2020

The Utilization of Eukaryotic Cell Types in Microbiomes for Semen, Vaginal Fluid, and Menstrual Blood Identification using 18S rDNA

Ines Benaissa
Virginia Commonwealth University

The Utilization of Eukaryotic Cell Types in Microbiomes for Semen, Vaginal Fluid, and Menstrual Blood Identification using 18S rDNA

FRSC 793 – Technical Report

Date: May 5, 2020

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Forensic Science at Virginia Commonwealth University.

By

Ines Benaissa

Bachelor of Science, University of South Carolina, 2017

Director: Dr. Sarah Seashols-Williams

Graduate Program Director, VCU Department of Forensic Science

Virginia Commonwealth University

Richmond, Virginia

May 2020

Acknowledgements

I would like to thank Dr. Sarah Seashols-Williams, Dr. Baneshwar Singh, and Dr. Tomasz Arodz for their patience, guidance, and the knowledge they have allowed me to gain throughout this project. I would like to thank the members of the Singh and Seashols-Williams laboratories, with special thanks to Denise Wohlfahrt, Claire Cartozzo, and Alyssa Daniels. They helped me greatly in this research by sharing their knowledge and always being willing to teach me new things. I would like to thank my friends in the VCU Forensics program who were always so supportive and kind to me. I want to give thanks to the VCU Forensics Department for allowing me to be here and study in this field. Finally, I would like to thank my family who are the reason I am here pursuing my dream. Their support and love have helped me get through the hardest of times and I cannot thank them enough.

Abstract

The accurate identification of body fluids is vital for crime scene investigation and forensic analysis. While many body fluids can be presumptively identified, there is a lack of confirmatory testing for verification. Therefore, the purpose of this study was to determine if body fluids (menstrual blood, vaginal fluid, and semen) could be differentiated based on their eukaryotic microbial communities. The target area of interest was the V9 hyper-variable region of 18S ribosomal DNA (18S rDNA). The samples were extracted, barcoded using a duel-index strategy, and PCR amplified before undergoing high throughput sequencing using the Illumina Miseq FGx instrument. The samples were then analyzed using an open source platform called Mothur (v1.39.5) alongside the R (v3.6.3) program. Using an Analysis of Molecular Variance (AMOVA) test, the results showed that vaginal fluid and semen were significantly different from one another, but menstrual blood could not be differentiated from vaginal fluid or semen. From the relative abundance data, it was determined that the most abundant eukaryotic families in vaginal fluid were Debaryomycetaceae and Saccharomycetaceae while in semen it was Malasseziaceae. Menstrual blood appeared to show combinations of Debaryomycetaceae, Saccharomycetaceae, and Malasseziaceae. This data helped show that vaginal fluid and semen could potentially be used for the differentiation of body fluids in crime scenes. In the future, this study may be combined with the previous 16S rDNA study to develop a confirmatory test for the accurate identification of human body fluids.

Keywords: microbiome, eukaryotes, 18S rDNA, vaginal fluid, menstrual blood, semen, forensic science, next generation sequencing, Mothur

Introduction

One of the biggest challenges that a crime scene investigator or a forensic scientist can face is identifying what body fluids may have been left behind by a perpetrator or victim at a crime scene. While most body fluids have presumptive tests that aim to indicate the type of fluid present, some lack confirmatory tests or are difficult to distinguish from other fluids. For example, a white stain could potentially be vaginal fluid or semen. Presumptively, both fluids have similar reactions under an alternate light source so further testing would be required to make a correct identification. A more difficult sample to distinguish is that of menstrual blood from venous blood. Being able to identify specific body fluids could define the type of crime that took place such as a rape or homicide. Thus, the determination of the body fluids present is crucial and has a huge impact in the reconstruction of a scene.

Serological Methods

Currently, body fluids are distinguished using presumptive chemical tests, which undergo color change reactions, or physical methods. The issue with these tests is that they may produce false positives because certain enzymes are present in more than one body fluid (1). For instance, the acid phosphatase enzyme commonly seen in semen, which is used for presumptive identification, can also be found in smaller concentrations in saliva, vaginal fluid, and various others (2). Semen stains fluoresce at 450 nm under an alternate light source but sweat, saliva, and mucus do as well (2). Unlike semen, whose presence can be confirmed using a p30 immunochromatographic card or by the microscopic identification of spermatozoa, vaginal fluid and menstrual blood do not have confirmatory tests. To date, there is no method for determining whether the origin of blood is venous or menstrual. Vaginal fluid is often difficult to identify because its cells are morphologically similar to buccal epithelial cells (3). Although confirmatory

tests are deemed more reliable, that is not always the case. The immunochromatographic cards used for confirming the presence of blood or semen are also capable of producing false positives (1). Sensitivity is yet another limitation that arises with serological testing. If a sample is degraded or does not have enough human cells, results may vary and are often inconclusive (4). Thus, it is important to find a reliable method to classify body fluids and no longer depend on tests which may provide inaccurate results.

Body Fluid Identification Techniques

In more recent years, there have been many experiments performed to classify and distinguish body fluids. Some included the use of microRNA, messenger RNA, ribosomal RNA, and DNA methylation measurements (1). Throughout trials and research, it was determined that mRNA and miRNA were able to detect and discriminate between the types of body fluids but did not have a dependable quantitation method (1). It was noted that there were variations in the number of RNA markers in a sample which could cause issues (1). Specifically, in the case of mixtures, having different amounts of the RNA markers made it hard to conclude whether the data was showing background noise or true signals (1). Another technique used to differentiate body fluids was the assessment of the levels of methylation in DNA using CpG islands (1). Methylation, a common process that occurs in both eukaryotes and prokaryotes, is a key factor in many body functions including aging, cell type expression, and the inactivation of the X chromosome (5). In mammals, a little over half of the DNA is unmethylated so when methylation occurs, it can be monitored (5). The degree of methylation varies between body fluids and tissues which was why it was a good method for discriminating between fluids but also why it caused complications (1). It was determined that, although it was successful in

differentiation between body fluids, the amount of methylation varied between individuals which made this technique less than optimal, especially when used alone (1).

The Human Microbiome

Regarding body fluid identification, microbiomes have been significantly studied. It is known that nearly all body fluids contain some quantity of microbes. Bacteria, for example, is present all over the body and distributed at various concentrations (4). In some fluids, there are more microbial cells than total human cells (e.g., feces), which could reduce the need to rely solely on information from human cells (4). Research using ribosomal RNA to classify the organisms in microbiomes has gained popularity and is continuously being tested. The increased interest in the effects of microbial communities on health and contributions to disease resulted in the conception of the Human Microbiome Project in 2008 and has caused the development of a reference database (6). This project aimed to determine if body fluids had main microbial communities (6). It began with the study of the microorganisms found in and on the human body which were collected from 300 individuals (6). Of the various microbiota found in body sites and fluids, bacteria were largely studied. The 16S ribosomal RNA gene was used to characterize the bacterial structure associated with each body sites (6). This 16S rDNA is an informative marker with 1500 base pairs and containing both conserved and hypervariable regions (7). It is continuously used in studies and it was determined that the function of this gene has not changed over time, so any variation in the DNA sequence could accurately depict evolutionary or phylogenic changes (7). The classification of bacterial organisms is generally most informative down to the genus level. It was found that more than 90% of 16S rDNA sequences can be identified at the genus level but that value drops greatly at the species level and some sequences

remain unclassified (7). To date, 16S rDNA studies are continuously performed and more information can be obtained using the bacterial microbiome.

18S rDNA

Unlike 16S rDNA, research on 18S rDNA in terms of body fluid identification has not been as widely explored. For this research, non-human eukaryotic microbial communities will be studied. Specifically, the species under the Kingdom Fungi may prove to be informative. This kingdom includes yeasts which can be detected in body sites such as male and female intimate areas, the mouth, and the armpits (8). These sites are vital as they produce the fluids in which forensic samples originate. Therefore, the main focus of this research will be on 18S ribosomal DNA to detect and identify the microbial taxa within each body fluid to determine if they can be differentiated.

The 18S region, which is also part of the overall 40S ribosomal subunit, is a site for protein synthesis (9). The 18S rRNA gene is one of the most used markers for phylogeny (10). The diversity of the eukaryotic organisms that are present in certain samples can be determined based on the variety of sequences from the hyper-variable regions of the 18S rRNA gene (10). Although it contains nine regions, the target of this research is the V9 hyper-variable region of the 18S rDNA. In a study done by Hadziavdic et al. (2014), the V9 region proved to be highly variable and very well fit for determining eukaryotic diversity(10).

Amplification and Mammalian Blocking Primers

To ensure the eukaryotic microbial biodiversity in body fluids is accurately depicted, specific primers need to be used that will allow the V9 hyper-variable region to be amplified properly (10). This can be accomplished by using primers that anneal to common eukaryotic sequences (10). Along with these primers, a mammalian blocking primer should be included to

inhibit the amplification of host (i.e., human) DNA (11). A blocking primer inhibits non-target DNA amplification either by blocking the elongation or annealing step during the PCR amplification process (11). Annealing inhibiting blockers function by competing with amplification primers to prevent binding (11). Elongation arrest blockers bind to the area in between two amplification primers to hinder the elongation process (11). These blocking primers are created similar to the average primer but include the addition of a C3 spacer at the 3'end, which is responsible for inhibiting the elongation of the primer (11). There is a short stretch of species-specific bases, such as mammalian-specific, at the 3' end of the blocking primer to make certain it attaches to the undesired DNA (11). Although blocking primers cannot prevent all human DNA from being amplified and present in the samples, they greatly reduce the quantity amplified and any that remain are generally eliminated during data analysis.

Sequencing

Studies of the phylogeny and taxonomy of the microbial world advanced with the use of Sanger sequencing (12). Although Sanger sequencing was very beneficial, over time, more efficient instruments were developed with high throughput sequencing capabilities, that reduced the time and cost of use (13). Next generation sequencing revolutionized how sequencing was performed and it is still used today. With these instruments (e.g., MiSeq, HiSeq, Ion Torrent, etc.) and by using a barcoding approach, now hundreds of microbial samples can be sequenced in a single run (12). Microbial samples can be sequenced either as single-index (only one primer is barcoded) or dual-index (both primers are barcoded) approaches (14). The dual-index approach, as developed by Kozich et al. (2013), is advantageous for multiplexing as it requires a lesser number of primers (40 primers can sequence 384 samples) than the single-index method

(minimum 384 primers are needed for 384 samples), and hence the dual-index method is in recent times a preferred method for 16S rDNA/18S rDNA sequencing (14).

Eukaryotic Taxonomy

Structurally, the taxonomic classification of eukaryotic organisms has undergone multiple changes over time and is yet to be complete. The ability to sequence DNA and better classify organisms created a shift within the taxonomical system which required further revision of original groupings. Taxon which were previously believed to descend from a common ancestor had to be modified based on the results of their DNA sequences. In 2005, it was proposed that the 4 kingdom system for classifying eukaryotes was outdated and a 6 group classification would be more fitting (15-16). Although classifications are improving, there is still a need to study eukaryotic organisms closer to correctly categorize them. The SILVA (v132) database is in workings with the UniEuk Project, whose mission is to maintain universal eukaryotic taxonomy, to stay up to date with advancements in classification (17).

16S rDNA Studies

To date, numerous studies have been performed using 16S rDNA. In a study performed by Giampaoli et al. (2012), it was discovered that certain strains of bacteria exhibited strong signs of linkage to saliva, fecal, and vaginal fluid samples (18). Research performed by Hanssen et al. (2017), concluded that the microbes present in saliva samples could be separated from those found on skin (1). While it was tested with saliva samples taken off skin surfaces, it was stated that this method can be replicated using other body fluids (1). Another study by Dobay et al. (2018), revealed the resilience of bacteria by displaying their ability to persist in six different body fluids for a period of 30 days under indoor environmental conditions (4). Studies have continuously shown that certain bacteria are heavily associated with the body areas they came

from (4). Further experimentation has displayed, statistically and with 3-D modeling (spatial scale), that the microbes cluster themselves together by the bodily area but also form a distinguishable pattern when observed on a spatial scale (4). Although results of these 16S rDNA studies show promise, a highly discriminatory and accurate method has yet to be perfected for use in casework.

As a predecessor to this 18S rDNA study, a bacterial study was conducted at the Seashols-Williams laboratory (19). This study was aimed to characterize bacterial communities in all forensically relevant body fluids including menstrual blood, vaginal fluid, semen, urine, feces, saliva and blood (19). Bacterial signatures were developed from all body fluids and a novel classification model was developed (19). For prediction modeling and precision testing, every body fluid could be identified with an accuracy of 90% or greater at the genus level (19). Unfortunately, vaginal fluid and menstrual blood could not be differentiated using this study, so they were grouped together as female intimate samples (19). When assessing the ratio of human DNA to bacterial DNA, it was observed that menstrual blood and vaginal fluid had high yields of both but the human DNA was greater (19). Semen was found to have very little bacterial DNA and a majority of human DNA (19). These results led to the idea that using 18S rDNA could potentially complement the bacterial study and solve issues that arose with it.

Aim of Study

The goal of this research is to characterize eukaryotic community structures associated with three human intimate biological samples (e.g., semen, vaginal fluid, menstrual blood) and to utilize this information to develop a predictive model for identification of these samples with known error rates. The fluids were selected because they are commonly found at crime scenes and both vaginal fluid and menstrual blood have no accurate means of identification. The aim is

to discover if there are eukaryotic microbes that are more common in a specific body fluid. It would also help to observe if any of those microbes are present in certain fluids and not in others.

Materials & Methods

Sample Collection

The samples that were used for this research (i.e., vaginal fluid, menstrual blood, and semen) were previously collected following a VCU approved IRB protocol (HM20002931) from volunteers, onto swabs and the DNA was extracted. The semen samples were first collected into a container and stored at 4°C (19). One hundred and fifty microliters of each semen sample was aliquoted onto a cotton swab and left to dry at room temperature (19). The vaginal fluid and menstrual blood samples were collected by the donors, placed in swab boxes, and stored at room temperature until needed (19). In total, 99 vaginal fluid, 94 menstrual blood, and 53 semen samples were used for this research (Table 1).

DNA Extraction

DNA from vaginal fluid, menstrual blood, and semen samples were extracted using the Qiagen® QIAamp DNA Investigator Kit (Qiagen, Hilden, Germany) on the QIAcube (Qiagen) robotic platform, as described in the manufacturer's protocol but with final elution volume of 30 µl (19-20). Those samples were then quantified for bacterial DNA yield on the Applied Biosystems® Quantstudio 6 Flex Real-Time PCR System using a previously optimized protocol (19). Reagent blanks were processed along with samples to ensure no contamination. A standard curve was created using the ZymoBIOMICS^{TM} Microbial Community DNA Standard D6306 (19). The microbial mock community was also used as a positive control throughout the workflow.

Barcoding and PCR Amplification

18S rDNA was amplified and sequenced using a dual-index strategy as described by Kozich et al. (2013) in the Earth Microbiome Project 18S rDNA protocol (14). A mammalian cell blocking primer, Mammal_block_I-short_1391f, was used along with forward and reverse primers (21). The primers included an Illumina adapter sequence (5'-AATGATACGGCGACCACCGAGATCTACAC-3' for forward and 5'-CAAGCAGAAGACGGCATACGAGAT-3' for reverse), an eight nucleotide index sequence, a ten nucleotide pad sequence (5'-TATCGCCGTT-3' for forward and 5'-AGTCAGTCAG-3' for reverse), a two nucleotide non-complementary linker sequence (CG for forward, CA for reverse), an 18S rRNA gene specific forward primer (1391f: 5'-GTACACACCGCCCGTC-3'), and a reverse primer (EukBr: 5'-TGATCCTTCTGCAGGTTCACCTAC-3') (21). These primers were used along with other reagents and samples to perform barcoding (21). Within each well of a 96 well plate, 12.5 µl of Promega® 2X PCR master mix (Promega, Madison, WI) was combined with 5.5 µl of nuclease free water, 1 µl each of forward and reverse primer (5 pmol/µl), 4 µl of blocking primer (10 pmol/µl), and 1 µl of sample (input: 0.02 ng for semen and 0.3 ng for menstrual blood and vaginal fluid) for a total of 25 µl. The DNA inputs were determined for each body fluid type based on previous 16S rDNA study. The samples were amplified using an Applied Biosystems Veriti^{TM} 96-Well Thermal Cycler (Thermo Fisher Scientific, Waltham, MA) (21).

Gel Electrophoresis

After PCR, sample products were loaded onto a gel for visualization. In a 96 well plate, 3 µl of sample (including mock communities, negatives, and reagent blanks) was added together with 7 µl of TAE buffer and 2 µl of 6X loading dye. Once mixed, each sample was loaded onto

the 1.6% agarose gel, ran for 30-60 minutes at 150V depending on the gel size, and visualized using the UVP Multi Doc-it ™ 125 Imaging System.

Gel Extraction and Purification

Following visualization in an agarose gel, the DNA was extracted from samples as described in the QIAquick® Gel Extraction Kit/QIAquick® PCR & Gel Cleanup Kit protocol (22). This gel extraction step was necessary due to double banding observed during gel electrophoresis. The target band was at the anticipated size of 260 base pairs and contained desired eukaryotic DNA. A secondary band occurred at a differing size, around 500 base pairs, but contained unwanted DNA or products. The desired bands were physically cut out of the gel, weighed, and purified as stated in the QIAquick® Gel Extraction Kit/QIAquick® PCR & Gel Cleanup Kit protocol (22). Ten microliters of each sample were further purified using the Agencourt® AMPure XP PCR Purification Kit (Beckman Coulter Life Sciences; Indianapolis, IN) and by following the manufacturer's protocol (23).

MiSeq® Library Preparation and Sequencing

Purified 18S rDNA amplified products were quantified using the Qubit® dsDNA HS Assay Kit (Thermo Fisher, USA) on the Qubit® fluorometer. The 18S rDNA library was prepared by pooling all samples in equimolar concentration (1 ng/µl). The quantities of reagent blanks and negatives that were added were based on average input volume of their associated sample types to the pool. Many samples required more input to reach 1 ng/µl for the final pool, so the average volume added for each sample type was calculated and used as the input for reagent blanks and negatives. For example, if the average volume added for all menstrual blood samples was 2.95 µl, reagent blanks that were processed with menstrual blood samples would be added to the pool at that volume. After the final pool was quantified, it was dried down by using

a vacuum centrifuge and quantified again until the pool reached 1 ng/ul. The 18S rDNA library was sequenced using the MiSeq v2 Reagent Kit on the MiSeq FGx (Illumina® Inc., USA) platform and by following the manufacturer's protocol (24). There were two total runs for this study, one with 116 samples and the other with 171, totaling 287 samples and included 26 extraction reagent blanks, 8 PCR negatives, and 7 mock communities. In total, 246 of those samples were body fluid samples (99 vaginal fluid, 94 menstrual blood, and 53 semen).

Data Analysis

Sequence data were analyzed using the 18S rDNA protocol as described in Forger et al. (2019), on the open source data analysis pipeline Mothur (version 1.39.5) (https://www.mothur.org) (21). The first step in analysis was creating a contig file which would pair the forward and reverse strands. The contigs with short reads were removed from analysis. The sequences were then aligned using the SILVA v132 reference database. Sequences that did not align with the target region of interest were removed. The sequences were also run through UCHIME to determine if there were any chimeras present (individuals with two or more sets of DNA) and remove them. Samples along with their sequences were obtained and those with under 200 reads were removed. The samples were classified (classify.seqs) and the unwanted taxon were removed to only leave eukaryotic cells (remove.lineage). This included archaea, bacteria, vertebrata, and other unknown or unclassified sequences. The removal of these undesired lineages ensured that only non-mammalian eukaryotic DNA would be analyzed. The eukaryotic family Aspergillaceae was removed because it was discovered to be in the reagent blanks across all body fluids that had sequences. Upon further analysis, it was concluded that the reagent blank sequences were due to this Aspergillaceae rather than contamination. The resulting taxonomy file was used to create a relative abundance chart. The taxonomic level that was used for analysis

was at the Family level (level 5). All taxa that had sequences below 100 reads were grouped together and labeled “Rare <100 reads” and those that were unclassified were labeled as such. The Log10 of those values was obtained to create a heat map. Those values that were at 0 were changed to 0.001. Alpha and beta diversity data was also assessed in Mothur v1.39.5 using operational taxonomic units (OTUs) and phylogenetic approaches. Alpha diversity was evaluated using OTUs at a 5% genetic distance. Beta diversity was analyzed using weighted UniFrac methods. The R software v3.6.3 was used for the statistical analyses and graphing of data (25). An Analysis of Molecular Variance (AMOVA) test was used to determine any significant differences in the microbial communities of the various body fluids (menstrual blood, vaginal fluid, and semen) (19). Using RStudio v3.6.3, a Principal Coordinate Analysis (PCoA) plot was developed using the ggplot2 package along with a heat map (gplots package) and boxplot to display data (25).

Results

Quantitation of Cleaned 18S rDNA Amplified Products

Following purification, 18S rDNA PCR products were quantitated using the Qubit® fluorometer. The results showed fairly similar quantitation ranges, and standard deviations, between the body fluids (Table 2). The majority of the samples in each body fluid were generally near the 0.3 ng/µl to 1.0 ng/µl range. The reagent blanks and negative controls had quantitative values <0.05 ng/µl.

Sequence Characteristics

Once running the Mothur v1.39.5 program and analyzing the data, the number of sequence reads for each sample was observed before classification (Table 3). As expected, the negatives and most of the reagent blanks were well under 200 reads, ranging from 6 to 129. Out

of 26 reagent blanks, 7 were above 200 reads, the highest being 1,902. After cleaning up the data by removing undesired lineages (archaea, bacteria, vertebrata, etc.), mocks, and sequences below 200 reads, the number of samples with desired requirements dropped dramatically from 287 to 36 after classification. The total remaining samples that were used for further analysis included 15 vaginal fluid, 13 menstrual blood, and 8 semen samples. This showed a sample loss of about 85% overall and within each body fluid (Table 4). The 36 remaining samples were classified and various numbers of taxa were identified at the different taxonomic levels. At the second level (Phylum), 20 different phyla were observed. At Level 3 (Class), 33 classes were identified and there were 54 orders at Level 4 (Order). Level 5 (Family) showed 62 families and there was a total of 77 genera found at Level 6 (Genus).

Eukaryotic Structure

The relative abundance of eukaryotic cells in each body fluid was graphed (Figure 1). The data was observed at the Family level (Level 5) for classification. The taxa presented on the graph were those which had over 100 sequences. Those that were below 100 reads (1.66%) or unclassified (0.11%) were labeled as such. That left the top 13 most abundant taxa across all three body fluids. In vaginal fluid, the most abundant eukaryotic families were Debaryomycetaceae (51.69%) and Saccharomycetaceae (26.05%). In semen, the most abundant was Malasseziaceae (49.88%). Menstrual blood had a variety of eukaryotic families and appeared to have mixes of Malasseziaceae (30.43%), Debaryomycetaceae (22.45%), and Saccharomycetaceae (21.14%). Malasseziacea and Saccharomycetaceae were observed in all three body fluids but at varying degrees. Debaryomycetaceae was not observed in semen but found in the female intimate samples. Overall, these taxa were amongst the top 5 eukaryotic families found (Figure 2).

Eukaryotic Diversity

For alpha diversity, the Inverse Simpson Index was used to determine differences in diversity within each sample type. A boxplot was created to visualize the data (Figure 3). The boxplot shows that all three fluids had very low alpha diversity and Inverse Simpson Index values below 2 and closer to 1. There was very little difference in alpha diversity between all three fluids; menstrual blood (1.46), semen (1.45), and vaginal fluid (1.29).

Eukaryotic beta diversity (weighted UniFrac) differed significantly between semen and vaginal fluid but didn't differ significantly between semen and menstrual blood or menstrual blood and vaginal fluid (Table 5).

Discussion

During gel electrophoresis, all samples including mock communities, negatives, and reagent blanks were monitored to determine if any failed to amplify or were contaminated. Any reagent blanks that showed banding were removed from analysis along with their associated samples. If banding appeared in a negative control, the batch of samples associate with that negative control were rerun to ensure no contamination. Once all negatives and reagent blanks were clean, samples were permitted to continue through the processes for future sequencing. Once samples were quantitated after purification, they all showed relatively similar ranges amongst body fluids. All reagent blanks and negative controls had quants <0.05 ng/µl. Once they were ready, samples were pooled, sequenced, and analyzed in Mothur v1.39.5.

During the data analysis using Mothur v1.39.5, it was observed that Aspergillaceae was the most dominant family in all reagent blanks. Upon further investigation, it was determined that this was not due to contamination because, except semen, this family was not abundant in any other samples included in this study.

With the removal of certain lineages (archaea, bacteria, vertebrata, etc.), mocks, and sequences below 200 reads, the number of samples to be used for this study dropped from 287 to 36 after classification. This was an 85% sample loss overall and between each body fluid (Table 4). Although a mammalian blocking primer was used, this loss shows that there was an overpowering amount of human DNA in the samples. It was expected that semen would have a majority of human DNA as observed in the bacterial study (19). Menstrual blood and vaginal fluid reflected these findings as they had a large amount of human DNA as well. The mammalian blocking primer may have prevented some human DNA from amplifying but could not block it all due to the overwhelming amount present in each body fluid. This is also observed when looking at the sequence reads for each fluid. The fluids had a range of reads which averaged around 21,000 – 25,000 but then dropped to less than about 1,000 reads (Table 3). We can see that samples were successfully amplified but those sequences were mostly human DNA.

Taking an overall view at the data and results helped show some trends and give information about these body fluids. Looking at the relative abundance charts (Figures 1 and 2), in vaginal fluid, the most prevalent eukaryotic families observed were Debaryomycetaceae and Saccharomycetaceae. Debaryomycetaceae are yeasts which can been found in the gut of wood feeding insects (26). Saccharomycetaceae are fermentation yeasts that are commonly found in the gastrointestinal tract, respiratory tract, and the vaginal mucosa (27). In semen, the most abundant eukaryotic family was Malasseziaceae, which are yeasts that are found on skin surfaces of humans of animals and can be known to cause skin diseases (28). Menstrual blood had many eukaryotic families but overall appeared to have mixes of mostly Debaryomycetaceae, Saccharomycetaceae, and Malasseziaceae. It was anticipated that many of the eukaryotes found in these body fluids would be yeasts, and our results support that idea.

The amount of diversity within the samples of each body fluid was observed in the boxplot (Figure 3). When the Inverse Simpson Index value for the specific body fluid is high, it indicates greater diversity. In this case, all three body fluids were below 2 on the index, which shows that there was not much diversity within the samples of the body fluids. This is understandable because looking at the relative abundance graphs, especially in the case of vaginal fluid and semen, most of the samples had a major family that they all shared. The boxplot shows that menstrual blood had a little more diversity with its samples, followed by semen then vaginal fluid.

The purpose of this study was to determine if body fluids could be differentiated based on their eukaryotic composition. This leads to the question of whether, statistically, the body fluids are truly different from each other. The AMOVA table shows comparisons of the body fluids against each other to see any significant differences (Table 5). The Bonferroni correction was used to relieve any issues encountered with multiple comparisons (29). Since we are comparing multiple sample types, there is a higher chance of incorrectly rejecting the null hypothesis (29). Thus, the Bonferroni correction accounts for that and tests each individual 'hypothesis' by the significance level (29). In this case, when semen was tested against menstrual blood, it was not significantly different, although the p-value (0.0638) was very close to Bonferroni pair-wise error rate (i.e., 0.0166667). On the other hand, vaginal fluid was clearly not significantly different from menstrual blood. When vaginal fluid was tested against semen, they were in fact found to be statistically different from each other.

The two-dimensional Principal Coordinate Analysis (PCoA) plot generated based on weighted UniFrac distances explained approximately 37% of variation associated with this data set (Figure 4). Compared to vaginal fluid samples, much greater dispersion from centroid was

observed for semen and menstrual blood samples. This level of dispersion might have occurred either because of the overall variability in multivariate datasets or because of greater beta diversity associated with semen and menstrual blood samples.

Conclusion

The goal of this study was to determine if body fluids such as menstrual blood, vaginal fluid, and semen could be differentiated from one another based on the eukaryotic communities they possess. The results showed that we can see differences between vaginal fluid and semen, but we cannot separate menstrual blood and vaginal fluid or menstrual blood and semen. This study demonstrates that it is possible to separate vaginal fluid and semen, which would a powerful aid in sexual assault investigations. In combination with the previous 16S rDNA data, an assay could be developed which could revolutionize how we look at crime scenes and differentiate body fluids. In the future, it may be beneficial to run this study using a larger and more diverse population size. The general demographic of participants for this study were mostly college aged students between 18 and 30 years old (Table 1). These results were obtained from testing 246 samples which narrowed to 36 samples for analysis. With a larger sample size, ranging hundreds to thousands more, it may provide clearer trends and a more detailed answer as to what eukaryotic families could be expected in body fluids. Another study that may be performed in the future is testing a different DNA marker to get better taxonomic results. The taxon that were most abundant in the body fluids tested were yeasts which are fungi. The 18S rDNA marker can be problematic when trying to identify fungi at lower taxonomic levels. Studying alternative markers, such as the internal transcribed spacer region (ITS), could aid in the differentiation of body fluids by further characterizing fungi at the genus or species level (30). The ITS region has been found to be discriminatory when identifying eukaryotes and was

proposed as a standard marker for fungi (30). With the use of this marker, there is potential for better identification and separation for body fluids, make a great impact on the forensic science community.

References

1. Hanssen EN, Avershina E, Rudi K, Gill P, Snipen L. Body fluid prediction from microbial patterns for forensic application. Forensic Sci Int Genet 2017;30:10–7.
2. Poyntz F, Martin P. Comparison of p30 and acid phosphatase levels in post-coital vaginal swabs from donor and casework studies. Forensic Sci Int 1984;24(1):17–25.
3. Anderson DJ, Marathe J, Pudney J. The Structure of the Human Vaginal Stratum Corneum and its Role in Immune Defense. Am J Reprod Immunol 2014;71(6):618–23.
4. Dobay A, Haas C, Fucile G, Downey N, Morrison HG, Kratzer A. Microbiome-based body fluid identification of samples exposed to indoor conditions. Forensic Sci Int Genet 2019;40:105–13.
5. Forat S, Huettel B, Reinhardt R, Fimmers R, Haidl G, Denschlag D, et al. Methylation markers for the identification of body fluids and tissues from forensic trace evidence. PLoS One 2016;11(2):1–19.
6. Human Microbiome Project Consortium. A framework for human microbiome research. Nature 2012;486(7402):215-21.
7. Janda JM, Abbott SL. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J Clin Microbiol 2007;45(9):2761-2764.
8. Aaron DM. Candidiasis (Yeast Infection) - Skin Disorders. Merck Manuals Consumer Version. 2020. Available at: https://www.merckmanuals.com/home/skin-disorders/fungal-skin-infections/candidiasis-yeast-infection
9. Hayashi R. Gene Expression and the Impact of Antioxidant Supplements in the Cataractous Lens. Handb Nutr Diet Eye 2014;517–24.

10. Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C. Characterization of the 18s rRNA Gene for Designing Universal Eukaryote Specific Primers. PLoS One 2014;9(2).

11. Boessenkool S, Epp LS, Haile J, Bellemain E, Edwards M, Coissac E, et al. Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. Molecular Ecology 2012;21:1806-15.

12. Gurson, N. When Do I Use Sanger Sequencing vs NGS? Behind the Bench. 2018. Available at: https://www.thermofisher.com/blog/behindthebench/when-do-i-use-sanger-sequencing-vs-ngs-seq-it-out-7/

13. Lie AAY, Liu Z, Hu SK, Jones AC, Kim DY, Countway PD, et al. Investigating Microbial Eukaryotic Diversity from a Global Census: Insights from a Comparison of Pyrotag and Full-Length Sequences of 18S rRNA Genes. Appl Environ Microbiol 2014;80(14):4363–73.

14. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl Environ Microbiol 2013;79(17):5112–20.

15. Kumar S, Rzhetsky A. Evolutionary relationships of eukaryotic kingdoms. J Mol Evol 1996; 42(2):183–93.

16. Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, et al. The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. J Eukaryot Microbiol 2005;52:399-451.

17. “Silva.” Silva, www.arb-silva.de/.

18. Giampaoli S, Berti A, Valeriani F, Gianfranceschi G, Piccolella A, Buggiotti L, et al. Molecular identification of vaginal fluid by microbial signature. Forensic Sci Int Genet 2012;6(5):559–64.

19. Seashols-Williams S, Green R, Wohlfahrt D, Brand A, Tan-Torres AL, Nogales F, et al. An accurate bacterial DNA quantification assay for HTS library preparation of human biological samples. Electrophoresis 2018;39(21):2824–32.

20. QIAGEN. QIAamp® DNA Investigator Sample & Assay Technologies. 2012;(June):1–60. Available at: https://www.qiagen.com/us/resources/download.aspx?id=dcc5a995-3743-4219-914d-94d6a28e49b3&lang=en

21. Forger L, Singh B, Simmons T, Connon CC. JLS. Long-Term Postmortem Interval Estimation Based on Eukaryotic Community Succession on Decomposing Porcine Remains. Virginia Commonwealth University. 2018.

22. QIAGEN. QIAquick® Gel Extraction Kit QIAquick® PCR & Gel Cleanup Kit. 2018;(July):1-3. Available at: https://www.qiagen.com/cn/resources/download.aspx?id=a72e2c07-7816-436f-b920-98a0ede5159a&lang=en

23. Agencourt. AMPure PCR Purification Instructions for Use. 2016;(August):1-36. Available at: https://www.beckmancoulter.com/wsrportal/techdocs?docname=B37419AA.

24. ILLUMINA. ForenSeq DNA Signature Prep Reference Guide. 2015;(September):1-36. Available at: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/forenseq/forenseq-dna-signature-prep-guide-15049528-01.pdf

25. R, D.C.T. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, http://www.R-project.org, Vienna, Austria. 2011).

26. Hui F, Ren Y, Chen L, Li Y, Zhang L, Niu Q. Molecular Phylogenetic Analysis Reveals the New Genus Hemisphaericaspora of the Family Debaryomycetaceae. PLoS One 2014;9(7):e103737.

27. Zhou X, Li Y, She Z. Chapter 5 - Oral Mucosal Microbes. Atlas of Oral Microbiology: from Healthy Microflora to Disease 2015;95–107.

28. Saunders CW, Scheynius A, Heitman J. Malassezia fungi are specialized to live on skin and associated with dandruff, eczema, and other skin diseases. PLoS Pathog 2012;8(6):e1002701.

29. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014;34:502–8.

30. Santamaria M, Fosso B, Licciulli F, Balech B, Larini I, Grillo G, et al. ITSoneDB: a comprehensive collection of eukaryotic ribosomal RNA Internal Transcribed Spacer 1 (ITS1) sequences. Nucleic Acids Res 2018;46(D1): D127–32.

Appendix I: Tables and Figures

Table 1: Sample Demographic Information by Body Fluid

	Semen	Menstrual Blood	Vaginal Fluid
Gender			
Female	0	94	99
Male	53	0	0
Unreported	0	0	0
Age Group (Years)			
<18	0	0	0
18-30	37	87	94
31-50	7	4	2
>50	3	0	0
Unreported	6	3	3
Ethnicity			
Caucasian	23	38	35
African American	17	19	21
Hispanic	2	8	8
Asian	4	17	16
Mixed	2	4	9
Other/Unreported	5	8	10
Total Samples	53	94	99

Table 2: Qubit Fluorometer Quantitation Range Values for Menstrual Blood, Vaginal Fluid, and Semen

	Range	Standard Deviations
Menstrual Blood	0.067 ng/µl - 1.31 ng/µl	±0.277
Vaginal Fluid	0.120 ng/µl - 1.62 ng/µl	±0.285
Semen	0.381 ng/µl - 1.43 ng/µl	±0.263

Table 3: Average Sequence Read Per Sample Before and After Removal of Vertebrate/Archaea/Bacterial Sequences

	Before		**After**	
	Range	**Average**	**Range**	**Average**
Menstrual Blood	46 – 98,709	21,996	215 – 1954	659
Vaginal Fluid	162 – 118,932	25,034	270 – 716	476
Semen	2,883 – 42,953	21,111	204 – 4566	1039

Table 4: Percent Loss After Removal of Vertebrate/Archaea/Bacterial Sequences

	Before	**After**	**Percent Loss**
Menstrual Blood	94	13	86.17%
Vaginal Fluid	99	15	84.85%
Semen	53	8	84.90%
Total	**246**	**36**	**85.37%**

Table 5: Analysis of Molecular Variance (AMOVA) Comparing Menstrual Blood, Vaginal Fluid, and Semen

Analysis of Molecular Variance (AMOVA) Table		
	Semen	Vaginal Fluid
Vaginal Fluid	0.0026*	
Menstrual Blood	0.0638	0.24236
*Indicates significance; Bonferroni pair-wise error rate: 0.0166667		

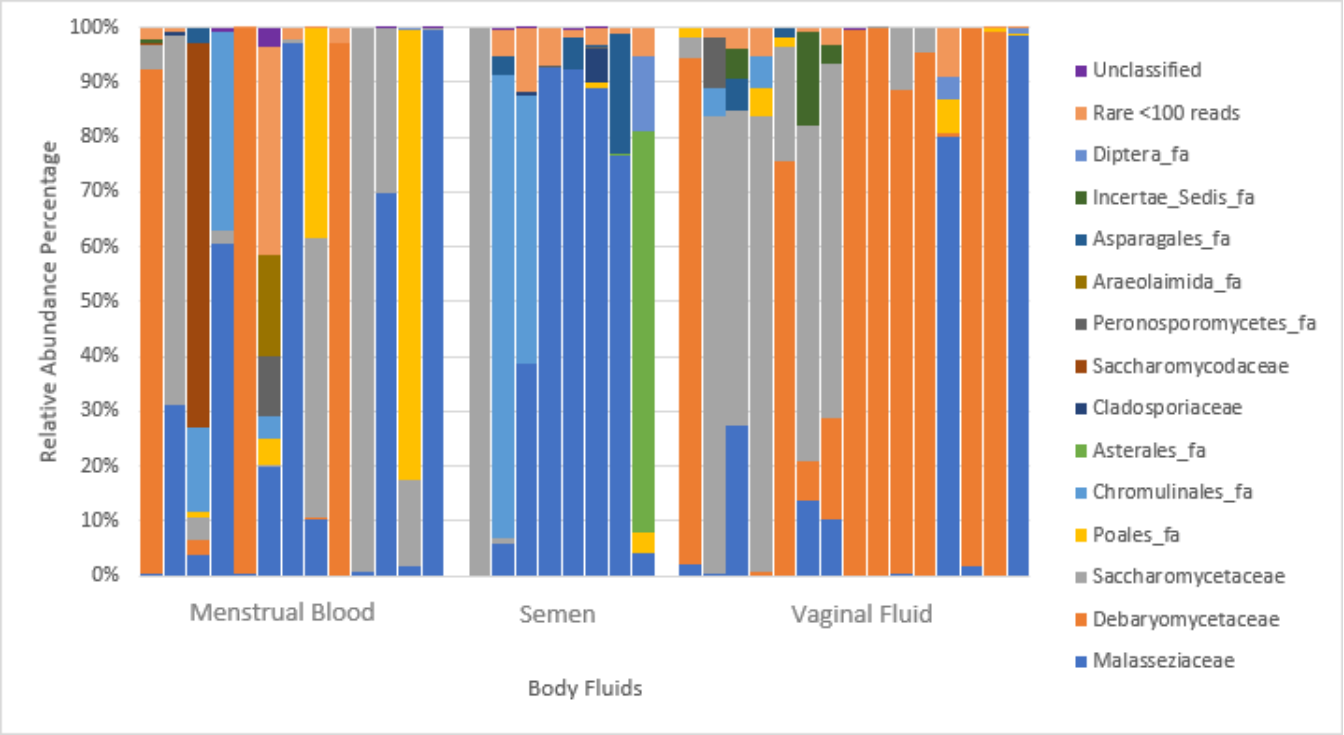

Figure 1: Relative abundance data for top 13 eukaryotic families found in menstrual blood (n=13), semen (n=8), and vaginal fluid (n=15). In vaginal fluid, the most abundant eukaryotic families were Debaryomycetaceae (51.69%) in dark orange and Saccharomycetaceae (26.05%) in gray. In semen the most abundant was Malasseziaceae (49.88%) in blue. Menstrual blood had a variety of eukaryotic families and appeared to have mixes of Malasseziaceae (30.43%), Debaryomycetaceae (22.45%), and Saccharomycetaceae (21.14%). Taxa below 100 reads were grouped together (1.66%) as were those unclassified (0.11%).

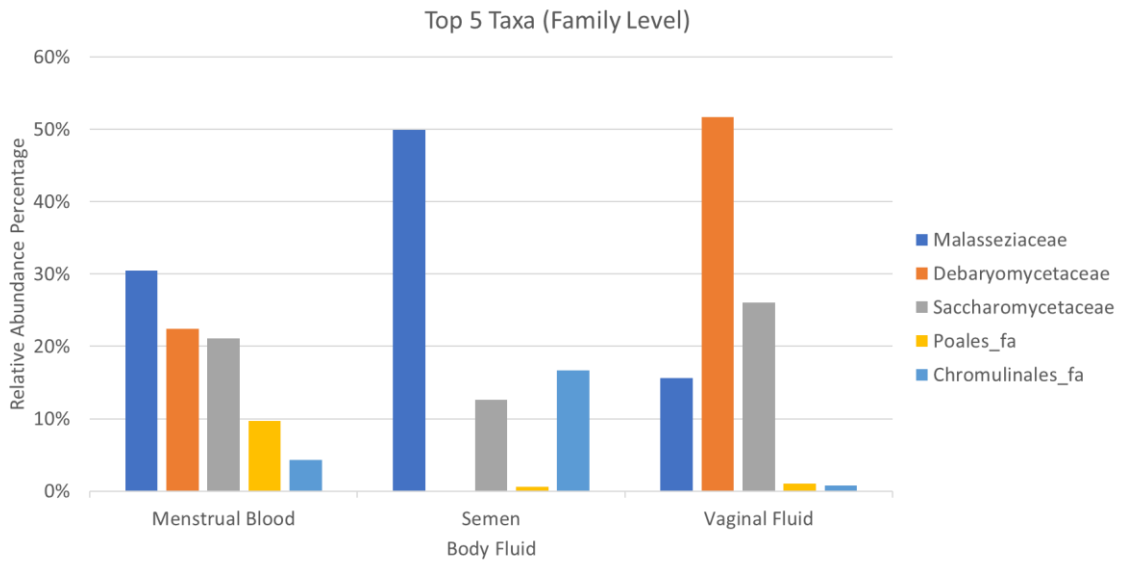

Figure 2: Relative abundance data for top 5 taxa found in menstrual blood (n=13), semen (n=8), and vaginal fluid (n=15). Vaginal fluid and semen each had one major taxa which composed about 50% of the body fluid. Menstrual blood had the top three taxa between 20-30% abundant. All three fluids contained Malasseziaceae and Saccharomycetaceae.

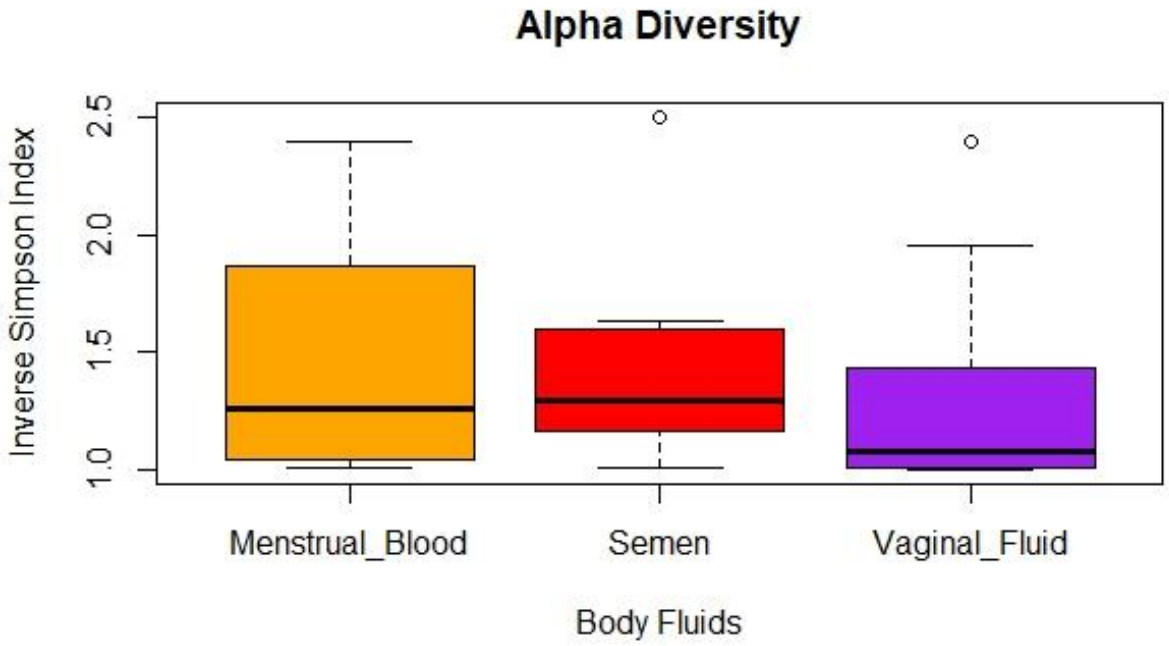

Figure 3: A menstrual blood outlier was removed from this figure. The Inverse Simpson Index shows that all three body fluids have very low alpha diversity but of the three, menstrual blood (n=13) was slightly higher than semen (n=8) and vaginal fluid (n=15). The boxplot shows that all three fluids had an Inverse Simpson Index value below 2 and closer to 1. The highest value average Inverse Simpson Index value was observed in menstrual blood (1.46), followed by semen (1.45) and vaginal fluid (1.29).

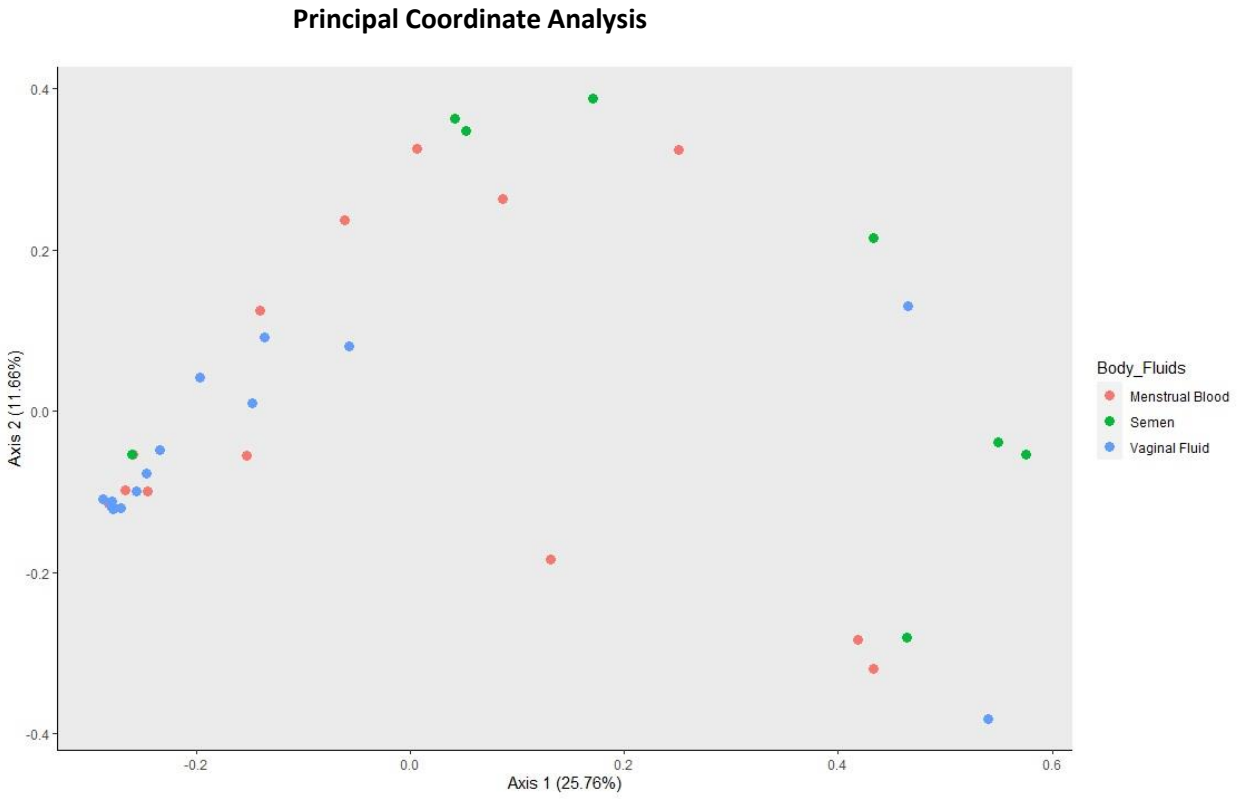

Figure 4: The PCoA plot shows the dissimilarities between samples of menstrual blood (n=13), semen (n=8), and vaginal fluid (n=15). Axis 1 shows that 25.76% of the data could be explained by the plot and in Axis 2, 11.66% could be explained. It could be seen that there was some clustering of the vaginal fluid samples but no real clustering of semen and menstrual blood samples.

Appendix II: Mothur Commands and RStudio Codes

MOTHUR COMMANDS

make.contigs(file=BfID_18S.txt, trimoverlap=t, insert=30, processors=16)

summary.seqs(fasta=BfID_18S.trim.contigs.fasta, processors=16)

screen.seqs(fasta=BfID_18S.trim.contigs.fasta, group=BfID_18S.contigs.groups, maxambig=0, minlength=50, maxlength=200)

summary.seqs(fasta=BfID_18S.trim.contigs.good.fasta)

unique.seqs(fasta=BfID_18S.trim.contigs.good.fasta)

summary.seqs(fasta=BfID_18S.trim.contigs.good.unique.fasta, name=BfID_18S.trim.contigs.good.names)

align.seqs(fasta=BfID_18S.trim.contigs.good.unique.fasta, reference=silva.nr_v132.align, processors=16)

summary.seqs(fasta=BfID_18S.trim.contigs.good.unique.align, name=BfID_18S.trim.contigs.good.names)

screen.seqs(fasta=BfID_18S.trim.contigs.good.unique.align, name=BfID_18S.trim.contigs.good.names, group=BfID_18S.contigs.good.groups, summary=BfID_18S.trim.contigs.good.unique.summary, start=42554, end=43116, maxhomop=8)

summary.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.align, name=BfID_18S.trim.contigs.good.good.names)

filter.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.align, vertical=T, trump=.)

unique.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.fasta, name=BfID_18S.trim.contigs.good.good.names)

summary.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.names)

pre.cluster(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.names, group=BfID_18S.contigs.good.good.groups, diffs=1)

summary.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.names)

chimera.uchime(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.names, group=BfID_18S.contigs.good.good.groups, dereplicate=t)

remove.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.fasta, accnos=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.accnos, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.names, group=BfID_18S.contigs.good.good.groups, dups=T)

summary.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta , name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.names)

count.groups(group=BfID_18S.contigs.good.good.pick.groups)

get.groups(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.names, group=BfID_18S.contigs.good.good.pick.groups, accnos=BfID_18S_reduced.txt)

count.groups()

classify.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.names, group=BfID_18S.contigs.good.good.pick.pick.groups, template=silva.nr_v132.align, taxonomy=silva.nr_v132.tax, cutoff=80, processors=4)

remove.lineage(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.names, group=BfID_18S.contigs.good.good.pick.pick.groups, taxonomy=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.nr_v132.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Archaea-Bacteria-unclassified-uncultured-Vertebrata-Aspergillaceae)

count.groups(group=BfID_18S.contigs.good.good.pick.pick.pick.groups)

get.groups(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.names, group=BfID_18S.contigs.good.good.pick.pick.pick.groups, taxonomy=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.nr_v132.wang.pick.taxonomy, accnos=BfID_18S_reduced2.txt)

classify.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.fasta, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.names, group=BfID_18S.contigs.good.good.pick.pick.pick.pick.groups, template=silva.nr_v132.align, taxonomy=silva.nr_v132.tax, cutoff=80)

count.groups(group=BfID_18S.contigs.good.good.pick.pick.pick.pick.groups)

Beta Diversity

dist.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.fasta, output=phylip)

clearcut(phylip=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.dist)

phylo.diversity(tree=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.pick.phylip.tre, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.names, group=BfID_18S.contigs.good.good.pick.pick.pick.pick.groups, rarefy=T, freq=100)

//UniFrac unweighted//

unifrac.unweighted(tree=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.tre, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.names, group=BfID_18S.contigs.good.good.pick.pick.pick.pick.groups, distance=lt, random=F)

tree.shared(phylip=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.tre1.unweighted.phylip.dist)

//UniFrac weighted//

unifrac.weighted(tree=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.tre, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.names, group=BfID_18S.contigs.good.good.pick.pick.pick.pick.groups, distance=lt, random=F)

tree.shared(phylip=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.tre1.weighted.phylip.dist)

Tests for Beta Diversity

amova(phylip=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.tre1.weighted.phylip.dist, design=IB_design_file.design, sets=all, iters=50000)

pcoa(phylip=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.tre1.weighted.phylip.dist)

nmds(phylip=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.phylip.tre1.weighted.phylip.dist, mindim=3, maxdim=3, iters=5000)

Alpha Diversity

dist.seqs(fasta=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.fasta, cutoff=0.20, countends=F)

cluster(column=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.dist, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.names, method=opti, cutoff=0.20)

make.shared(list=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.opti_mcc.list, group=BfID_18S.contigs.good.good.pick.pick.pick.pick.groups, label=0.05)

list.seqs(list=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.opti_mcc.list)

get.seqs(group=BfID_18S.contigs.good.good.pick.pick.pick.pick.groups, accnos=current)

classify.otu(list=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.opti_mcc.list, name=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.names, taxonomy=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.nr_v132.wang.taxonomy, label=0.05)

make.shared(list=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.opti_mcc.list, group=BfID_18S.contigs.good.good.pick.pick.pick.pick.pick.groups, label=0.05)

collect.single(shared=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.opti_mcc.shared, calc=chao-ace-invsimpson-shannon-shannoneven, freq=10)

rarefaction.single(shared=BfID_18S.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.pick.opti_mcc.shared, calc=sobs, freq=10)

summary.single(calc=nseqs-coverage-sobs-invsimpson-shannon-shannoneven, iters=1000)

R CODES

//PCOA//

pcoaw<-read.csv(file.choose(),header=TRUE)

install.packages("ggplot2")

library("ggplot2")

require(ggplot2)

ggplot()+geom_point(aes(x=axis1,y=axis2,col=Body_Fluids),size=3,data=pcoaw) + xlab("Axis 1 (25.76%)") + ylab("Axis 2 (11.66%)") + theme() + theme(axis.line = element_line(colour = "black"),panel.grid.minor = element_blank(),panel.grid.major = element_blank())

//Boxplot//

alpha<-read.csv(file.choose(),header=TRUE)

head(alpha)

bf<-(alpha$Body_Fluid)

is<-(alpha$invsimpson)

boxplot(is~bf, main="Alpha Diversity",xlab="Body Fluids", ylab="Inverse Simpson Index",col=c("orange","red","purple"))

//Heat Map//

dlevel5<-read.csv(file.choose(),header=TRUE)

library(gplots)

library("RColorBrewer")

row.names (dlevel5) = dlevel5$taxon

dlevel5= dlevel5 [2:37]

dlevel5_matrix = data.matrix (dlevel5)

heatmap.2 (dlevel5_matrix,Rowv=FALSE, Colv=FALSE, scale="none",cexRow=0.80,cexCol=1.0,margins=c(6,15), key=TRUE, keysize= 2.0, symkey=FALSE, trace="none", density.info="none", col=brewer.pal(9,"YlGnBu"))

Vita

Ines Benaissa was born on January 20, 1995 in Constantine, Algeria and is an American citizen. She graduated from Woodland High School in Dorchester, South Carolina in 2013. She graduated in 2017 from the University of South Carolina in Columbia, South Carolina with a Bachelor of Science degree in Biology and a Chemistry minor. She is currently attending Virginia Commonwealth University in Richmond, Virginia pursuing a Master of Science degree in Forensic Science and is set to graduate May 2020. During her time at VCU in the Forensic Science department, she worked as an Office Clerk and later became a Research Associate in the Seashols-Williams and Singh laboratories performing her Directed Research project.