**Metagenomic Comparison between the Gut Microbiota**

**of Yellow Fever Mosquitos *(Aedes aegypti)***

**San Jose State University**

Uyen Hoang ([uyen.hoang@sjsu.edu](mailto:uyen.hoang@sjsu.edu))

Ngoc Lu (​​[ngoc.lu@sjsu.edu](mailto:ngoc.lu@sjsu.edu))

Fabiola Briseno ([fabiola.briseno@sjsu.edu](mailto:fabiola.briseno@sjsu.edu))

Dang Minh Nhu Nguyen ([dang.m.nguyen@sjsu.edu](mailto:dang.m.nguyen@sjsu.edu))

**CS123B: Bioinformatics**

**Dr. Wendy Lee**

**November 30, 2021**

**Table of Contents**

[**Introduction:**](#_pfl9jwe1a7fn) **2**

[**Methods and Materials:**](#_fborew98lt8a) **3**

[Methods:](#_d6wbcmkcvl1i) 3

[Materials:](#_j4k5fykv4c42) 6

[**Results and Analysis:**](#_xaavtqbf8mfp) **7**

[Results:](#_2tp15nkso10k) 7

[Analysis:](#_ki3z2v8dhqd) 8

[**Discussion:**](#_4ht807qu9pj5) **10**

[**Appendix**](#_4t59p08dcvqr) **12**

[**References**](#_km9wij5epgkr) **14**

# Introduction:

Mosquitos are known as one of the most efficient biological vectors for transmitting deadly human pathogens, including: Malaria, Dengue, Yellow Fever, Chikungunya, Zika, and West Nile. More than 700,000 people die from mosquito-borne diseases and millions are infected every year [1]. In an effort to control these vector-borne diseases, traditional methods use chemical insecticides, yet these practices often have negative effects on the environment. Another issue with pesticides is the development of resistance within the mosquito population. Recently, many researches have shown that mosquito microbiota hold a great potential for reducing transmission of mosquito-borne pathogens. Mosquito microbiota are the microorganisms found in the mosquito’s midgut contributing to the processes of food digestion and egg production in female mosquitos. In this project, we focus on the midgut microbiota of Yellow Fever mosquitoes, *Aedes aegypti.* Previous studies indicated that the food source, especially blood-meal source, affects the microbial composition in the female mosquitos’ midgut. The effect of midgut microbiota on mosquito parasite infection is complicated. Some of them were found to contribute in defending the mosquito against viral infection [2][3]; others facilitated the arboviral infection [4].

Thus, investigating the mosquito midgut microbiome, or the combined genome of microorganisms in the mosquito, allows us to alternatively reduce the transmission of blood-borne pathogens. Our goal is to obtain data from NCBI (SRA Bioproject PRJNA494958) produced using next-generation sequencing platform, followed by 16S metagenomic alignment and comparison using the Galaxy (<https://usegalaxy.org/>) web-based platform for bioinformatic analysis to reproduce the characterization of *A. aegypti* midgut microbiota and examine the effect of food sources on the composityion of microorganisms within the mosquito’s digestive tracts.Based on the data obtained, we will analyse the results obtained and use the data obtained to build a Krona pie chart to visualize our results. We hypothesize that the mosquitos with the most similar dietary restrictions will have the highest similarity in microorganism composition in their digestive tracts. We propose that this is due to only certain microorganisms being able to thrive in the digestive tracts when certain food sources are available.

Previously, several studies had revealed that the independent laboratory-specific conditions may have significant influence on the Yellow Fever Mosquito gut microbiomes [5]. In order to determine the core microbiome taxa involved in the blood-borne diseases transmission processes, it’s crucial to improve the characterization of mosquito gut bacterial communities. Thus, identifying the conserved bacterial taxa could reveal some critical characteristics of mosquito biology that drive the global spread of mosquito-borne pathogens. For future studies, genetic engineering could be used to modify certain symbiotic bacteria to produce antipathogenic agents.

# Methods and Materials:

## Methods:

First, we needed to build small databases using the Run selector from NCBI SRA. The database will consist of five 16S rRNA sequences collected from an independent laboratory-based mosquito Bioproject conducted by Frankel-Bricker et al [5]. To download the SRA accession lists, we utilized the Run Selector Feature on the NCBI SRA website and uploaded our data to Galaxy using the “Upload Data” window. We adapted our protocol from the “BIOL/CS123B Metagenomics Data Analysis” by Wendy Lee and Cleber Ouverney (2020) to identify the similarities and differences between the gut microbiota of Yellow Fever Mosquitos by building a Krona Pie Chart using Galaxy [7].

After uploading, we used the **Faster Download and Extract Reads in FastQC** function on Galaxy to obtain the Fastq files. Since our data is pair-end data, each sample consists of two separate fastq files: one containing the forward reads and one containing the reverse reads. We then performed quality control of our data using the **Make.contigs** tool, which will combine the forward and reverse reads for each sample, and also combine the resulting contigs from all samples into a single file. In this project, paired-end sequencing of the ~253 bp V4 region of the 16S rRNA gene was performed from either end of each fragment. Because the reads are about 250 bp in length, this results in a significant overlap between the forward and reverse reads in each pair. We now have a combined paired collection of 10x2 or 5x2 FASTQ files to a single FASTA file. The next step is data cleaning with the purpose of improving the quality of our data. To do this, we first need a statistics summary of the output file by using the *trim.contigs.fasta* file that was generated by **Make.contigs** to run **Screen.seqs** which will filter reads based on quality and length. Specifically, this cleaning tool will remove sequences with ambiguous bases (*maxambig*) and contigs longer than a given threshold (*maxlength*). Since the V3-V4 is about 443 base pairs, we will be using the *maxlength* as “470” and *maxambig* as “0”. Furthermore, to refine the unique sequences in our sample files, we will be running the **Unique.seqs** tool as that not only reduces the duplicated reads but also helps us find unique sequences. Due to the numerous organism sequences included in our samples, there may be many identical sequences, so it is critical to only analyze the unique sequences. Then, to speed up the computational job as well as reduce file sizes for analysis, we use the **Count.seqs** tool to summarize the number of duplicates of each sequence observed across our samples.

Next, we performed alignment to a reference database file using **Align.seqs** and the Silva.123 reference database. The output file generated from the align tools can be used to make a summary log file using **Summary.seqs** to review our analysis so far. To ensure all our reads overlap our region of interest, we need to conduct post-alignment data cleaning. This step will remove any reads not overlapping the region V4 region and any overhang sequences on either end of the V4 region. The three tools, **Screen.Seqs, Filter.seqs** and **Unique.Seqs,** were performed again to clean the data. Afterwards, we will use **Pre.cluster** to merge near-identical sequences into one sequence. Because our contigs are ~442 bp long, we will set the threshold to “4” mismatches and homopolymer to “8”. Since we know that our reference database does not contain any homopolymer stretches longer than 8 bases, any reads containing such long stretches are likely the result of technical errors in the lab. Next, we will need to remove a class of sequencing artifacts known as *chimeras*, which occurs when the polymerase falls off one sequence and then reattaches to another during a single round of PCR. This results in a seemingly unique sequence, but is, in fact, two different sequences merged together. Chimera removal will be performed using the VSEARCH algorithm within mothur which is called the **Chimera.vsearch** tool. This tool will split the data by sample and check for chimeras using the abundant sequences as our reference. However, this only stores the *chimeric* sequences in a count table file, so we will need to physically remove them from the fasta file using **Remove.seqs.**

To further clean our sequences, we used **Classify.seqs** and **Remove.lineage** to remove non-bacterial sequences, such as organisms belonging to Archaea, chloroplasts, and mitochondria. Running Classify.seqs involves using the count table generated by Remove.seqs and the reference data that we loaded into this project.

In this project, we use a 16S metagenomics approach which requires OTUs clustering of similar sequence variants of the 16S rRNA gene sequence, aiming to cluster taxonomic units of bacterial species or genus depending on the sequence similarity threshold. Cluster.split was used to perform clustering by assigning a “0.03” cutoff and a tax-level set to “4.” To determine how many sequences within each OTU, we utilized the **Make.shared** tool and afterwards, differentiated each taxonomic level with **Classify.otu**. We finally normalized performing subsampling, utilizing the **count.group** and **sub.samples** tools. The last step is visualizing the composition of our data using **Krona**. The resulting file will contain an interactive Krona pie chart that helps us compare results for each type of sample.

## Materials:

In order to examine the effect of food sources on the composition of microorganisms within the Yellow Fever mosquitos’ midgut, we first required data to analyze which we found in a study performed by Frankel-Brickel J. and Frankel L. K [5]. In their study, the scientists collected adult mosquitos (Gainesiville strain) immediately after hatching from pupaes and extracted their microbial RNA samples. Frankel-Bricker et al. utilized five different experimental conditions to perform their analysis, obtaining samples from yellow fever mosquitos raised with human blood, sugar, chicken blood, rabbit blood, and mosquitos that had been unfed as a negative control. They obtained a total of 148 samples that were sequenced by an Illumina MiSeq instrument using paired-end sequencing. This data is publicly available in the NCBI SRA Bioproject PRJNA541017.

In order to perform their metagenomic analysis, Frankel-Bricker et al. designed sequence specific primers to amplify the 16s V3-V4 region of the small ribosomal subunit. As a result, our analysis needed to similarly target the 16S V3-V4 region by utilizing a reliable reference alignment file. After reading a study by Bukin, Y.S. et al, we selected the Silva.123 database which was released on July 23, 2015 as it was similarly used by the scientists to align their 16S V3-V4 amplified RNA samples [6]. The Silva.123 was downloaded from the SILVA 123 online database (<http://arb-silva.de>). Furthermore, to perform taxonomic classification of our sequences, we also required two taxonomy data files which we obtained from the “BIOL/CS123B Metagenomics Data Analysis” website created by Wendy Lee and Cleber Ouverney [7].

Rather than performing the analysis on all 148 samples from the Frankel-Brickel study, we performed separate metagenomic analyses using between 5-10 random samples from each condition, as seen in *Table 1*. This was purposely done because when attempting to perform the metagenomic analysis on all 219 samples, we continuously ran out of space on our Galaxy servers. By reducing the number of samples we analyzed, we were able to perform all five separate metagenomic analyses in order to compare the composition of microorganisms within the Yellow Fever mosquitos’ digestive tracts when raised on different food sources.

# Results and Analysis:

## Results:

As previously stated, to elucidate the effect of different food sources on the composition of microorganisms within the mosquito’s digestive tracts, we ran five separate 16S metagenomic analyses which generated 5 Krona Pie Charts for each feeding condition. The main core bacterial phyla between the five conditions were *Bacteroidetes* and *Proteobacteria*, which accounted for 49% - 96% of the gut microorganism populations. Particularly, mosquitos raised on chicken blood exhibited the least bacterial diversity with 96% of its gut bacterial population belonging to the *Bacteroidetes* Phylum, specifically *Chryseobacterium* and *Elizabethkingia* as observed in *Figure 2*. The remaining 4% of the bacterial composition in these mosquitoes belonged to Proteobacteria (2%) and Actinobacteria (2%). Similarly, we found that mosquitos raised on human blood also had low bacterial diversity, but not as low as the mosquitoes raised on chicken blood. The mosquitos raised on human blood had 91% of its gut bacterial population composed of *Proteobacteria*, specifically *Serratia*, which belongs to the *Chryseobacterium* genus as observed in *Figure 3*. The other two significant groups that composed these mosquitos’ gut bacterial population were *Bacteroidetes* (7%), specifically *Chryseobacterium*, and others (2%), which included *Actinobacteria* and *Acidobacteria*. Moreover, the mosquitos raised on rabbit blood had 82% of its bacterial gut population composed of Bacteroidetes, specifically *Elizabethkingia*, which belong to the *Bacteroidetes* phyla as observed in *Figure 2*. The mosquitoes also had 17% of its bacterial gut diversity belonging to *Proteobacteria* (i.e. *Gammaproteobacteria*) and 1% of others, including *Pantoea* and *Enterobacteriaceae*. On the other hand, mosquitoes raised on a sugar diet or were unfed had the highest amount of bacterial diversity as observed in *Figure 1*. Most notably, mosquitos fed with sugar had 70% of its gut bacterial composition belonging to *Actinobacteria*, specifically *Nocardioides*, as seen in Figure 1. These mosquitoes also exhibited a significant portion of its bacterial gut composition belonging to *Proteobacteria*, specifically *Asaia*, at 30%, and less than 0.5% composed of *Clostridia* and *Bacteroidetes*. Furthermore, the mosquitos that were unfed had an even higher bacterial diversity with 49% *Actinomycetales* (i.e. *Nocardioides* and *Corynebacterium* ), 29% *Firmicutes* (i.e. *Streptococcus*), 18% *Proteobacteria* (i.e. *Neisseria*), and 4% *Bacteroidetes* (i.e. *Flavobacterium*) as seen in *Figure 1*.

## Analysis:

To determine whether we can accept or reject our hypothesis that the mosquitoes with the most similar dietary restrictions will have the highest similarity in microorganism composition in their digestive tracts, we need to draw a comparison between our results. Most notably, the mosquitoes fed with chicken and rabbit blood both had *Bacteroidetes* (96% and 82% respectively)as the majority of the bacterial gut composition which supported our hypothesis. Also, these two mosquito populations had *Proteobacteria* (2% and 17% respectively) as their second most type of bacteria, further proving our hypothesis.

However, contrary to our hypothesis we found that mosquitoes fed with human blood had a completely different gut bacterial composition in comparison to the mosquitoes fed with rabbit and chicken blood. In fact, the majority of the gut bacterial composition was composed of *Proteobacteria* (91%) while in the chicken and rabbit blood fed mosquitoes the majority was *Bacteroidetes*. In the human blood fed mosquitoes, *Bacteroidetes* was only 7% of its gut bacterial composition, opposite to the other blood fed mosquitoes. These results are interesting because both of these diets contained blood, yet the mosquitos’ gut bacterial compositions were opposite. This could indicate that a component or substance present in human blood may be preventing the *Bacteroidetes*, which were previously abundant in the chicken and rabbit blood fed, from thriving. However, further research must be conducted in order to determine this.

Furthermore, the differences between non-blood fed mosquitoes and blood fed mosquitoes also supported our results as they resulted in varying bacterial growth in the digestive tracts of these mosquitos. Specifically, in the Sugar-fed and unfed mosquitoes, we see a wider diversity of bacterial organisms, but in the blood-fed mosquitoes, we mainly observe a single type of bacteria being more abundant than the rest. In the Sugar-fed and unfed mosquitoes, the most abundant bacteria are Actinobacteria (70%) and Actinomycetales (49%) respectively. In the blood-fed mosquitoes, we do not see these bacteria which could suggest that feeding on blood may eliminate these organisms. Particularly, in the unfed mosquitos, we see *Firmicutes*, such as *Streptococcus* and *Lactobacilli*, present in the mosquitos’ guts which is strikingly similar to the bacteria present in human digestive tracts. However, the presence of *Firmicutes* are greatly reduced when the mosquitoes are fed with sugar.

Of note, we did use more sequences in our analysis of the unfed and sugar-fed mosquitoes, so that may contribute to the wider bacterial diversity that is observed. Nonetheless, our results lead us to accept our hypothesis that the mosquitos with the most similar dietary restrictions will have the highest similarity in microorganism composition in their digestive tracts.

# Discussion:

The midgut’s microbial communities of anautogenous insects such as mosquitoes support many crucial functions for the host, and their diversity can be strongly influenced by both genetic and environmental factors. In this experiment, we’re investigating the effect of different food sources on the composition of bacterial communities in the *Aedes Aegypti* mosquitoes. In order to achieve this, we compared the gut microbiota of adult female mosquitoes that were fed on sugar, human blood, chicken blood, rabbit blood, and unfed mosquitoes. In the results, we hope to detect some core microbiota that may affect the host physiology. We also hypothesized that mosquitoes that fed on different food sources with high similarities such as rabbit and human blood will resemble each other’s gut microorganism composition. Interestingly,we found a high similarity between the microbiota composition in mosquitoes fed on chicken blood and rabbit blood. The metagenomic visualization from Krona pie charts indicated that Chryseobacterium accounts for a large portion of the bacterial population in these female mosquitoes. Since Chryseobacterium plays an important role in supporting the egg production process in *A. aegypti* [8], targeting this bacterial genus could be an effective approach for vector control. In addition, results from 16S metagenomic analysis showed a correlation between gut microbiota in mosquitoes fed on sugar and unfed mosquitoes. In these hosts, *Actinobacteria* were the dominant microorganisms, instead of bacteroidetes in the blood-fed mosquitoes. The metabolism of carbohydrate-rich and protein-rich food sources leading to different digestive tract conditions could have affected the proliferation of some specific bacterial taxa. Most importantly, mosquitoes fed on human blood showed a high composition of gut *Proteobacteria*, a bacterial phylum contributing to viral infection in hematophagous insects. This explains how humans are the most likely to be infected by blood-borne pathogens.

Nevertheless, our analysis holds some limitations. Specifically, due to limited data storage, only 10 SRAs sequences were run for sugar, and unfed samples; 5 sequences were run for chicken fed, rabbit fed, and human fed samples. Since our sample sizes were small, our metagenomics analysis results might not present a comprehensive relationship between the microbial communities.

In conclusion, our outcomes demonstrated that the host blood meal sources possess a strong impact on the composition of bacterial communities in mosquito midgut. Different bacterial taxa may affect the mosquito’s ability to transmit blood-borne pathogens. Thus, targeting the midgut microbiota appeared to be an efficient strategy for reducing vector-borne diseases.

# Appendix

| | **Feeding Condition** | **SRA Accession Numbers** | | --- | --- | | Sugar | SRR7975705, SRR7975706, SRR7975707, SRR7975708, SRR7975709, SRR7975710, SRR7975711, SRR7975712, SRR7975713, SRR7975714 | | Unfed | SRR7975717, SRR7975718, SRR7975723, SRR7975724, SRR7975725, SRR7975726, SRR7975727, SRR7975728, SRR7975729, SRR7975730 | | Rabbit Blood | SRR7975732, SRR7975733, SRR7975734, SRR7975735, SRR7975736 | | Chicken Blood | SRR7975695, SRR7975696, SRR7975697, SRR7975698, SRR7975699 | | Human Blood | SRR7975758, SRR7975788, SRR7975789, SRR7975791, SRR7975792 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Table 1: Table depicting the SRA accession numbers used in the metagenomic analyses. While the Sugar and Unfed conditions analyzed 10 samples, Rabbit Chicken and Human blood conditions analyzed 5 samples each.* |

|  |
| --- |
| *Figure 1: Screenshots illustrating the Krona trees generated for Sugar (left) and Unfed (right) feeding conditions.* |

|  |
| --- |
| *Figure 2: Screenshots illustrating the Krona trees generated for Rabbit blood (left) and Chicken blood (right) feeding conditions.* |

|  |
| --- |
| *Figure 3: Screenshot illustrating the Krona tree generated for the human blood feeding condition.* |

# References

[1]World Health Organization. (n.d.). *Vector-borne diseases*. World Health Organization. Retrieved November 30, 2021, from https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases.

[2]Barletta, A. B., Nascimento-Silva, M. C., Talyuli, O. A., Oliveira, J. H., Pereira, L. O., Oliveira, P. L., & Sorgine, M. H. (2017). Microbiota activates the IMD pathway and limits Sindbis infection in Aedes aegypti. *Parasites & vectors*, *10*(1), 103. https://doi.org/10.1186/s13071-017-2040-9

[3] Huang, W. (2020). *Use of Microbiota to Fight Mosquito-Borne Disease*. Frontiers. https://www.frontiersin.org/articles/10.3389/fgene.2020.00196/full

[4] Bando H., Okado K., Guelbeogo W. M., Badolo A., Aonuma H., Nelson B., et al. . (2013). Intra-specific diversity of *Serratia marcescens* in Anopheles mosquito midgut defines Plasmodium transmission capacity. *Sci. Rep.* 3:1641. 10.1038/srep01641

[5] Frankel-Bricker, J. (2021, April 2). *Re-Analysis of 16S rRNA Gene Sequence Data Sets Uncovers Disparate Laboratory-Specific Microbiomes Associated with the Yellow Fever Mosquito (Aedes aegypti)*. Microbial Ecology. https://link.springer.com/article/10.1007/s00248-021-01739-2?error=cookies\_not\_supported&code=26503f03-e2ec-43f6-865a-05fb592a649e

[6] Bukin, Y., Galachyants, Y., Morozov, I. *et al.* The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data* 6, 190007 (2019). https://doi.org/10.1038/sdata.2019.7

[7] Lee, W. (2020). *METAGENOMICS DATA ANALYSIS*. Google Sites. Retrieved November 29, 2021, from https://sites.google.com/sjsu.edu/sjsu123b-metagenomics/home

[8] Huang, W., Wang, S., & Jacobs-Lorena, M. (2020). Use of Microbiota to Fight Mosquito-Borne Disease. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00196

[9] Ephantus J Muturi, Christopher Dunlap, Jose L Ramirez, Alejandro P Rooney, Chang-Hyun Kim, Host blood-meal source has a strong impact on gut microbiota of *Aedes aegypti*, *FEMS Microbiology Ecology*, Volume 95, Issue 1, January 2019, fiy213, https://doi.org/10.1093/femsec/fiy213

[10]Moreira, L. A., Iturbe-Ormaetxe, I., Jeffery, J. A., Lu, G., Pyke, A. T., Hedges, L. M., Rocha, B. C., Hall-Mendelin, S., Day, A., Riegler, M., Hugo, L. E., Johnson, K. N., Kay, B. H., McGraw, E. A., van den Hurk, A. F., Ryan, P. A., & O'Neill, S. L. (2009). A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium. *Cell*, *139*(7), 1268–1278. https://doi.org/10.1016/j.cell.2009.11.042