.296</td>
<td>0.092</td>
<td></td>
</tr>
</tbody>
</table>
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.