Identification of unique microbial signatures pre- and post-coitus in male-female pairings by massively parallel sequencing and its potential to detect sexual contact.

By

Ruby DIXON

Bachelor of Science

This thesis is presented for the degree of

Bachelor of Science Honours in Forensic Biology and Toxicology

School of Medical, Molecular and Forensic Sciences

Murdoch University, Perth

2022

Supervisors:

Brendan Chapman (Murdoch University)

Siobhon Egan (Murdoch University)

Declaration

I declare this Scientific Manuscript is my own account of my research and contains as its main content, work which has not been previously submitted for a degree at any tertiary education institution.

Ruby Dixon

Acknowledgements

Firstly, I'd like to thank my supervisors, Brendan Chapman and Siobhon Egan. Brendan, you have been a lifesaver this last year. Your support has been second to none, even during what felt like many crises. Thank you for being so approachable and helping me to grasp all the concepts and ideas that have helped shape this project. Siobhon, your sheer knowledge, support and effort in teaching me how to do a meaningful analysis has helped me to create a thesis that I am very proud of. When I say I don't know what I would have done without you, I mean it!

To my partner Leigh, you have been my greatest support. You have listened to me talk and often complain about everything going on, even when you don't know what it means. You were there to listen irrespectively and support me in whatever way you could. I would not be where I am today if it weren't for your support throughout my university journey. Even when I moved across the country to finish my degree, your support never wavered. I am eternally grateful to you.

Of course, I would also like to thank my family, who have supported me since the beginning. Thank you for giving me the kick I needed to keep going even when I felt overwhelmed. Finally, to my brother Mark. Seeing your perseverance this last year and your determination to shape a life we are all proud of has been nothing short of inspirational.

Table of Contents

List of Figures ***6***

List of Tables ***8***

Chapter 1: Literature Review ***10***

1. Introduction **10**

2. Clinical Management Post Assault **10**

3. Investigation of Sexual Assault **11**

3.1. Sexual Assault Prosecution 12

3.2. Importance of Forensic Evidence in Sexual Assault Cases 13

3.3. Detection of Spermatozoa in a Forensic Setting 14

3.3.1. DNA Analysis 14

3.4. Limitations of Current Perpetrator Identification Methods 15

3.4.1. Mixed DNA Profiles 15

3.4.2. STR Profiles from the Y-chromosome 16

3.4.3. No Recoverable Male DNA 17

4. Bacterial Profiling **18**

4.1. 16S Ribosomal RNA 19

4.2. Sequencing Techniques 20

4.2.1. Short Read Massively Parallel Sequencing 21

4.2.1.1. Illumina™ MiSeq™ System 21

4.2.1.2. Ion Torrent 16S™ Metagenomics PGM System 22

4.2.2. Long Read Massively Parallel Sequencing 22

4.2.2.1. Pacific Biosciences SMRT® Sequel II Technology 23

4.3. Microbiome Bioinformatics 24

4.4. Statistical Analysis of Microbial Communities 25

5. Human Microbiome Project **25**

5.1. Skin Microbiome 26

5.2. The Female Vaginal Microbiome 27

5.3. The Male Penile Microbiome 28

6. Forensic Microbiome Studies **29**

6.1. The “Sexome” 29

7. Conclusion **30**

8. Research Aims **31**

Chapter 2: Scientific Manuscript *49*

Abstract **49**

1. Introduction **50**

2. Materials and Methods **52**

2.1. Participant Recruitment and Selection 52

2.2. Collection of Intimate Samples 52

2.3. Bacterial DNA Extraction for MPS 53

2.4. Quality Control of DNA Extracts 53

2.4.1. PCR of 16S rRNA Gene Targeting V4 Region 54

2.4.2. Gel Electrophoresis 55

2.5. Library Preparation 55

2.6. Massively Parallel Sequencing 57

2.7. Processing of HiFi Read Files 57

2.8. Statistical Analysis 57

3. Results **58**

3.1. Participant Summary 58

3.2. 16S Microbial Diversity 60

3.2.1. Bioinformatics 60

3.2.2. Rarefaction and Alpha Diversity 60

3.2.3. Taxonomy 61

3.2.4. Beta Diversity 64

3.2.5. Shared Microbiota 65

4. Discussion **68**

4.1. Sample Diversity in Male and Female Pairs 68

4.2. Microbial Transfer Post-Coitus 72

4.3. Forensic Microbiome Studies 73

4.4. Future Research Considerations 74

5. Conclusion **74**

6. Acknowledgements **75**

Supplementary Information **81**

List of Figures

Chapter 1

Figure 1: Schematic of the human sperm cell visualised using a longitudinal cross section. Image credit Alves et al. (57) 17

Figure 2: (a) Cartoon rendering of the 70S prokaryotic ribosome. Large subunit (50S) comprised of 5S and 23S rRNA. Small subunit (30S) comprised of 16S rRNA. Image adapted from Russell et al. (67). (b) Secondary structure of the 16S rRNA of Escherichia coli including hypervariable regions V1-9. Image credit Yarza et al. (68) 19

Figure 3: Full 16S ribosomal gene depicting primer sets and approximate location of each region, V1-V9 (71-74). For each commonly targeted hypervariable region, or collective regions, the most effective massively parallel sequencing (MPS) platform for analysis is described. Image adapted from Chakravorty et al (72)..... 20

Figure 4: Bacterial distribution of different skin sites and their corresponding skin type (sebaceous, moist and dry). Image credit Grice et al. (118)..... 26

Chapter 2

Figure 1: Alpha diversity analysis in male vs female samples displayed through (a) observed ASVs, (b) Chao1, (c) Shannon and (d) Inverse-Simpson diversity measures. 61

Figure 2: Taxonomic summary of bacteria in samples displayed with a heatmap of identified taxa. Couples are displayed side by side for each time point (*before* and *after*). (a) Relative abundance of top 15 taxa grouping at a genus level (including phylum annotation). (b) Relative abundance of top 30 taxa grouping at an ASV level

(including genera annotation). Unique ASVs are indicated on the figure for the female of couple 17, along with their presence in the male sample post-coitus. 63

Figure 3: Beta diversity displayed through Principle Coordinate Analysis (PCoA) using Bray-Curtis (a) and UniFrac (b) distance measures. Clustering of couple 6 and 17 are circled in the figure. 65

Figure 4: Venn diagrams indicating overlap of taxa (at ASV level) in male and female pairings both *before* and *after* coitus. All couples are displayed. Couples 6, 9, 17 and 19 indicate an increase in the amount of shared ASVs post-coitus. 66

Supplementary Figure S1: Method for female self-sampling as per PathWest low vaginal swab technique. 81

Supplementary Figure S2: Method for male self-sampling of penile skin. 82

Supplementary Figure S3: Rarefaction curve of observed ASV reads in male and female samples. This is used to assess richness and sequencing depth. Curve plateaus at a depth of 2,200 reads. 85

Supplementary Figure S4: The taxonomic distribution of samples visualised using distribution plots for males and females at both time points (*before* and *after* intercourse). Grouped by phylum level identity. 86

Supplementary Figure S5: Bar plots displaying differential abundance of top nine genera grouped by male and female samples. Statistical analysis was performed using Wilcoxon pair-wise significance test. 87

List of Tables

Chapter 1

Table 1: Percentage of sperm positive vaginal swab samples across three different time points within a 72 hour period. The swabs include high, low and external vaginal swabs (11). 11

Table 2: Price comparison of short read and long read massively parallel sequencing platforms for 16S gene sequencing. Information was gathered directly from manufacturer websites. Ion Torrent™ has been excluded from this table as it is no longer recommended for short read analysis. Cost for Illumina™ has been converted from USD to AUD. Cost for Pacific Biosciences™ has been acquired from the Australian Genome Research Facility's (AGRF)16S Sequencing Service. 23

Table 3: Table displaying bacterial composition commonly found at the urogenital tract of males and females and shared bacteria across genders. Mean percentage contribution of bacteria in vaginal and penile samples above 0.30%. Data retrieved from two separate datasets (123, 124) investigating the microbiome of the vagina and penis in individuals with and without bacterial vaginosis. Male and female rows depict composition from highest mean contribution to lowest. 27

Chapter 2

Table 1: Bacterial 16S rRNA primers used. The V4 primers used were supplied by Sigma-Aldrich® and target a 291bp fragment of the 16S gene. Primers are compatible with Illumina™ short read sequencing platforms. Forward primer catalogue name Illumina515F16sV4. Reverse primer catalogue name Illumina806R16sV4. Primers targeting bacterial V1-9 hypervariable regions used to prepare samples for full length 16S rRNA sequencing, Primers are compatible with PacBio® long read sequencing platform. 54

Table 2: Summary of participant information (including samples from previous recruitment cohort). A more detailed table can be found in the supplementary information [Supplementary Table S1]. Information not available/unknown is marked by an asterisk (*)..... 59

Table 3: ASVs that were not identified as shared at baseline for couples 6, 9, 17 and 19. This does not include taxa that were identified as shared at both time points. Text in bold indicates ASVs that were present in the female-only at baseline. Underlined text indicates ASVs that were present in the male-only at baseline. Regular text indicates ASVs that were not detected in either male or female samples at baseline. Couples 13 and 14 were excluded from this table as there was an insignificant change in shared ASVs. Genus level identification is listed in square brackets. 67

Supplementary Table S1: Detailed summary of participant survey results from 2021 cohort..... 83

Supplementary Table S2: Summary of sequences obtained from MPS of V3-4 regions of bacterial 16S rRNA gene using Illumina™ MiSeq™ platform. 84

Chapter 1: Literature Review

1. Introduction

With the increase in new scientific discoveries, forensic investigation has become more rigorous and cutting edge. Despite this, a significant area of investigation that still yields lower conviction rates than that of any other serious crime is sexual assault (1). Sexual assault exists by many definitions, most commonly as nonconsensual sexual contact through physical force or incapacitation (2). The issue remains much more nuanced than this, however. A broader, more encompassing definition can be elucidated as any sexual contact that occurs in the absence of "active, ongoing affirmative agreement" (2, 3).

According to the Australian Institute of Health and Welfare (4), in 2018, the number of police-recorded sexual assaults against people over 15 years of age was 23.5 people per 100,000 in males and 154.4 people per 100,000 in females. While it is important to note that sexual assault occurs irrespective of gender, women are more likely to experience sexual assault than men (5). This data also detailed that in 2018-19 only 38.9% of sexual assault criminal cases were finalised by a guilty verdict. Low conviction rates are believed to be attributed to a lack of scientific evidence (6, 7) and due to the confronting legal procedure, seeking a conviction can be re-victimising for a victim (6).

In a forensic investigation, obtaining DNA from spermatozoa is a principle objective as it is critical in identifying a perpetrator. Retrieving human biological evidence is not always possible, as factors such as condom use and a lack of spermatozoa in semen (oligospermia) (8) among other medical conditions will result in the absence of DNA. This becomes particularly difficult in an investigation where the perpetrator is not known to the victim.

2. Clinical Management Post Assault

In Australia, sexual health clinics support victims of sexual assault. An example is the Sexual Assault Resource Centre (SARC) (9) which is located in Perth, Western Australia.

Victims can access their services at any point after an incident has occurred. When contact is made within two weeks post-assault, a range of emergency services are offered. A medical assessment is offered in which doctors provide free testing for pregnancy, sexually transmitted infections (STI's) and/or, blood-borne infections such as human immunodeficiency virus (HIV), hepatitis B virus and hepatitis C virus, as well as options for emergency contraception (9). Counselling services are offered to support patients throughout any stage of examination and after the emergency appointment (9).

3. Investigation of Sexual Assault

A forensic examination is offered if the victim wishes to report the incident to the police. During the forensic examination, physical evidence is collected that can be later used during legal proceedings. Biological samples are collected using a full forensic kit (FFK), which also collects toxicological samples for analysis. Collecting biological specimens are time sensitive. While semen can stay present for up to five days in the vagina, the longer the delay in collecting biological specimens, the less chance there is of obtaining a semen sample (10). A study by Casey et al (11) found that within a 24 hour period following penetrative intercourse, 34% of vaginal swabs were sperm positive. This further decreased in the 24 - 48 hour time frame to 28%. Lastly, in the 48 - 72 hour timeframe, only 6% of vaginal swabs were sperm positive [Table 1].

***Table 1**: Percentage of sperm positive vaginal swab samples across three different time points within a 72 hour period. The swabs include high, low and external vaginal swabs (11).*

Time Since Intercourse (Hours)	0-24	24-48	48-72
Positive	889	113	13
Negative	1734	294	195
Total	2623	407	208
% Positive	34	28	6

If there is a delay in procuring a suitably trained doctor or nurse to conduct a full forensic examination, patients are offered an early evidence kit (EEK) in which biological samples are self-collected (12). After post-assault medical care has been completed and the victim chooses to, a police investigation is initiated. Samples are sent for analysis by trained forensic scientists who perform DNA profiling of biological specimens and toxicological analysis. Once expert reports have been produced, all evidence is examined to assess eligibility to pursue a conviction, in which case it is referred to prosecution.

3.1. Sexual Assault Prosecution

Globally, sexual assault remains the most underreported occurrence of all serious crimes (13). An inquiry by the Victorian Law Reform Commission in 2004 found that only 17% of sexual assault victims reported the incident to police, compared to the rate of reported robbery, which sat at 50% and assaults at 30% (14, 15). Such low reporting figures can be attributed to multiple factors, namely what the victim perceives as a lack of evidence or lack of severity, the thought of not being believed and the thought of going to court to testify (16). This can also mean a delay in reporting, which is detrimental to the case if a victim has not sought help through forensic examination and reporting sooner.

An article published in 2009 stated that of the female population over the age of 18 in the United States, only 19% of victims report sexual assault to the police (17). Of this figure, only 18-44% of reported incidents were referred to prosecution, and warrants are issued in 46-72% of the referred cases. It was initially believed that sexual assaults committed by a person unknown to the victim had higher prosecution rates; however, this has since been proved false (17). In most circumstances, whether or not a case is referred to prosecution relies on the presence/severity of injuries sustained, whether a weapon was used by the assailant and witness credibility (17). These three factors do not encompass the full breadth of sexual assault cases, yet they often determine if a case will progress. When a witness's

credibility is called into question, even strong physical evidence will not necessarily equal a conviction (17).

In an Australian context, it has been found that when an offender has been identified and the case moves to trial, unless the defendant pleads guilty, they are more likely to have an acquittal outcome (15). A lack of consent and the offender's awareness of this must be proven beyond reasonable doubt during the trial. This can become ambiguous, especially when considering witness credibility and rape myths (15). Rape myths include the notion of a victim “consenting” because they did not fight or specifically say no to the sexual contact (15). A lawyer will review the information at hand and advise the defendant whether to plead guilty or not guilty. A conviction, therefore, relies heavily upon a guilty plea in this context (15).

3.2. Importance of Forensic Evidence in Sexual Assault Cases

Forensic evidence still remains a vital component of prosecution even when factors such as witness credibility and consent are called into question. The collection of high-quality forensic evidence can identify an offender based on the biological specimens left at a scene or recovered from a victim. In cases of “stranger assaults” in which the victim does not know the assailant, DNA evidence is imperative in identifying the perpetrator (18). The collection of DNA can also connect the offender to other cases of sexual assault which may have been committed by a serial offender or historical cases with a previously unidentified perpetrator.

A milestone case that holds significant value to the evolution of sexual assault investigation was that leading to the conviction of Colin Pitchfork. In 1983 and 1986, two underage women from the Leicestershire area in England were raped and murdered (19). DNA evidence was collected from both scenes and later profiled after a new technique of DNA fingerprinting was developed (20). Once the DNA was profiled, the evidence indicated that a single offender was responsible for both crimes, which were previously thought to be unrelated (21). After a mass collection effort across the region, DNA evidence eventually led to Mr Pitchfork. Forensic DNA evidence was instrumental in the arrest and conviction of Mr Pitchfork

and the exoneration of Richard Buckland, who was initially the main suspect (20). Without forensic DNA evidence, rates of unidentified sexual offenders would be far more significant, as would the rate of unprosecuted sexual assault crimes (22-24).

3.3. Detection of Spermatozoa in a Forensic Setting

In sexual assault investigation, spermatozoa from semen is the primary source of DNA evidence and isolating it from a scene, or evidence kit is paramount. Current methods for detecting the presence of semen in a crime scene or biological sample include a range of presumptive assays, confirmatory assays and DNA short tandem repeat (STR) typing (25). In the context of a crime scene evaluation, a visual examination is performed initially, using an approximately 450nm wavelength alternative light source along with an orange barrier filter (26). At this wavelength, the excitation of flavin and choline-conjugated (27) molecules in semen emit light that appears as a fluorescence, illuminating seminal fluid that is not visible to the naked eye (28). A presumptive test can be carried out once a stain of interest has been identified. The most common presumptive test for identifying seminal fluids is the acid phosphatase (AP) test (25). If a positive result is obtained, the sample can be collected for further confirmatory testing and STR typing. Current confirmatory testing is ideally completed in a laboratory setting. It involves the identification of spermatozoa under a microscope through staining and visual examination (29) or by rapid antigen tests such as RSID-semen (30).

3.3.1. DNA Analysis

After the presence of spermatozoa has been confirmed, DNA can be extracted for STR typing. A widely adopted method involves the use of ionic detergent, sodium dodecyl sulphate (SDS), Proteinase K (Pro-K) and dithiothreitol (DTT) (31). Following extraction, the sample is quantified to determine the quantity of DNA present, typically by a quantitative polymerase chain reaction (qPCR). Current STR analysis kits are optimised for DNA concentrations

between 0.5 - 2ng of DNA (32). Amplification is carried out using STR kits that specifically target certain loci on chromosomes known to contain microsatellite markers (33). Traditional DNA typing involves the use of primers that are carefully multiplexed and carry a range of fluorescent markers that are used to separate target regions by size and fluorophore during capillary electrophoresis (34). The variation in STR length (length polymorphism) on each target loci is what forms an allele, and within each STR marker, there is a pair of alleles. Across a person's entire genotype, there are various STR markers that can be targeted depending on the STR amplification kit used. Examples include highly optimised PowerPlex® (Promega) STR kits which can target up to 26 loci (35) while GlobalFiler® (Applied Biosystems) STR kits can target up to 21 loci (36). The Australian National Criminal Investigation DNA Database (NCIDD) uses 18 core loci (37) for generating their DNA database while the Federal Bureau of Investigation's (FBI) Combined DNA Index System (CODIS) uses 20 core loci (38). Statistical examination of a person's unique profile typically yields somewhere in the order of a 1 in 1 billion chance of having the same profile as any other person in the population (39), making it a valuable piece of evidence for forensic investigation.

3.4. Limitations of Current Perpetrator Identification Methods

While current methods of forensic DNA profiling have proven to be of significant value to the investigation and prosecution of criminal offenders, they are not without limitations. A frequent occurrence in samples obtained from a sexual assault victim is the presence of mixed DNA or unrecoverable male DNA.

3.4.1. Mixed DNA Profiles

The most prolific limitation in DNA profiling is the existence of mixed DNA samples, which is recurrent in sexual assault cases. The presence of female DNA in the samples obtained from FFK's and EEK's can affect the ability to obtain a correct male DNA profile as the female DNA overwhelms the presence of male DNA (40). Due to the nature in which a

sample is obtained, usually through post-assault examination of a female victim, mixed DNA is inevitable. To counter this issue, a differential extraction method is employed to separate the male sperm cells from other DNA containing cells in the sample (41). As previously discussed, DNA analysis for sexual assault samples involves the use of ionic detergent, SDS, Pro-K and DTT (31). The sample is first subject to a preferential lysis which breaks open any non-sperm cells (42). Next, the cells are digested using Pro-K and SDS which leaves the sperm cells untouched (42). After the digestion process, the sample is subject to a number of centrifugation and washing cycles that aim to remove any non-sperm DNA contaminants from the sample (43). Once this process is completed, the sperm cells are lysed using an ionic detergent, Pro-K and DTT (44). The remaining sample contains the perpetrators sperm cell fraction for further STR typing. While this method is widely adopted, it is not without its limitations. This method of separation is laborious and can take up to 8 hours to complete which contributes greatly to the backlog issue with sexual assault investigation (45). A study by Gill et al (46) found that the differential method for isolating male sperm fractions from a vaginal swab had a mean success rate of 62%. It has also been found that the conventional differential method can cause a loss of 60-90% of the male DNA in a sample (40, 44, 47, 48). A method to selectively extract DNA from male spermatozoa has not yet been completely refined.

3.4.2. STR Profiles from the Y-chromosome

One of the to avoid mixed DNA samples is through the analysis of STRs located solely on the Y-chromosome (Y-STR) (49, 50). This method has effectively bypassed issues associated with mixed DNA samples as the male karyotype has an XY chromosome configuration and females have a pair of X chromosomes, or XX (51). Therefore, females will not generate a Y-STR profile. Again, this method is not without its obstacles. Due to the paternal inheritance pattern of the Y chromosome, Y-STR haplotypes for related males will be the same (52). This can cause issues, especially when there is a succession of males in a

single-family tree, as the Y-STR result obtained from a sample cannot differentiate between related male individuals (53). For example, brothers will exhibit the same Y-STR haplotype as their father, paternal grandfather and paternal uncles (among others). Consequently, these profiles are not as informative as STR profiles; however, they are valuable in circumstances where female DNA overwhelms the presence of male DNA (54).

3.4.3. No Recoverable Male DNA

DNA from spermatozoa is found in the nucleus located in the head of the sperm cell [Fig 1] (55). The DNA found in the sperm head is highly dense and tightly packed to maximise the DNA transported to the egg during fertilisation (56). Due to the high concentration of DNA in sperm cells, they are the ideal target for DNA analysis in sexual assault biological specimens. It holds the greatest chance of recovering a male STR profile when compared to the male epithelial cells.

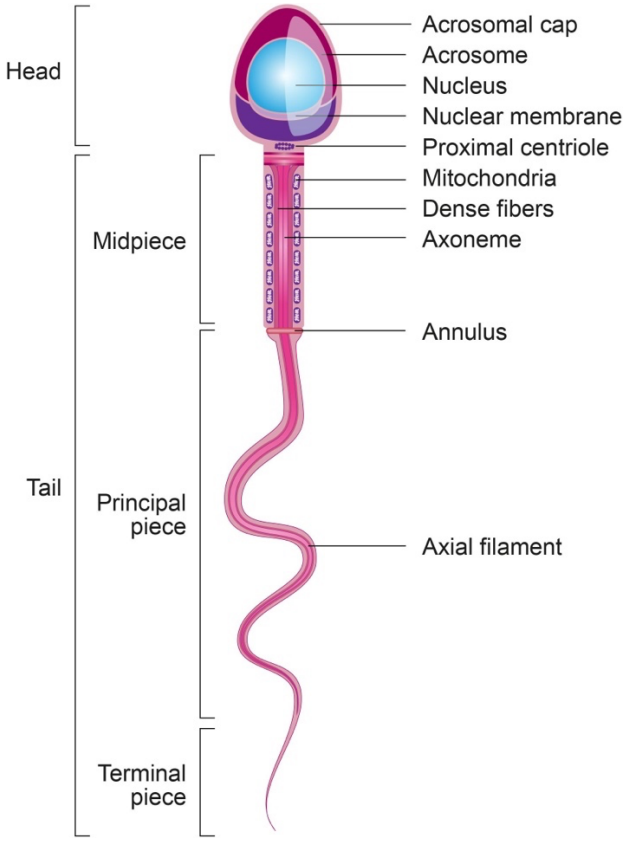


***Figure 1**: Schematic of the human sperm cell visualised using a longitudinal cross section. Image credit Alves et al. (57)*

There are instances where no male DNA can be recovered from a biological sample. The use of a condom during intercourse is 96-99% effective in preventing pregnancy (58) by controlling the amount of sperm left by the male individual. This is through capture of seminal fluids by the condom, preventing it from entering the vaginal opening. Based on this, when condoms are used during a sexual assault, it prevents the spermatozoa from being collected post-assault. Males can also suffer from medical conditions such as azoospermia (absence of semen in ejaculate) and oligospermia (deficiency of semen in ejaculate) which greatly reduces the availability of male DNA isolated from sperm cells (54). A vasectomy, which is the male sterilisation procedure, will also reduce the presence of semen after ejaculation (54).

4. Bacterial Profiling

Microorganisms are ubiquitous with life and the microbiome is defined as the collection of genes from microbiota including bacteria, fungi, protozoa and viruses (59). Historically, there has always been a fundamental need to understand and identify bacteria. Scientists first developed a technique of categorisation based on morphological features and behaviour. This was critical in relation to public health and remains an active area of research (60). Techniques to accurately identify bacteria to species and strain level have become more accessible as new DNA sequencing techniques are introduced and evaluated (61).

Traditional culturing techniques used to identify bacteria, became of little value when faced with a bacterium that could not be grown through typical methods (62). Additionally, given the diversity of different areas of the body, cultivation methods do not provide the high throughput needed and are selective based on the target bacterium. As such, culture-based methods are not useful at cataloguing entire populations. After the Human Genome project (63) reached completion, bacterial sequencing became more readily available and cost effective. Through investigation of bacterial genes, RNA from the 16S ribosomal subunit (16S rRNA) was proven to be the most valuable source of diversity for phylogenetic analysis (64).

4.1. 16S Ribosomal RNA

The ribosome is an organelle found in cells that functions to synthesise proteins from messenger RNA (mRNA) (65). In prokaryotic organisms, it is comprised of a large and small subunit, 50S and 30S respectively. Within the 50S subunit lies 5S and 23S rRNA while the 30S subunit contains 16S rRNA [Fig 2a]. The 16S rRNA is responsible for binding to Shine-Dalgarno sequences on mRNA, which is crucial for the initiation of translation and protein biosynthesis (65, 66). It also facilitates the binding of the small and large subunits by interacting with the 23S rRNA located in the large subunit (65).

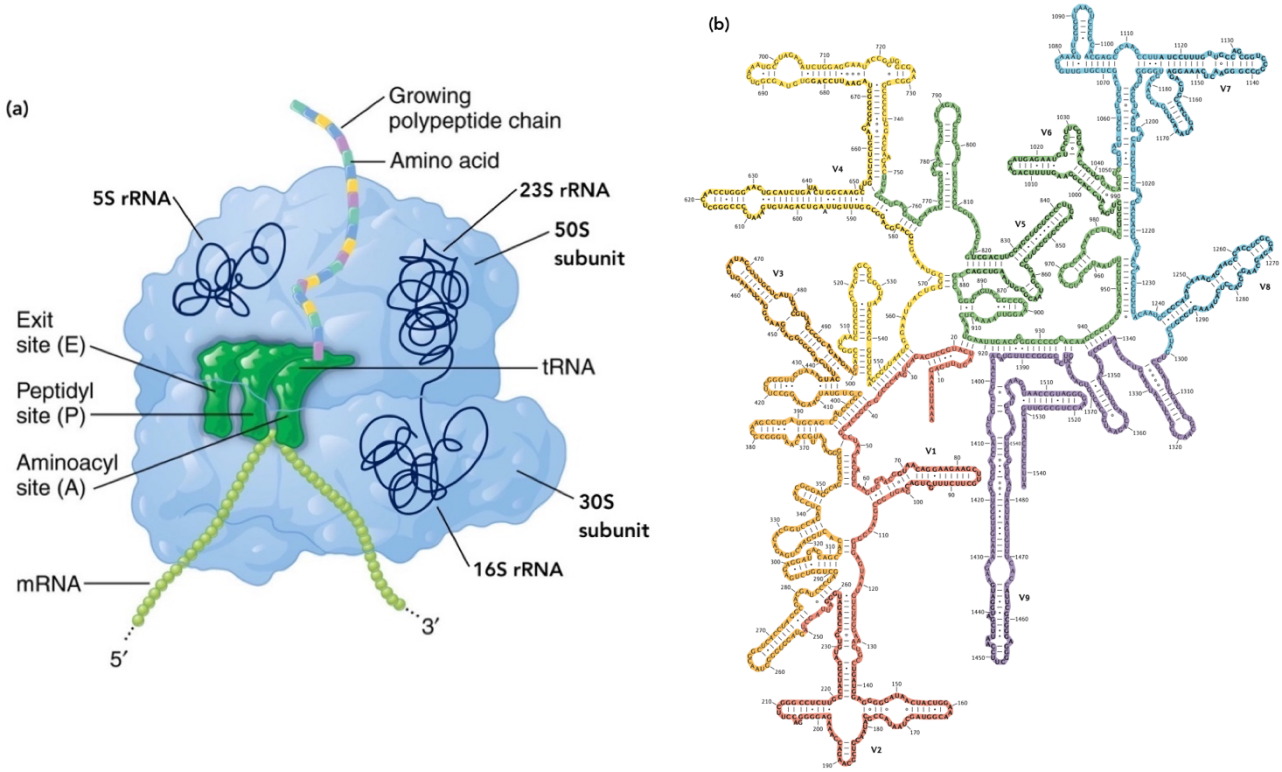


***Figure 2**: (a) Cartoon rendering of the 70S prokaryotic ribosome. Large subunit (50S) comprised of 5S and 23S rRNA. Small subunit (30S) comprised of 16S rRNA. Image adapted from Russell et al. (67). (b) Secondary structure of the 16S rRNA of Escherichia coli including hypervariable regions V1-9. Image credit Yarza et al. (68)*

Bacterial 16S ribosomal RNA is useful for analysis as the gene contains regions which are highly conserved across bacteria taxa (69), and has a relatively slow evolutionary rate (70). This gene also contains variable regions that can distinguish between bacterial species

along with flanking regions that are conserved between genera and provide primer binding opportunities. There are nine hypervariable regions of the 16S rRNA that can be targeted during sequencing [Fig 2b, Fig 3].

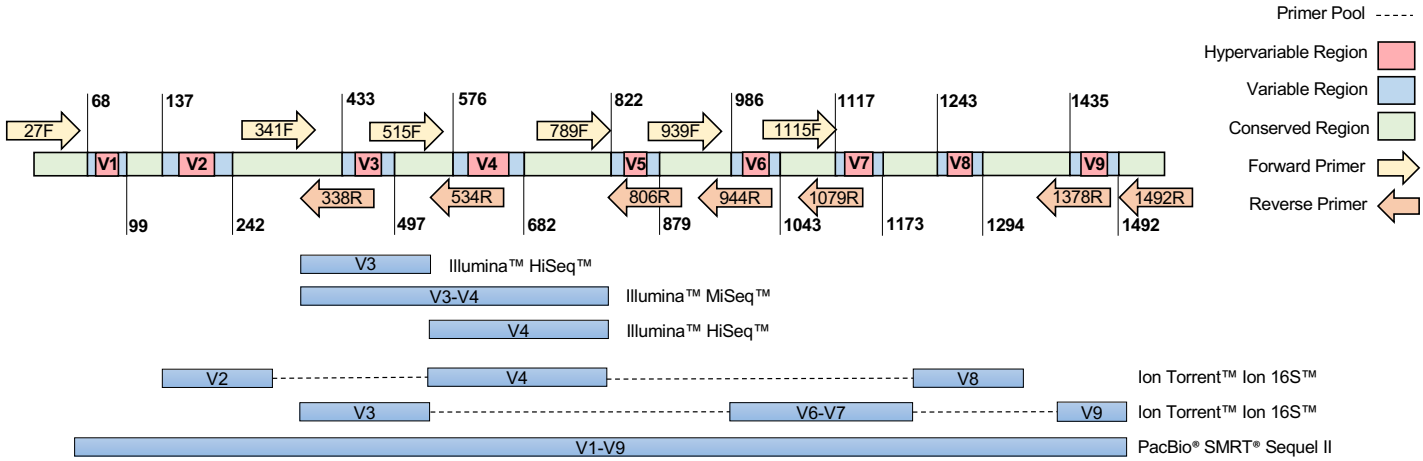


Figure 3: *Full 16S ribosomal gene depicting primer sets and approximate location of each region, V1-V9 (71-74). For each commonly targeted hypervariable region, or collective regions, the most effective massively parallel sequencing (MPS) platform for analysis is described. Image adapted from Chakravorty et al (72).*

4.2. Sequencing Techniques

Sanger sequencing was once the preferred method for identification of microorganisms but this has now been overlapped by short and long read technology. While traditional sanger sequencing remains widely used, it lacks the ability to identify more than one species per primer pair making it time consuming and laborious for investigating entire microbial communities (75). Next generation sequencing (NGS), also known as massively parallel sequencing (MPS), which it will be referred to herein, have revolutionised the genomic analysis of complex sample mixtures.

4.2.1. Short Read Massively Parallel Sequencing

Now referred to as “second-generation” sequencing, short read sequences allows the generation of millions of sequences in parallel. Bacterial short read sequencing targets partial fragments of the 16S rRNA gene. This technique uses PCR with specific primers to target the region of interest. Short read sequencing has been instrumental in microbiome research due to the low cost and high throughput compared to traditional Sanger sequencing (76). The introduction of short read sequencing to microbiota analyses resulted in a greater depth of sequencing, which has permitted the recognition of rare populations that appear in low abundance (77). Commonly targeted variable regions for short read microbiome studies include the V1-2, V3-4 and V4 regions [Fig 3] (78). While the taxonomic classification and resolution is variable among different sample types, most short read sequence studies are limited to family or genus level classification (78). This is a significant limitation of short read sequencing because it is challenging to obtain a species identification in bacteria that has not been previously characterised. The consequence of this is that there is a large proportion of database entries that remain unclassified. Many of these entries are predicted to be artifacts caused by the formation of chimeric sequences (79). A lack of adequate databases that contain true and accurate data further inhibits species identification. An amalgamation of all these factors influences the ability to categorise bacteria at a species level using short read analysis, further reinforcing the need for a sequencing technique able to accomplish this.

4.2.1.1. Illumina™ MiSeq™ System

MPS technology has been at the forefront of microbiome research in the past decade, and the technology has facilitated accurate data production and has been extensively and critically evaluated. The Illumina™ MiSeq™ sequencing platform can generate up to 25 million reads (15Gb of sequence data) and has a maximum sequencing length of 600 bp (2 x 300 bp paired end) (80). This system is commonly used to amplify regions V3-4 of the 16S rRNA targeting an approximately 465 bp fragment for short read analysis. Sample preparation for

microbiome studies using the MiSeq™ platform relies on a PCR amplification which includes primer sets 341F (5` - CCTACGGGNGGCWGCAG - 3`) and 806R (5` - GGACTACNVGGGTWTCTAAT – 3`) (72). Illumina™ sequencing technology works by visualising fluorescently labelled terminators after the addition of each deoxynucleoside triphosphate (dNTP) (81). This method reduces errors through the base-by-base protocol, enabling the generation of accurate data (81).

4.2.1.2. Ion Torrent 16S™ Metagenomics PGM System

ThermoFisher Ion 16S™ Metagenomic technology targets seven hypervariable regions V2-9 for analysis using two different primer pools (82). Where Illumina™ and many other sequencing technologies rely on light emission to detect nucleotide base addition during sequencing by synthesis, Ion Torrent™ is light independent (83). Instead, it detects the change in pH due to hydrogen ion release as nucleotides are incorporated into wells within the system (83, 84). This light independent technology has been shown to have a higher error rate than that of Illumina™ technology making it less preferable for use in 16S gene analysis (85).

4.2.2. Long Read Massively Parallel Sequencing

Long read sequencing, or third-generation sequencing, overcomes limitations of short-read sequencing while still maintaining a high-throughput platform capable of generating millions of sequences. In recent years more microbiome studies have adapted to these platforms for the applications of 16S rRNA sequencing. This technique takes into account the full length of the 16S rRNA gene by using primers that amplify regions V1-9, in an attempt to expand taxonomic classification ability (76). Previous research looking at short reads has found it challenging to generate a species-level identification of bacterial communities. Identifying the species of bacteria is paramount in emerging forensic microbiome studies as there is a need to resolve species that may be unique to an individual.

4.2.2.1. Pacific Biosciences SMRT® Sequel II Technology

Pacific Biosciences (PacBio®) Single-Molecule Real-Time (SMRT®) Sequel II System is the latest in third-generation sequencing. This technology allows for the analysis of up to 192 16S rRNA samples in parallel. Purified DNA samples undergo a two-step PCR where amplicons are barcoded and SMRTbell™ adapters are attached before amplicons are pooled (86). The SMRTbell™ adapter sequences allow for the formation of circular consensus sequences (CCS) which allow each nucleotide position to be read approximately 10 times within each amplicon (87). The core equipment used for Sequel II systems is the SMRT Cell. It contains 8 million wells called zero-mode waveguides (ZMWs) and single molecules of DNA are immobilized in each well (88). As each base is incorporated into the molecule, light is emitted and measured by the system (89). Errors with the SMRT sequencing platform are randomly distributed across reads and as such consensus sequences can still be achieved with adequate read depth (90). However, because each amplicon in a CCS is read more than once, the likelihood of an error being repeated for multiple reads is unlikely (90). The cost of this new technology is still relatively high for 16S analysis [Table 2], and read accuracy is analogous to that of short read analysis.

Table 2: *Price comparison of short read and long read massively parallel sequencing platforms for 16S gene sequencing. Information was gathered directly from manufacturer websites. Ion Torrent™ has been excluded from this table as it is no longer recommended for short read analysis. Cost for Illumina™ has been converted from USD to AUD. Cost for Pacific Biosciences™ has been acquired from the Australian Genome Research Facility's (AGRF)16S Sequencing Service.*

Sequencing Platform		Maximum Samples per Run	Maximum Reads per Run	Maximum Read Length for 16S	Cost Calculation	Cost	Base Call Accuracy	References
Illumina™	MiSeq v3	96	25 Million	600 bp (2 x 300 bp)	• MiSeq System • 96 samples • 2 × 300 bp read length • Nextera XT index primers • MiSeq Reagent v3 600-cycle kit	$26 AUD per Sample $2511 AUD per Run $4.19 per Base Pair	99.9% at Q30	(91-93)

Pacific Biosciences™	SMRT Sequel II	192	4 Million	1500 bp	• Sequel II System • HiFi reads • 192 samples • 1500 bp read length • Universal barcoded primers • Template Prep • 1 x SMRT Cell	$68 AUD per Sample $13056 per Run $8.70 per Base Pair	99.95% at Q33	(94, 95)

4.3. Microbiome Bioinformatics

Analysis of human microbiome data is constantly evolving, and much of the analysis has evolved from ecological studies. Commonly used pipelines to analyse microbiome data include DADA2 (96), QIIME2 (97), USEARCH (98) and MOTHUR (99). Raw data generated from short and long read sequencing must first be demultiplexed, a process which allows separation of sequences per unique sample, and for paired end analysis sequences must also be merged. Following this, sequences are quality filtered where chimeras (non-biological sequences) and low abundant reads are removed (100). Finally, sequences are either clustered or denoised to generate a list of final sequences that are taxonomically classified. A common approach is to run through a denoising algorithm to remove PCR sequencing errors (101) producing amplicon sequence variants (ASVs). This process can differentiate between sequences with only one nucleotide base difference, and common systems for denoising include DADA2 and IPED (102). Before denoising to produce ASVs became the standard across microbiome research, analysis was based on clustering sequences according to their likeness with a threshold of 97%+ similarity (103). These clusters are referred to as operational taxonomic units (OTUs), which are further compared with reference databases to deduce possible taxonomy (101). An ASVs approach removes the risk of coupling species together due to their sequence similarity providing higher resolution. Taxonomy is subsequently classified using a reference database such as SILVA (104) or Greengenes (105), which is further cross referenced with BLAST analysis through the National Centre for Biotechnology Information (NCBI) database (105, 106).

4.4. Statistical Analysis of Microbial Communities

Downstream analysis of microbiome data varies depending on the study aim; however, it is common for studies to include taxonomic or composition summary and comparison of diversity within and among samples. Alpha diversity analysis looks at the richness and diversity within a sample (107). The most simplest measure is the number of taxa (e.g. OTUs or ASVs) present, but can also include indexes such as Choa1, Shannon and Inverse Simpson measures (108). Beta diversity analysis assesses the variation of bacterial communities between samples, and is often combined with ordination analysis. Beta diversity measures include (i) Bray-Curtis dissimilarity which incorporates sequence abundance information (109), (ii) Jaccard index which assesses only presence/absence (110) and (iii) UniFrac (111) measure which uses phylogenetic information to compare samples. To visualize similarity of sample group these measures can be displayed using principal coordinate analysis (PCoA) (108, 111) and non-metric multidimensional scaling (NMDS) (112). Additional statistical analysis such as correlation, differential abundance and network analysis can also useful to identify biomarkers or predictors of the microbiome (111). These methods are particularly useful where only a small number of bacteria taxa are correlated between samples, this information can get lost when using diversity measures.

5. Human Microbiome Project

The human microbiome project (HMP) brought about the need to understand the bacterial communities that co-inhabit the body (113) and was the natural progression from the Human Genome Project (HGP). The data generated from the HMP research has shown the diverse microbial community among body sites. The creation of MPS technology allowed for high-throughput analysis of the complex microbial communities in the human body, rather than the characterisation of single organisms in isolation (114). The microbes the co-exist with the body have been linked to health as well as disease. Across the entire project, a large emphasis has been on the gastrointestinal (GI) microbiome. Low diversity within the GI

microbiome has been linked to obesity and inflammatory bowel disease (115, 116). The skin microbiome is another focus of the HMP due to the nature of pathogenetic bacteria that exist in this environment. *Staphylococcus aureus* is commonly found on the skin, and can reside there without causing illness. As soon as it breaches the skins surface and enters the body, it is capable of causing severe disease such as pneumonia and sepsis (117). Naturally, this brought forth the need to understand the skin microbiome and remains a large area of research.

5.1. Skin Microbiome

From the moment of birth, the skin becomes completely colonised by microorganisms and continues to change immensely during the first year of life (60). The diverse community of microorganisms that occupy the skin environment change depending on the conditions they thrive in. The bacterial composition changes between sebaceous, dry and moist environments [Fig 4].

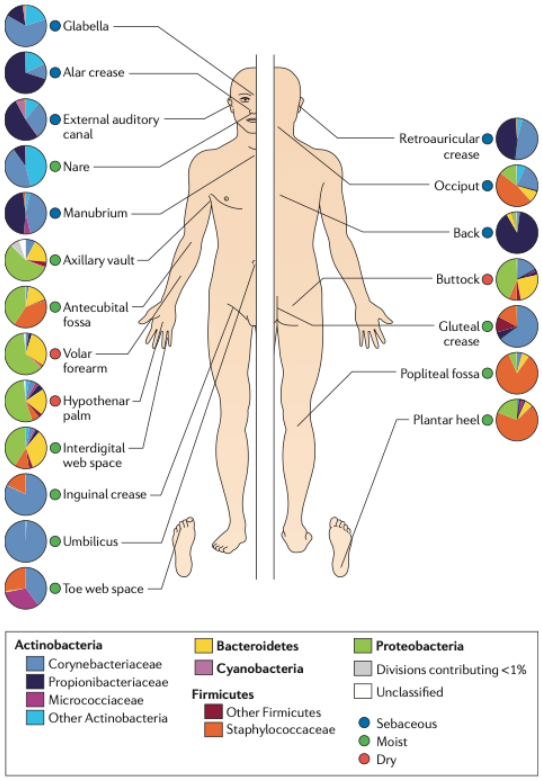


***Figure 4**: Bacterial distribution of different skin sites and their corresponding skin type (sebaceous, moist and dry). Image credit Grice et al. (118).*

While the skin microbiome varies among body site and between individuals, it remains relatively stable over time (119). Microbes have a huge impact on skin health and as such have been extensively studied. The most dominant phyla associated with the skin microbiome are Actinobacteria (51.8%), Firmicutes (24.4%), Proteobacteria (16.5%), and Bacteroidetes (6.3%) (119). Of these, the genera which form the majority of the skins microbiome is *Corynebacterium*, *Propionibacterium* and *Staphylococcus (119)*. *Staphylococcus epidermis* is a bacterium commonly attributed to skin health and protection from pathogenetic microbes. Acne present in pubescent children is often attributed to the presence of *Propionibacterium acnes*.

5.2. The Female Vaginal Microbiome

Research into the bacterial communities present in the female urogenital tract have been thoroughly investigated due to the relative importance of microbial health for women during pregnancy (120). The vaginal microbiome mainly consists of bacteria belonging to the *Lactobacillus* genus (121). In healthy women, the most common bacterial species include *Lactobacillus crispatus, Lactobacillus gasseri, Lactobacillus iners and Lactobacillus jensenii (122).* Individuals with bacterial vaginosis also commonly exhibit the presence of *Gardnerella vaginalis* and *Atopobium vaginae (122).* While these six bacterial species are the most documented in relation to the vaginal microbiome, other genera have been identified as co-inhabiting the vaginal region [Table 3] (123).

***Table 3**: Table displaying bacterial composition commonly found at the urogenital tract of males and females and shared bacteria across genders. Mean percentage contribution of bacteria in vaginal and penile samples above 0.30%. Data retrieved from two separate datasets (123, 124) investigating the microbiome of the vagina and penis in individuals with and without bacterial vaginosis. Male and female rows depict composition from highest mean contribution to lowest.*

	Genus level identity	Species level identity
Shared bacteria across genders	*Lactobacillus, Staphylococcus, Streptococcus, Prevotella Escherichia, Gardnerella*	*Lactobacillus iners*
Female specific bacteria	*Ureaplasma, Shuttleworthia, Veillonella, Sneathia, Atopobium, Klebsiella, Bifidobacterium, Megasphaera*	*Lactobacillus crispatus, Lactobacillus gasseri, Lactobacillus jensenii*
Male specific bacteria	*Corynebacterium, Peptoniphilus, Anaerococcus, Porphyromonas, Proteus, Enterobacter, Facklamia, Dialister*	*Prevotella bivia, Prevotella disiens*

After evaluating the contribution of bacteria across multiple vaginal samples in this dataset, the mean contribution of *Lactobacillus* is far greater than any other bacterium at 64.25% (123). *Streptococcus* and *Provotella* are the next most abundant genus with 9.58% and 6.96% mean contribution, respectively (123). Given the abundance of these genus and the sex specific species identified, there may be future opportunities to differentiate an individual donor based on sub-species or strain level identification of microbiota.

5.3. The Male Penile Microbiome

The male urogenital tract microbiome has not been as thoroughly investigated as the female genital region. Studies investigating the microbiome of penile skin found that the most abundant bacteria were *Corynebacterium*, followed closely by *Peptoniphilus*, *Anaerococcus* and *Staphylococcus (123, 124)*. This indicates a similarity to the skin microbiome [Section 4.1] with the only difference being a change in percentage composition of taxa. The male urogenital tract is classified as a moist environment [Fig 4]. With reference to the female vaginal microbiome, the male penile microbiome differs in diversity. When considering the most abundant genera, there is no substantial overlap between male and female genital microbiome. In a forensic context, this is particularly meaningful because it presents an opportunity to investigate microbiome transfer between individuals.

6. Forensic Microbiome Studies

Forensic applications of the human microbiome are emerging and remain an active area of research. Previous studies of the human GI microbiome have indicated that the subspecies level of bacterial taxa are unique between individuals (125). This presents a research opportunity to ascertain whether other areas of the human body contain unique bacterial subspecies that could aid in the identification of an individual based on their microbiome. If this hypothesis can be validated, there is the potential to be able to use the human microbiome to link a suspect/s to criminal activities (126). A pilot study by Procopio et al (127) investigated the transferability of fingerprint microbes to a glass slide with the aim to imitate “trace evidence” collected from a crime scene. They reported a reduced number of ASVs extracted from the fingerprint sample deposited on the glass slide when compared with the corresponding skin swab sample. They inferred that the microbiome from the fingerprint did transfer but did not represent the entire microbial community. A study by Neckovic et al (128) reported that in non-coinhabiting pairs the skin microbiome could transfer between two individuals hands, disrupting the bacterial communities of specific skin locations. Direct skin-to-skin contact facilitated this transfer, while indirect transfer was not conducive to the disruption. Both of these studies confirmed that the transferability of the human microbiome and suggested further research to investigate subspecies identification of bacterial taxa for use in forensic investigation.

6.1. The “Sexome”

The female genital microbiome has been extensively studied to date, but at this point, none have investigated the microbiota during sexual intercourse or specifically aimed to determine pre and post-coitus transfer. Several studies have investigated the microbiota of co-inhabiting and non-coinhabiting couples (128, 129). Previous research has determined that couples share bacterial communities but hypothesised that this was due to environmental factors and not direct contact (130). The study did not set parameters for participants to adhere

to in order to gain data that simulated a once off sexual experience, and swabs were taken from the pubic mound rather than directly from the genitalia. Data regarding sexual activity prior to the samples being collected varied, and couples who were more sexually active *before* sampling, exhibited a higher percentage of cluster sharing (130). This area of research holds significant value in forensic medicine. It could provide a means of identifying an offender in sexual assault cases where the perpetrator may not be known, denies the contact occurred or a male STR profile was not obtained. Further research into the detection of unique microbial signatures post coitus are important for the forensic development of sexual assault investigation.

7. Conclusion

The future of microbiome research lies in exploring the uniqueness of an individual's microbial communities. Specific applications within the forensic field would be advantageous in sexual assault investigations. Currently, there are many limitations to typical methods of perpetrator identification in sexual assault cases. Namely, the suboptimal method for isolating DNA from spermatozoa in mixed biological samples, the paternal inheritance of Y-STR profiles and the inability to recover spermatozoa in a sample. Investigation of the human genital microbiome and the ability to detect unique microbial signatures post intercourse offers an adjunct method to aid in perpetrator identification. For microbiome analysis to be used in a forensic context, a method to accurately provide species, sub-species and strain level classification is needed. Long-read analysis of the 16S rRNA gene using the PacBio SMRT Sequel II system has greater taxonomic ability while maintaining the high accuracy and throughput of short-read analysis. The application of long-read analysis to detect unique microbial signatures that indicate sexual contact provides a progressive step in the development of forensic microbiome research.

8. Research Aims

The pilot "Sexome" study by Nye involved the investigation of bacterial communities through short read 16S rRNA analysis of the hypervariable regions V3-4 (131). Data generated could not classify bacterial communities with greater resolution than family or genus level. This research project aims to characterise the bacterial communities from vaginal and penile samples of participants using long read analysis of the V1-9 regions of 16S rRNA in an attempt to expand taxonomic classification. Further, it aims to compare differences in bacterial communities present on penis (males) and vagina (females) samples both pre and post intercourse. More specifically for forensic applications, this study aims to identify specific bacterial sequences (using an ASV approach) that are unique to each individual and identify specific bacterial species theorised to have transferred through intercourse within each male-female pairing.

From reviewing the literature associated with the human reproductive microbiome and its relevance to forensic investigation, the research hypotheses for this study are:

1. There are bacterial taxa found in the genital region that are statistically unique to each sex prior to sexual intercourse.
2. There are bacterial taxa found in the genital region that are statistically unique to each individual prior to sexual intercourse.
3. Within each male-female pairing, there is a transfer of the statistically identifiable male genital microflora to the female (or inverse) that can be detected from post-coitus low vaginal or shaft and glans sampling.

Word Count: 6,344

References

1. Larcombe W. Falling Rape Conviction Rates: Feminist Aims and Measures for Rape Law. Feminist legal studies. 2011;19(1):27-45.

2. Graham LM, Treves-Kagan S, Magee EP, DeLong SM, Ashley OS, Macy RJ, et al. Sexual Assault Policies and Consent Definitions: A Nationally Representative Investigation of U.S. Colleges and Universities. Journal of school violence. 2017;16(3):243-58.

3. Cantor D, Fisher B, Chibnall S, Townsend R, Lee H, Bruce C, et al. Report on the AAU Campus Climate Survey on Sexual Assault and Sexual Misconduct. Rockville, Maryland: The Association of American Universities; 2015.

4. Welfare AIoHa. Sexual assault in Australia2020. Available from: https://www.aihw.gov.au/getmedia/0375553f-0395-46cc-9574-d54c74fa601a/aihw-fdv-5.pdf.aspx?inline=true.

5. Statistics ABo. Sexual Violence - Victimisation 2021 [Available from: https://www.abs.gov.au/articles/sexual-violence-victimisation.

6. Martin LJ. Forensic evidence collection for sexual assault: a South African perspective. International journal of gynecology and obstetrics. 2002;78(1):S105-S10.

7. Davis RC, Wells W. DNA testing in sexual assault cases: When do the benefits outweigh the costs? Forensic science international. 2019;299:44-8.

8. Eisenberg MLMD, Chen ZPD, Ye APD, Buck Louis GMPD. Relationship between physical occupational exposures and health on semen quality: data from the Longitudinal Investigation of Fertility and the Environment (LIFE) Study. Fertility and sterility. 2015;103(5):1271-7.

9. Health Do. Sexual Assault Resource Centre (SARC) Perth, Western Australia: Government of Western Australia; 2021 [updated 17/12/2021.

10. Freedman E. Clinical management of patients presenting following a sexual assault. Australian journal of general practice. 2020;49(7):406-11.

11. Casey DG, Domijan K, MacNeill S, Rizet D, O'Connell D, Ryan J. The Persistence of Sperm and the Development of Time Since Intercourse (TSI) Guidelines in Sexual Assault Cases at Forensic Science Ireland, Dublin, Ireland. Journal of forensic sciences. 2017;62(3):585-92.

12. Centre SAR. Procedure for Responding to a Recent Sexual Assault. Department of Health; 2017.

13. Welfare AIoHa. Family, domestic and sexual violence in Australia: continuing the national story 2019. Canberra: AIHW; 2019.

14. Friedman N, Langan A, Little H, Neave MA. Sexual Offences: Law and Procedure Final Report. Melbourne, Australia; 2004.

15. Naylor B. Effective justice for victims of sexual assault : taking up the debate on alternative pathways. University of New South Wales law journal. 2010;33(3):662-84.

16. Khan SR, Hirsch JS, Wamboldt A, Mellins CA. "I didn't want to be 'that girl'": The social risks of labeling, telling, and reporting sexual assault. Sociological science. 2018;5(19):432-60.

17. Campbell R, Patterson D, Bybee D, Dworkin ER. Predicting Sexual Assault Prosecution Outcomes: The Role of Medical Forensic Evidence Collected by Sexual Assault Nurse Examiners. Criminal justice and behavior. 2009;36(7):712-27.

18. Henry TKS, Jurek AL. Identification, Corroboration, and Charging: Examining the Use of DNA Evidence by Prosecutors in Sexual Assault Cases. Feminist criminology. 2020;15(5):634-58.

19. Phillips ML. Crime Scene Genetics: Transforming Forensic Science through Molecular Technologies. Bioscience. 2008;58(6):484-9.

20. Jobling MA, Gill P. Encoded evidence: DNA in forensic analysis. Nature reviews Genetics. 2004;5(10):739-51.

21. Goodwin W. DNA profiling: The first 30 years. Science & justice. 2015;55(6):375-6.

22. Wilson DB, Weisburd D, McClure D. Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. Campbell systematic review. 2011;7(1):1-53.

23. Makasa I, Heathfield LJ. A retrospective study of sexual offences in Zambia. Journal of forensic and legal medicine. 2018;54:23-33.

24. Wilson DB, McClure D, Weisburd D. Does Forensic DNA Help to Solve Crime? The Benefit of Sophisticated Answers to Naive Questions. Journal of contemporary criminal justice. 2010;26(4):458-69.

25. Sakurada K, Watanabe K, Akutsu T. Current Methods for Body Fluid Identification Related to Sexual Crime: Focusing on Saliva, Semen, and Vaginal Fluid. Diagnostics (Basel). 2020;10(9):693.

26. Nolan A, Speers SJ, Murakami J, Chapman B. A pilot study: The effects of repeat washing and fabric type on the detection of seminal fluid and spermatozoa. Forensic science international. 2018;289:51-6.

27. Sheppard K, Cassella JP, Fieldhouse S, King R. The adaptation of a 360° camera utilising an alternate light source (ALS) for the detection of biological fluids at crime scenes. Science & justice. 2016;57(4):239-49.

28. Kobus HJ, Silenieks E, Scharnberg J. Improving the effectiveness of fluorescence for the detection of semen stains on fabrics. Journal of forensic sciences. 2002;47(4):819.

29. Deshmukh S, Inci F, Karaaslan MG, Ogut MG, Duncan D, Klevan L, et al. A confirmatory test for sperm in sexual assault samples using a microfluidic-integrated cell phone imaging system. Forensic science international : genetics. 2020;48:102313-.

30. Boward ESBA, Wilson SLMS. A comparison of ABAcard® p30 and RSID™-Semen test kits for forensic semen identification. Journal of forensic and legal medicine. 2013;20(8):1126-30.

31. Wu H, de Gannes MK, Luchetti G, Pilsner JR. Rapid method for the isolation of mammalian sperm DNA. BioTechniques. 2015;58(6):293-300.

32. Biosystems A. AmpFlSTR™ NGM™ PCR Amplification Kit User Guide. Cheshire, nUnited Kingdom: ThermoFisher Scientific; 2018. p. 127.

33. Cho S, Shin KJ, Bae SJ, Kwon YL, Lee SD. Improved STR analysis of degraded DNA from human skeletal remains through in-house MPS-STR panel. Electrophoresis. 2020;41(18-19):1600-5.

34. Butler JM. Fundamentals of forensic DNA typing. Amsterdam;Boston;: Academic Press/Elsevier; 2010.

35. Promega. PowerPlex® Fusion 6C System: Promega Corporation; 2022 [Available from: https://www.promega.com.au/products/forensic-dna-analysis-ce/str-amplification/powerplex-fusion-6c-system/?catNum=DC2705.

36. Biosystems A. GlobalFiler™ PCR Amplification Kit: Thermo Fisher Scientific; 2022 [

37. Wilson-Wilde LK, Mojca; Walsh, Simon. DNA Profiling Success Rates In Volume Crime Cases To Determine The Optimal Number And Type Of Samples That Should Be Analysed Per Case. AustraliaAustralia New Zealand Policing Advisory Agency 2013.

38. Investigation FBo. Frequently Asked Questions on CODIS and NDIS [Available from: https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet.

39. Ossowski A, Diepenbroek M, Szargut M, Zielińska G, Jędrzejczyk M, Berent J, et al. Population analysis and forensic evaluation of 21 autosomal loci included in GlobalFiler™ PCR Kit in Poland. Forensic science international : genetics. 2017;29:e38-e9.

40. Vuichard S, Borer U, Bottinelli M, Cossu C, Malik N, Meier V, et al. Differential DNA extraction of challenging simulated sexual-assault samples: A Swiss collaborative study. Investigative genetics. 2011;2(1):11-.

41. Xu K, Clark CP, Poe BL, Lounsbury JA, Nilsson J, Laurell T, et al. Isolation of a Low Number of Sperm Cells from Female DNA in a Glass–PDMS–Glass Microchip via Bead-Assisted Acoustic Differential Extraction. Analytical chemistry (Washington). 2019;91(3):2186-91.

42. Clark C, Turiello R, Cotton R, Landers JP. Analytical approaches to differential extraction for sexual assault evidence. Analytica chimica acta. 2021;1141:230-45.

43. Yoshida K, Sekiguchi K, Mizuno N, Kasai K, Sakai I, Sato H, et al. The modified method of two-step differential extraction of sperm and vaginal epithelial cell DNA from vaginal fluid mixed with semen. Forensic science international. 1995;72(1):25-33.

44. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA 'fingerprints'. Nature (London). 1985;318(6046):577-9.

45. Inci F, Ozen MO, Saylan Y, Miansari M, Cimen D, Dhara R, et al. A Novel On-Chip Method for Differential Extraction of Sperm in Forensic Cases. Advanced science. 2018;5(9):1800121-n/a.

46. Gill P, Lygo JE, Fowler SJ, Werrett DJ. An evaluation of DNA fingerprinting for forensic purposes. Electrophoresis. 1987;8(1):38-44.

47. Norris JV, Manning K, Linke SJ, Ferrance JP, Landers JP. Expedited, Chemically Enhanced Sperm Cell Recovery from Cotton Swabs for Rape Kit Analysis. Journal of forensic sciences. 2007;52(4):800-5.

48. Lounsbury JA, Nambiar SM, Karlsson A, Cunniffe H, Norris JV, Ferrance JP, et al. Enhanced recovery of spermatozoa and comprehensive lysis of epithelial cells from sexual assault samples having a low cell counts or aged up to one year. Forensic science international : genetics. 2013;8(1):84-9.

49. Hall A, Ballantyne J. Novel Y-STR typing strategies reveal the genetic profile of the semen donor in extended interval post-coital cervicovaginal samples. Forensic Science International (Online). 2003;136(1):58-72.

50. Roewer L. Y chromosome STR typing in crime casework. Forensic science, medicine, and pathology. 2009;5(2):77-84.

51. Burg A, Kahn R, Welch K. DNA testing of sexual assault evidence: The laboratory perspective. Journal of forensic nursing. 2011;7(3):145-52.

52. Bredemeyer S, Roewer L, Willuweit S. Next generation sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis. Forensic sciences research. 2021;ahead-of-print(ahead-of-print):1-6.

53. Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A, et al. A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating

male relatives and paternal lineages. Forensic science international : genetics. 2011;6(2):208-18.

54. McDonald A, Jones E, Lewis J, O'Rourke P. Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa. Forensic science international : genetics. 2014;15:84-9.

55. Alberts BJ, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter. Molecular Biology of the Cell. 4th ed. New York: Garland Science; 2002.

56. Küçük N. Sperm DNA and detection of DNA fragmentations in sperm. Türk üroloji dergisi. 2018;44(1):1.

57. Alves MBR, Celeghini ECC, Belleannée C. From Sperm Motility to Sperm-Borne microRNA Signatures: New Approaches to Predict Male Fertility Potential. Frontiers in Cell and Developmental Biology. 2020;8.

58. Fitch JT, Stine C, Hager WD, Mann J, Adam MB, McIlhaney J. Condom Effectiveness: Factors That Influence Risk Reduction. Sexually transmitted diseases. 2002;29(12):811-7.

59. Amon P, Sanderson I. What is the microbiome? Archives of disease in childhood Education and practice edition. 2017;102(5):257-60.

60. Roux P-F, Oddos T, Stamatas G. Deciphering the Role of Skin Surface Microbiome in Skin Health: An Integrative Multiomics Approach Reveals Three Distinct Metabolite–Microbe Clusters. Journal of investigative dermatology. 2022;142(2):469-79.e5.

61. Aagaard KM, Segars JH. What is the microbiome and how do we study it? Seminars in reproductive medicine. 2014;32(1):3.

62. Kong HH. Skin microbiome: genomics-based insights into the diversity and role of skin microbes. Trends in molecular medicine. 2011;17(6):320-8.

63. Schloss JA, Gibbs RA, Makhijani VB, Marziali A. Cultivating DNA Sequencing Technology After the Human Genome Project. Annual Review of Genomics and Human Genetics. 2020;21(1):117-38.

64. Clarridge JE. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. Clinical Microbiology Reviews. 2004;17(4):840-62.

65. Afonina E, Chichkova N, Bogdanova S, Bogdanov A. 30S ribosomal subunits with fragmented 16S RNA: a new approach for structure and function study of ribosomes. Biochimie. 1991;73(6):777-87.

66. Rodnina MV. Translation in Prokaryotes. Cold Spring Harbor perspectives in biology. 2018;10(9):a032664.

67. Russell PJa, Wilbur Be. iGenetics: a molecular approach. Third, Pearson international ed. San Francisco: Benjamin Cummings is an imprint of Pearson; 2010.

68. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature reviews Microbiology. 2014;12(9):635-45.

69. Martijn J, Lind AE, Schön ME, Spiertz I, Juzokaite L, Bunikis I, et al. Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. Environmental microbiology. 2019;21(7):2485-98.

70. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. Journal of clinical microbiology. 2007;45(9):2761-4.

71. Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. mSphere. 2021;6(1).

72. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of microbiological methods. 2007;69(2):330-9.

73. Wang Y, Qian P-Y. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PloS one. 2009;4(10):e7401-e.

74. VandeWalle JL, Goetz GW, Huse SM, Morrison HG, Sogin ML, Hoffmann RG, et al. Acinetobacter, Aeromonas and Trichococcus populations dominate the microbial community within urban sewer infrastructure. Environmental microbiology. 2012;14(9):2538-52.

75. Lamoureux C, Surgers L, Fihman V, Gricourt G, Demontant V, Trawinski E, et al. Prospective Comparison Between Shotgun Metagenomics and Sanger Sequencing of the 16S rRNA Gene for the Etiological Diagnosis of Infections. Frontiers in microbiology. 2022;13:761873-.

76. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. PeerJ (San Francisco, CA). 2016;2016(3):e1869-e.

77. Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, et al. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. Journal of applied microbiology. 2013;115(5):1147-58.

78. Van Der Pol WJ, Kumar R, Morrow CD, Blanchard EE, Taylor CM, Martin DH, et al. In Silico and Experimental Evaluation of Primer Sets for Species-Level

Resolution of the Vaginal Microbiota Using 16S Ribosomal RNA Gene Sequencing. The Journal of infectious diseases. 2019;219(2):305-14.

79. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome research. 2011;21(3):494-504.

80. Illumina. 16S Metagenomic Sequencing Library Preparation. Illumina; 2013.

81. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Briefings in Bioinformatics. 2019;22(1):178-93.

82. Torrent I. Ion 16S™ Metagenomics Kit. ThermoFisher Scientific; 2015.

83. Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, et al. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. Journal of microbiological methods. 2012;91(1):80-8.

84. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature (London). 2011;475(7356):348-52.

85. Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, et al. Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Applied and Environmental Microbiology. 2014;80(24):7583-91.

86. Pacific Biosciences of California I. Procedure & Checklist - Preparing SMRTbell® Libraries using PacBio® Barcoded Universal Primers for Multiplexing Amplicons. 2 ed. California, US2020.

87. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nature biotechnology. 2019;37(10):1155-62.

88. Pacific Biosciences of California I. Sequencing 101: from DNA to discovery — the steps of SMRT sequencing 2020 [Available from: https://www.pacb.com/blog/steps-of-smrt-sequencing/.

89. Adewale BA. Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? African journal of laboratory medicine. 2020;9(1):e1-e5.

90. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. Frontiers in genetics. 2019;10:426-.

91. Illumina. Specifications for the MiSeq System 2022 [Available from: https://sapac.illumina.com/systems/sequencing-platforms/miseq/specifications.html.

92. Illumina. MiSeq Reagent Kit v3 2022 [Available from: https://sapac.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/miseq-reagent-kit-v3.html.

93. Illumina. Measuring sequencing accuracy: Illumina; 2022 [Available from: https://sapac.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html.

94. Pacific Biosciences of California I. Metagenomics 2022 [Available from: https://www.pacb.com/products-and-services/applications/complex-populations/microbial/.

95. Facility ANG. Quotation For Genomic Services. 2022.

96. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods. 2016;13(7):581-3.

97. Bolyen E, Rideout JR, Dillon MR, Abnet CC, Al-Ghalith GA, Alexander H, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology. 2019;37(8):852-7.

98. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.

99. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology. 2009;75(23):7537-41.

100. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic acids research. 2019;47(18):e103-e.

101. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nature communications. 2019;10(1):5029-11.

102. Mysara M, Leys N, Raes J, Monsieurs P. IPED: A highly efficient denoising tool for Illumina MiSeq Paired-end 16S rRNA gene amplicon sequencing data. BMC bioinformatics. 2016;17(1):192-.

103. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal. 2017;11(12):2639-43.

104. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research. 2013;41(Database issue):D590-D6.

105. McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME Journal. 2012;6(3):610-8.

106. Meola M, Rifa E, Shani N, Delbès C, Berthoud H, Chassard C. DAIRYdb: A manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products. BMC genomics. 2019;20(1):560-.

107. Liu Y-X, Qin Y, Chen T, Lu M, Qian X, Guo X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein & Cell. 2021;12(5):315-30.

108. Yaghjyan L, Mai V, Wang X, Ukhanova M, Tagliamonte M, Martinez YC, et al. Gut microbiome, body weight, and mammographic breast density in healthy postmenopausal women. Cancer causes & control. 2021;32(7):681-92.

109. Baselga A, Freckleton R. Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. Methods in ecology and evolution. 2013;4(6):552-7.

110. Mach N, Berri M, Estellé J, Levenez F, Lemonnier G, Denis C, et al. Early-life establishment of the swine gut microbiome and impact on host phenotypes. Environmental microbiology reports. 2015;7(3):554-69.

111. Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses [version 1; peer review: 3 approved]. F1000 research. 2016;5:1492-.

112. Torondel B, Ensink JHJ, Gundogdu O, Ijaz UZ, Parkhill J, Abdelahi F, et al. Assessment of the influence of intrinsic environmental and geographical factors on the bacterial ecology of pit latrines. Microbial biotechnology. 2016;9(2):209-23.

113. Sciences IfG. NIH Human Microbiome Project: University of Maryland School of Medicine; 2022 [Available from: https://hmpdacc.org/hmp/.

114. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature. 2007;449(7164):804-10.

115. Creasy HH, Fitzgerald MG, Fulton RS, Giglio MG, Magrini V, Martin JC, et al. Structure, function and diversity of the healthy human microbiome. Nature (London). 2012;486(7402):207-14.

116. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature (London). 2009;457(7228):480-4.

117. Crandall H, Kapusta A, Killpack J, Heyrend C, Nilsson K, Dickey M, et al. Clinical and molecular epidemiology of invasive Staphylococcus aureus infection in Utah children; continued dominance of MSSA over MRSA. PloS one. 2020;15(9):e0238991-e.

118. Grice EA, Kong HH, Turner ML, Segre JA, Conlan S, Deming CB, et al. Topographical and Temporal Diversity of the Human Skin Microbiome. Science (American Association for the Advancement of Science). 2009;324(5931):1190-2.

119. Schommer NN, Gallo RL. Structure and function of the human skin microbiome. Trends in microbiology (Regular ed). 2013;21(12):660-8.

120. Stout MJMDM, Zhou YMD, Wylie KMP, Tarr PIMD, Macones GAMDM, Tuuli MGMDMPH. Early pregnancy vaginal microbiome trends and preterm birth. American journal of obstetrics and gynecology. 2017;217(3):356.e1-.e18.

121. Benschop CCG, Quaak FCA, Boon ME, Sijen T, Kuiper I. Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? International journal of legal medicine. 2012;126(2):303-10.

122. Choi A, Shin K-J, Yang WI, Lee HY. Body fluid identification by integrated analysis of DNA methylation and body fluid-specific microbial DNA. International journal of legal medicine. 2013;128(1):33-41.

123. Komesu YM, Dinwiddie DL, Richter HE, Lukacz ES, Sung VW, Siddiqui NY, et al. Defining the relationship between vaginal and urinary microbiomes. American journal of obstetrics and gynecology. 2020;222(2):154.e1-.e10.

124. Zozaya M, Ferris MJ, Siren JD, Lillis R, Myers L, Nsuami MJ, et al. Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. Microbiome. 2016;4(1):16-.

125. Zhu A, Sunagawa S, Mende DR, Bork P. Inter-individual differences in the gene content of human gut bacterial species. Genome Biology. 2015;16(1):82.

126. Neckovic A, van Oorschot RAH, Szkuta B, Durdle A. Investigation into the presence and transfer of microbiomes within a forensic laboratory setting. Forensic science international : genetics. 2021;52:102492-.

127. Procopio N, Lovisolo F, Sguazzi G, Ghignone S, Voyron S, Migliario M, et al. “Touch microbiome” as a potential tool for forensic investigation: A pilot study. Journal of forensic and legal medicine. 2021;82:102223-.

128. Neckovic A, van Oorschot RAH, Szkuta B, Durdle A. Investigation of direct and indirect transfer of microbiomes between individuals. Forensic science international : genetics. 2020;45:102212-.

129. Ross AA, Doxey AC, Neufeld JD. The Skin Microbiome of Cohabiting Couples. mSystems. 2017;2(4).

130. Williams DW, Gibson G. Classification of individuals and the potential to detect sexual contact using the microbiome of the pubic region. Forensic science international : genetics. 2019;41:177-87.

131. Nye N. The sexome: identifying unique microbial signatures in male and female pairings: Murdoch University; 2020.

This page has been intentionally left blank

Chapter 2: Scientific Manuscript

Abstract

IDENTIFICATION OF UNIQUE MICROBIAL SIGNATURES PRE- AND POST-COITUS IN MALE-FEMALE PAIRINGS BY MASSIVELY PARALLEL SEQUENCING AND ITS POTENTIAL TO DETECT SEXUAL CONTACT.

Background: The capture of male DNA, post-assault, is important in sexual assault investigation**,** particularly where an offender is unknown to the victim. The recovery of DNA often occurs when the female victim undergoes a forensic medical assessment. Analysis regularly results in mixed autosomal DNA profiles. As these results contain both victim and perpetrator DNA, they are often difficult to interpret a searchable male profile. While STR profiling of the male Y-chromosome is often used to overcome this, the successful identification of an individual can be hindered by the paternal inheritance pattern of Y-STRs. An adjunct method of perpetrator identification lies with microbiome analysis using massively parallel sequencing.

Aims: This study aimed to identify ASVs that were unique to each participant and compare the bacterial communities found on the genitals pre- and post-coitus. From the sequence data derived, statistical analysis was performed to investigate if bacteria sequences could be used to infer contact between each male-female pairing.

Content: Samples were collected from 14 male-female pairings across two recruitment cohorts. Volunteers were asked to self-collect samples pre- and post-coitus. Samples were extracted using PureLink™ Microbiome DNA Purification Kit. Extracted DNA underwent library preparation using primers targeting the V1-V9 hypervariable regions of the bacterial 16S rRNA gene (~1,449 bp). Libraries were sequenced by PacBio® SMRT Sequel II sequencing platform. Unique bacterial signatures were detected in low frequencies (<1%) in male and female participants pre-coitus. The data indicates a disruption to microbial composition post-coitus. Further genomic analysis is needed to confirm species and subspecies classification of bacteria.

1. Introduction

The prevalence of sexual assault remains a significant public health and safety issue globally. Sexual assault is the broad term used to encompass any non-consensual sexual contact (1). This can range from assault through physical force and incapacitation to contact without ongoing consent. While sexual assault occurs irrespective of gender, women are most at risk. In Australia, 1 in 6 women over the age of 15 years have experienced sexual assault, whilst the figure for men remains lower at 1 in 25 (2). While victims are most often women, the majority of sexual offenders are recorded as male (2). Therefore, the investigation of sexual assault relies heavily on isolating perpetrator DNA from spermatozoa. This is typically obtained through a forensic medical examination (3). Biological specimens must be acquired as soon as appropriate as vaginal swabs taken after 24 hours are less likely to contain sperm that can be isolated and analysed (4). DNA evidence in sexual assault cases is essential when the perpetrator is not known to the victim or denies the contact occurred. Obtaining a DNA profile is important; however, it is not without limitations.

Due to the nature in which biological specimens are obtained, from the victim post-assault, mixed DNA is unavoidable. The presence of female DNA overwhelms the male DNA, which influences the ability to obtain a male autosomal short tandem repeat (STR) profile (5). To counter this issue, male DNA is isolated through a process called differential extraction. While this is the standard method for DNA isolation in sexual assault cases globally, it is laborious, time-consuming and can cause degradation of the male DNA in the sample (5-9). To avoid issues with mixed DNA samples, the STR profile of the Y-chromosome can instead be targeted. The Y-chromosome is male-specific, meaning female DNA will not generate a Y-STR profile (10). However, due to the paternal inheritance pattern of the Y-chromosome, the Y-STR profile of related males will be the same (11). Another issue is having no male DNA recovered from a biological specimen. Using a barrier contraceptive, a vasectomy, and medical conditions such as azoospermia and oligospermia will inhibit the presence of sperm.

This illustrates the need to develop an adjunct method that will aid in identifying a perpetrator in sexual assault investigations.

The application of the human microbiome in forensic investigation is an active area of research. Previous studies have reported that, while vastly interconnected, the bacterial composition of human body sites differs between individuals (12). This creates an opportunity to assess whether individuals carry a unique microbial signature that can be detected and used forensically. The bacterial 16S ribosomal RNA (16S rRNA) is a ~1,500 bp gene commonly targeted in microbiome studies. This is due to its regular occurrence across bacterial taxa and its highly conserved regions that allow for universal primer binding opportunities (13). The development of Massively Parallel Sequencing (MPS) platforms has allowed for high-throughput cataloguing of entire microbial communities. Illumina™ short read analysis is common for microbiome studies due to its high accuracy and affordability. However, identifying bacterial taxa below a genus level is challenging (14). For the microbiome to be used forensically to identify unique microbial signatures in individuals, species, and subspecies resolution is vital. Pacific Biosciences (PacBio®) Single-Molecule Real-Time (SMRT®) Sequel II system allows for full-length sequencing of the 16S rRNA. Long read sequencing aims to increase taxonomic ability by considering the entire 16S rRNA gene.

The analysis of vaginal and penile microbiomes has only briefly been examined for use in a forensic context. Few studies have investigated the shared microbiome in coinhabiting and non-coinhabiting couples. These studies investigated the transfer of the skin microbiome in differing locations on the body, including the pubic mound (15). None have aimed to identify the microbial transfer that occurs during penetrative intercourse, nor have adequate parameters been set to imitate a one-off sexual encounter. The pilot study by Nye (16) utilised short-read analysis of the 16S rRNA region V3-4 to investigate the microbial transfer during intercourse. However, it could not identify bacterial taxa below a genus level. For the microbiome to be used forensically, a method must be developed to correctly identify species, sub-species and strain levels of bacterial taxa. This study investigates the application of long-read 16S sequencing in identifying microbial signatures that transfer during intercourse.

2. Materials and Methods

2.1. Participant Recruitment and Selection

Human vaginal and penile skin samples were collected from consensual male/female couples who were self-reported as healthy, with no history of sexually transmitted infections or reproductive medical conditions. Each couple was required to be in a long term monogamous relationship (+12 months) with both participants above 19 years of age. Each male and female volunteer provided written consent to participate in the study, which was approved by the Human Research Ethics Committee at Murdoch University (Protocol 2020/059). Recruitment occurred over two cohorts, the first conducted in 2020 and the second in 2021. The first recruitment cohort received samples from 9 male-female pairings. The second cohort received samples from 5 male-female pairings.

2.2. Collection of Intimate Samples

Microflora from the genital region was collected by each volunteer participant in a male and female pairing. Female participants used sterile Copan rayon swabs to self-collect samples, following the technique for Low Vaginal Swab (LVS) self-sampling by PathWest Laboratory Medicine WA (17) [Supplementary Fig S1]. Male participants wetted sterile Copan rayon swabs with DNase/RNase free water immediately before self-collection. The method for male sampling followed a process of swabbing up and down the penis shaft and around the glans of the penis. Details of this procedure are described in the volunteer information pack [Supplementary Fig S2]. Baseline samples were collected after at least two to four days of abstinence from sexual intercourse. In addition, females were asked to complete baseline sample collection at least three to four days post menses to control against cyclic variation in the vaginal microenvironment associated with menses (18). Five replicate swabs were collected for each participant and timepoint, totaling 20 per couple (10 male, 10 female). Participants were asked to collect post-intercourse samples within 3 – 12 hrs post penetrative

intercourse. Volunteers returned their kits to a nominated third party, who stored samples in a locked -20° C freezer until processing.

2.3. Bacterial DNA Extraction for MPS

DNA from bacterial communities was extracted using the PureLink™ Microbiome DNA Purification Kit (Invitrogen™) as described in the user guide for buccal, vaginal and skin swab samples (19). During the lysis step, samples underwent a bead beat using horizontal agitation. This was performed on a TissueLyser II (Qiagen®) at 25 Hz for 2 x 3 mins. Penile skin (n = 87) and vaginal samples (n = 87) were processed in batches by couple to reduce the risk of cross-contamination between couples. During extraction, some discoloration/growth on swabs were observed. These samples were still processed, and a note was made so DNA could be reassessed for quality control post extraction. A positive control (n = 1) consisting of a known mock bacterial community was used for quality control. A 75ml aliquot of the ZymoBIOMICS™ Microbial Community Standard, which consists of ten inactivated microorganisms (eight bacteria and two fungi) was processed alongside samples as the positive control. Additionally, extraction control blanks (n = 14) were processed with each extraction batch to assess background microbe levels and for quality control. For continuity, the blank control used was a sterile Copan rayon swab identical to those used to collect the intimate samples. Purified DNA was stored at -20° C prior to further analysis.

2.4. Quality Control of DNA Extracts

Due to budget limitations, only one sample from each time point and individual could be sent for sequencing. DNA extracts were therefore assessed to identify the replicate with the most suitable quantity of bacterial DNA.

2.4.1. PCR of 16S rRNA Gene Targeting V4 Region

To evaluate successful isolation of bacterial DNA, a PCR was conducted amplifying the V4 region of the 16S rRNA gene. A total of 231 samples were processed, including controls and replicate samples. Reactions were carried out in 25 µL volumes consisting of 5 µL purified DNA sample, 12.5 µL of 2X GoTaq® Green Master Mix (Promega™), 0.4 µM each of forward (515F) and reverse (806R) primers (Sigma-Aldrich®) [Table 1].

Table 1: *Bacterial 16S rRNA primers used. The V4 primers used were supplied by Sigma-Aldrich® and target a 291bp fragment of the 16S gene. Primers are compatible with Illumina™ short read sequencing platforms. Forward primer catalogue name Illumina515F16sV4. Reverse primer catalogue name Illumina806R16sV4. Primers targeting bacterial V1-9 hypervariable regions used to prepare samples for full length 16S rRNA sequencing, Primers are compatible with PacBio® long read sequencing platform.*

Assay	Primer Location	Universal Sequence	16S Target Specific Sequence
Short Read 16S V4	515F	5`-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3`	5' – GTGCCAGCMGCCGCGGTAA – 3'
	806R	5`-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3`	5' – GGACTACHVGGGTWTCTAAT – 3'
Long Read 16S V1-9	27F	5'-GCAGTCGAACATGTAGCTGACTCAGGTCAC-3'	5'- AGRGTTYGATYMTGGCTCAG-3'
	1492R	5'-TGGATCACTTGTGCAAGCATCACATCGTAG-3'	5'- RGYTACCTTGTTACGACTT-3'

Thermal cycling on an Applied Biosystems™ ProFlex™ 96-well PCR system was performed with initial denaturation at 95° C for 3 mins, followed by 30 cycles of 95° C for 30 secs, 55° C for 30 secs, 72° C for 30 secs, with a final extension at 72° C for 5 mins. These

conditions mimic that of the Nye study (16) to ensure data was comparable. Pre-PCR and post-PCR procedures were conducted in separate laboratories to reduce the risk of amplicon contamination.

2.4.2. Gel Electrophoresis

PCR products amplified using the 515F/806R primers [Table 1] was expected to produce a ~291 bp product. To evaluate this, amplicons were visualized by gel electrophoresis on a 1% (*w/v)* agarose. A tris-acetate-EDTA (TAE) buffer was used for gel casting and as the running buffer. SYBR™ Safe DNA Gel Stain (Invitrogen™) was added during gel casting at a concentration of 0.1 µL per 1 mL of buffer. A 5 µL aliquot of PCR product was loaded on the gel and separated at 100 volts for 40 mins on a Bio-Rad Sub-Cell® GT alongside an Axygen® 100 bp – 3 kB DNA ladder (Fisher Biotech™). Gels were visualised on a Dark Reader® transilluminator (Clare Chemical Research™). After gel electrophoresis minimal amplification was observed in the negative controls, therefore, DNA from these extracts was pooled into a single sample prior to amplicon library preparation and sequencing. Samples were selected based on the presence and intensity of a band at ~300 bp with the more prominent band at each replicate chosen to progress to sequencing. Swabs that had previously been noted as having discoloration/growth showed a more prominent band. It was inferred that this may have been due to an increased microbial presence from the growth and these samples were excluded from further analysis. In some cases band intensity alone did not identify the desired amplicons in any replicate samples, in these circumstances samples were selected based on chronological order.

2.5. Library Preparation

Purified DNA samples selected from section 2.4 followed a two-step PCR library preparation method. Hypervariable regions V1-9 of the 16S rRNA locus was amplified using the primers 27F/1492R with a universal UNITAG sequence and amine block attached to the

5′ ends of each primer [Table 1] (20). Primary PCR was carried out in 30 µL reactions containing 0.3 µM each of the forward and reverse primers, 15 µL of 1X AccuStart™ II ToughMix® (Quantabio™), 0.75 µL each of ArcticZymes dsDNase and DTT, 6.6 µL nuclease-free water and 6 µL of purified DNA extraction or nuclease-free water (negative control). The dsDNase and DTT were included as part of the ArcticZymes PCR Decontamination Kit™, which was used to remove contaminating DNA in PCR master mixes without reduction of PCR sensitivity. The master mix was incubated at 37°C for 20 min (dsDNase activation), followed by incubation for 20 min at 60°C (dsDNase inactivation), ensuring that any template added after inactivation remains safe from digestion. One negative template control was included. The PCR cycling conditions consisted of an initial denaturation step at 94 °C for 3 mins, followed by 30 cycles at 94°C for 30 secs, 52 °C for 30 secs, and 72 °C for 2 mins and a final extension step of 72 °C for 5 mins. Primary PCR products were purified using NucleoMag NGS magnetic beads (Macherey-Nagel™) at a 0.6X ratio and eluted in 30 µL low Tris-EDTA (TE) buffer. Primary PCR products were visualized on a QIAxcel® capillary gel electrophoresis system (Qiagen™) to confirm the presence and size of amplicons. Amplicons were normalized to 1 ng/µL prior to barcoding PCR. Barcoding PCR reactions, allow indexing of individual samples, and were carried out in 25 µL reactions containing 0.3 µM each of the forward and reverse barcoded primers, 12.5 µL of 1X AccuStart™ II ToughMix® (Quantabio™), 3 µL nuclease-free water and 2 µL of template or nuclease-free water (negative template control). PCR cycling conditions were the same as described above, but with reduced 8 cycles. PCR products were quantified using QIAxcel® capillary gel electrophoresis system. Barcoded PCR amplicons were pooled in equimolar concentrations of ~1000 ng per pool based on QIAxcel quantification of the target ~1500 bp band. Amplicon pools were purified using NucleoMag® NGS magnetic beads (Macherey-Nagel™) at a 0.55X ratio and eluted into 50 µL TE buffer. Pools were normalised to ~500 ng of DNA in 37 µL and used for sequencing.

2.6. Massively Parallel Sequencing

Purified amplicon pools were sequenced at the Australian Genome Research Facility (AGRF) QLD, Australia. The amplicon library was sequenced using Pacific Biosciences single molecule real-time (SMRT) hi-fidelity (HiFi) sequencing on a single SMRT® cell using the PacBio® Sequel II System. Raw data files were provided in BAM format.

2.7. Processing of HiFi Read Files

Raw data files received from AGRF were first demultiplexed and converted to FASTQ format. Sequence data was analysed using the DADA2 pipeline (21). First, primers were removed and sequences were quality filtered. The DADA2 denoise algorithm was then used to produce amplicon sequence variants (ASVs) (22). Taxonomy was assigned using the Bayesian classifier with reference to curated SILVA v128 database (23). Taxonomic species-level assignments were then confirmed using BLAST analysis against the National Center for Biotechnology Information (NCBI) nucleotide (nt) database (24).

2.8. Statistical Analysis

RStudio (v1.4) and R version 4.1.1 (25) was used to analyse the data and focus on establishing similarities and divergences of profiles across samples. Alpha diversity measures were carried out to assess the richness and diversity of individual samples. Alpha diversity was measured using observed ASV's, Choa1, Shannon index and Inverse Simpson (26). The taxonomic distribution of samples was visualised using distribution plots for males and females at both time points (*before* and *after* intercourse) [Supplementary Fig S4]. Bar plots were used to visualise the difference in the microbial composition of male and female samples. A Wilcoxon test observed differential abundance for the top nine genera grouped by male and female [Supplementary Fig S5]. A heatmap was used to observe microbial composition grouping samples at an ASV level. Beta diversity measures were carried out to assess the richness and diversity of samples across different communities. Beta diversity was measured

using Principle Coordinates Analysis (PCoA) with Bray-Curtis (27) and UniFrac (28) distance measures. Permutational analyses of variance (PERMANOVA) were used to assess multivariate community level differences between males and females.

3. Results

3.1. Participant Summary

In the 2020 recruitment cohort, nine males and nine females within each couple provided five replicate penile skin and five replicate vaginal swabs both prior to and post penetrative intercourse (totalling 10 per person, five *before* and five *after*). During the pilot study by Nye (16) 1-2 swabs per participant and time point were extracted, and the V3-4 region was sequenced using the Illumina™ MiSeq™ system. The remaining swabs were stored at –20° C. After 2021 recruitment, all remaining swabs from the 2020 cohort were extracted along with all replicate swabs from the 2021 cohort. From the 2020 cohort, couples 14 and 19 had 4 replicate swabs per participant and timepoint (eight total per participant, four *before* and four *after*). All other couples from this cohort had three replicate swabs per participant and timepoint (six total per participant, three *before* and three *after*). In the 2021 recruitment cohort, five males and five females within each couple provided five replicate penile skin and five replicate vaginal swabs both prior to and post penetrative intercourse (totalling 10 per participant, five *before* and five *after*). A total of 216 swabs underwent extraction (2020 = 116, 2021 = 100). Couples in the 2020 cohort were aged between 22 and 30 years of age, and used a variety of different contraceptive methods such as the oral contraceptive, condom and withdrawal method [Table 2]. All couples from this cohort abstained from intercourse for between three and 14 days. Couples in the 2021 cohort were aged between 20 and 28 years of age, and all used some form of oral contraceptive pill [Table 2]. All couples from this cohort abstained from intercourse for between two and four days.

Table 2: *Summary of participant information (including samples from previous recruitment cohort). A more detailed table can be found in the supplementary information [Supplementary Table S1]. Information not available/unknown is marked by an asterisk (*).*

Recruitment Cohort	Couple	Participant Information	Contraception	Time Abstinent Prior to Intercourse
2020	1	Male = unknown yrs Female = 22 yrs	Contraceptive Pill	5 days
	2	Male = 25 yrs Female = 26 yrs	Withdrawal Method	14 days
	4	Male = 25 yrs Female = 25 yrs	Contraceptive Pill	10 days
	6	Male = 30 yrs Female = 28 yrs	Condom (started without)	10 days
	9	Male = 23 yrs Female = 26 yrs	Condom	5 days
	13	Male = 22 yrs Female = 22 yrs	Contraceptive Pill & Condom	6 days
	14	Male = 24 yrs Female = 22 yrs	Condom	3 days
	17	Male = unknown yrs Female = unknown yrs	Vasectomy	* *
	19	Male = 28 yrs Female = 25 yrs	Condom	7 days
2021	2	Male = unknown yrs Female = 20 yrs	Contraceptive Pill	4 days
	16	Male = 23 yrs Female = 22 yrs	Contraceptive Pill	2 days
	17	Male = 23 yrs Female = 22 yrs	Contraceptive Pill	2 days
	18	Male = 22 yrs Female = 22 yrs	Contraceptive Pill	3 days
	19	Male = 22 yrs Female = 28 yrs	Contraceptive Pill	3 days

A word to the examiners

Due to unforeseen circumstances, the data provided by AGRF was not received in time for submission of this thesis document. The pilot study by Nye which investigated the V3-4 region of the 16S rRNA gene was identical in the recruitment and extraction procedure. However, adequate statistical analysis was not completed at the time. To account for the V1-9 results that were not acquired in time for submission, statistical analysis of the pilot data is discussed from herein. Materials and methods for the pilot research are discussed by Nye (16), which is available from the Murdoch University Research Repository (http://researchrepository.murdoch.edu.au/id/eprint/59536).

3.2. 16S Microbial Diversity

3.2.1. Bioinformatics

Metabarcoding produced 1,817,828 raw sequences from 24 samples [Supplementary Table S2]. Sequences were quality filtered and the DADA2 denoise algorithm was used to produce ASVs. The production of ASVs provides a higher-resolution version of traditional operational taxonomic units (OTUs) which cluster sequences based on 97% similarity rather than unique sequence variants (ASV method). A total of 864 bacterial ASVs were identified from 24 samples. There were a total of 284,053 sequences identified *after* taxonomic assignment with a mean read count of 11,836 reads.

3.2.2. Rarefaction and Alpha Diversity

Rarefaction curves showed the number of ASVs plateaued at a depth of 2,200 reads [Supplementary Fig S3] indicating that adequate sequencing depth was reached.

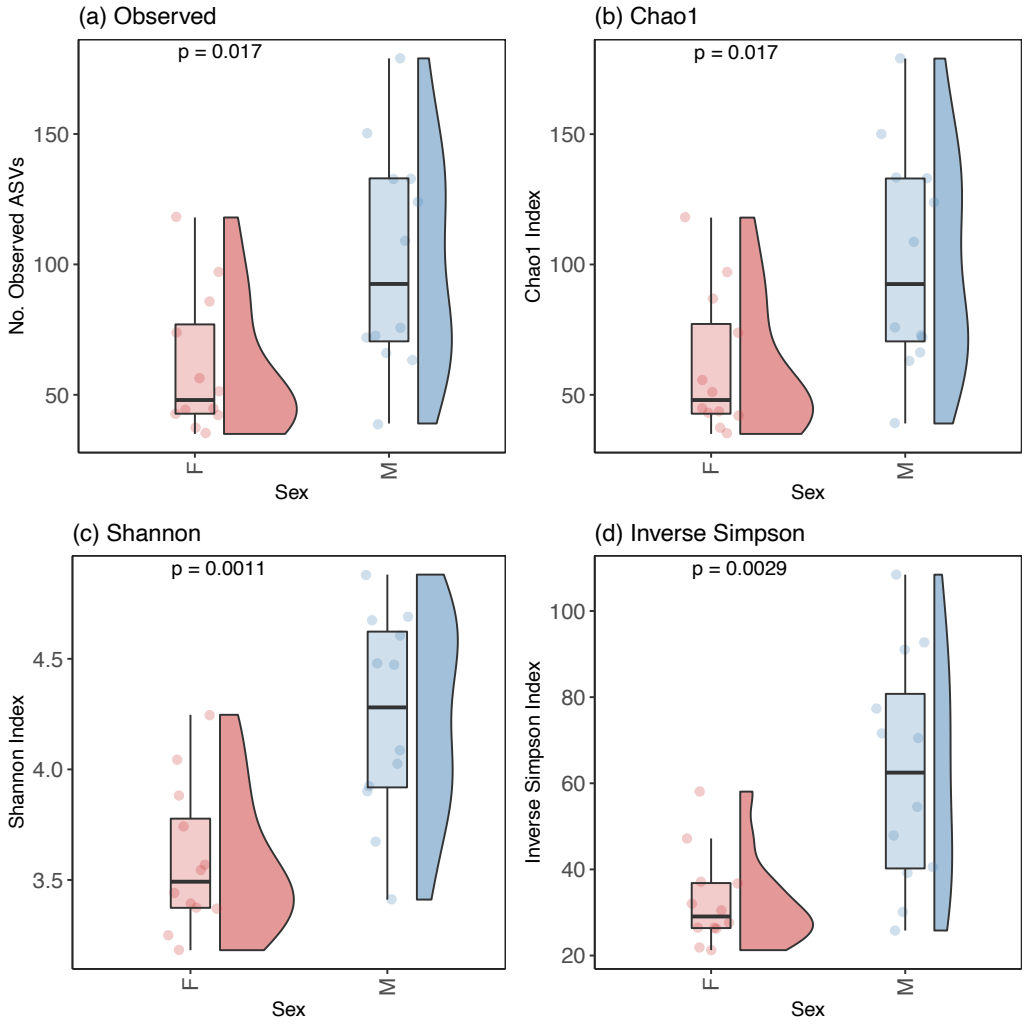


Figure 1: *Alpha diversity analysis in male vs female samples displayed through (a) observed ASVs, (b) Chao1, (c) Shannon and (d) Inverse-Simpson diversity measures.*

Alpha diversity analysis identified significant variation between male and female samples (Wilcoxon pair-wise test). The highest number of reads was observed in female vaginal samples. Penile skin communities were significantly more diverse, having a higher number of bacterial taxa (ASVs) in total (n = 726) than vaginal samples (n = 274). Male samples showed the highest alpha diversity across observed ASVs, chao1, Shannon and Inverse Simpson indices (P <0.05 in all measures) [Fig].

3.2.3. Taxonomy

The vaginal microbial communities of 12 female samples from six individuals was characterised. Overall, 15 genera were observed in baseline samples for females, dominated

by Lactobacillus (98.31%). The next most abundant genera were Finegoldia (0.70%), Peptoniphilus (0.34%), Corynebacterium (0.21%) and Gardnerella (0.11%). Post-coitus, a total of 32 genera were observed, dominated by Lactobacillus (98.15%). The next most abundant genera were Finegoldia (0.44%), Gardnerella (0.22%), Escherichia/Shigella (0.13%) and Peptoniphilus (0.11%).

The penile skin microbial communities of 6 males were observed. Overall, a total of 47 genera were observed in baseline samples, dominated by *Lactobacillus* (24.84%). The next most abundant genera were *Staphylococcus* (12.73%), *Finegoldia* (12.38%), *Peptoniphilus* (11.51%) and *Ezakiella* (4.67%). Post-coitus, a total of 57 genera were observed, dominated by *Lactobacillus* (34.93%). The next most abundant genera were *Peptoniphilus* (12.13%), *Finegoldia* (10.75%), *Ezakiella* (5.06%) and *Anaerococcus* (3.06%).

A heatmap displaying bacterial composition grouping at a genus level indicated changes in the abundance of taxa following intercourse [Fig 2a]. The most notable change was in the abundance of *Lactobacillus* in the male *before* and *after* samples of couple 17. Notable shifts *after* intercourse in vaginal bacterial communities appeared in females of couples 17 and 19. They were related to an increased relative abundance of *Finegoldia*. Notable shifts *after* intercourse in penile skin bacterial communities appeared in the male samples of couples 9, 17 and 19. They were related to an increased relative abundance of *Lactobacillus*. A heatmap displaying bacterial composition grouping at an ASV level indicated changes in the abundance of *Lactobacillus* ASVs following intercourse [Fig 2b]. The top 30 most abundant ASVs are displayed in this graph figure and indicate ASVs unique in the female participant of couple 17. ASV28, ASV29 and ASV30 are detected in the female of couple 17 only and are present in baseline and post-coitus samples. These three ASVs are not detected in the male baseline sample, but are detected in the male sample post-coitus. ASV22 was also detected in the female baseline and post-coitus sample and male post-coitus sample of couple 17 but was not unique across the cohort.

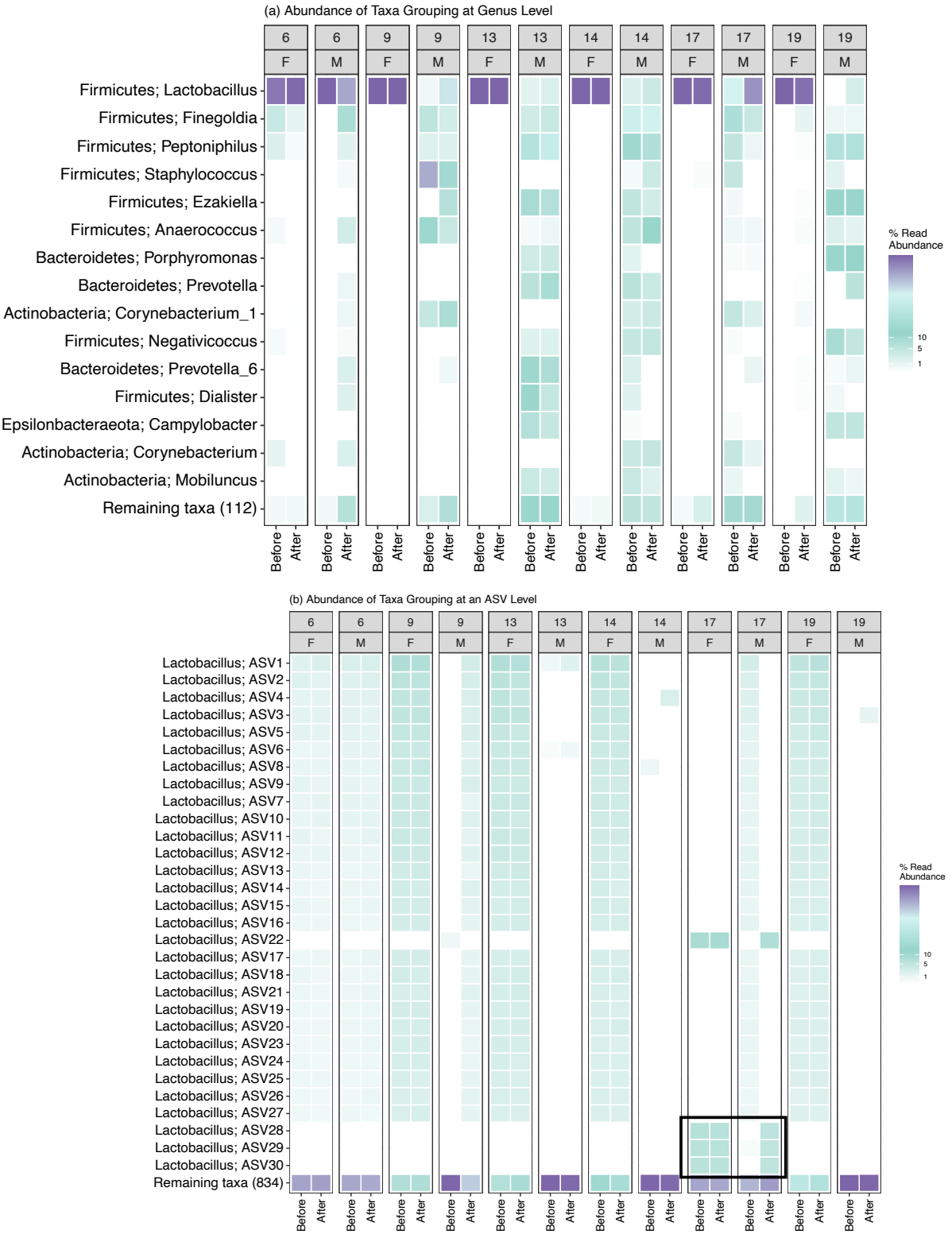
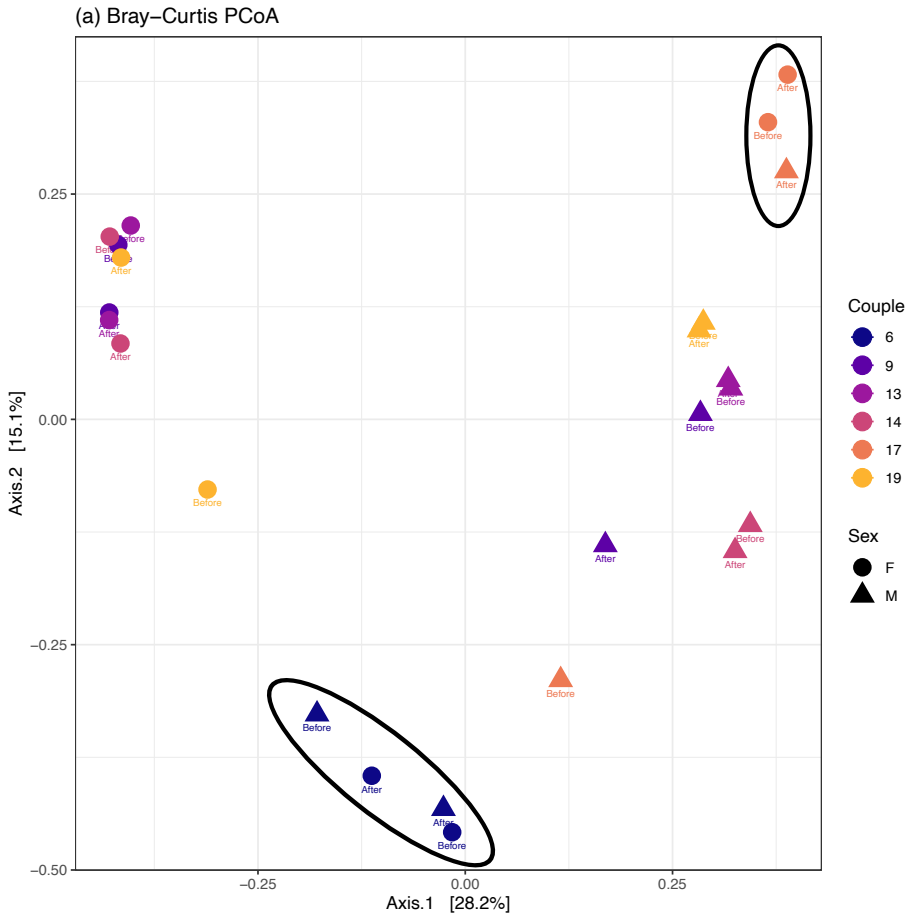


***Figure 2**: Taxonomic summary of bacteria in samples displayed with a heatmap of identified taxa. Couples are displayed side by side for each time point (before and after). (a) Relative abundance of top 15 taxa grouping at a genus level (including phylum annotation). (b) Relative abundance of top 30 taxa grouping at an ASV level (including genera annotation). Unique ASVs are indicated on the figure for the female of couple 17, along with their presence in the male sample post-coitus.*

3.2.4. Beta Diversity

Beta-diversity measures shown by ordination analysis indicated a difference in the composition of bacterial taxa in male and female samples. Principal coordinate analysis (PCoA) using the Bray-Curtis metric, which assesses dissimilarity based on overall ASV count data (i.e. includes abundance information), revealed sample clustering based on gender. Clustering based on couple was less apparent [Fig 3a]. PCoA using the weighted UniFrac metric, which incorporates both phylogenetic information of taxa and information based on abundance, was assessed [Fig 3b]. Using the Bray-Curtis metric, axis 1 (28.2%) drove the division of samples based on gender, and axis 2 (15.1%) drove the division based on baseline and post-coitus sampling. The weighted UniFrac metric noted the same sample division based on gender for axis 1 (47.2%) and sampling time for axis 2 (24.8%). These graphs indicated that sex had a significant effect on the overall microbial composition and was statistically confirmed using a PERMANOVA significance test (P=0.002).



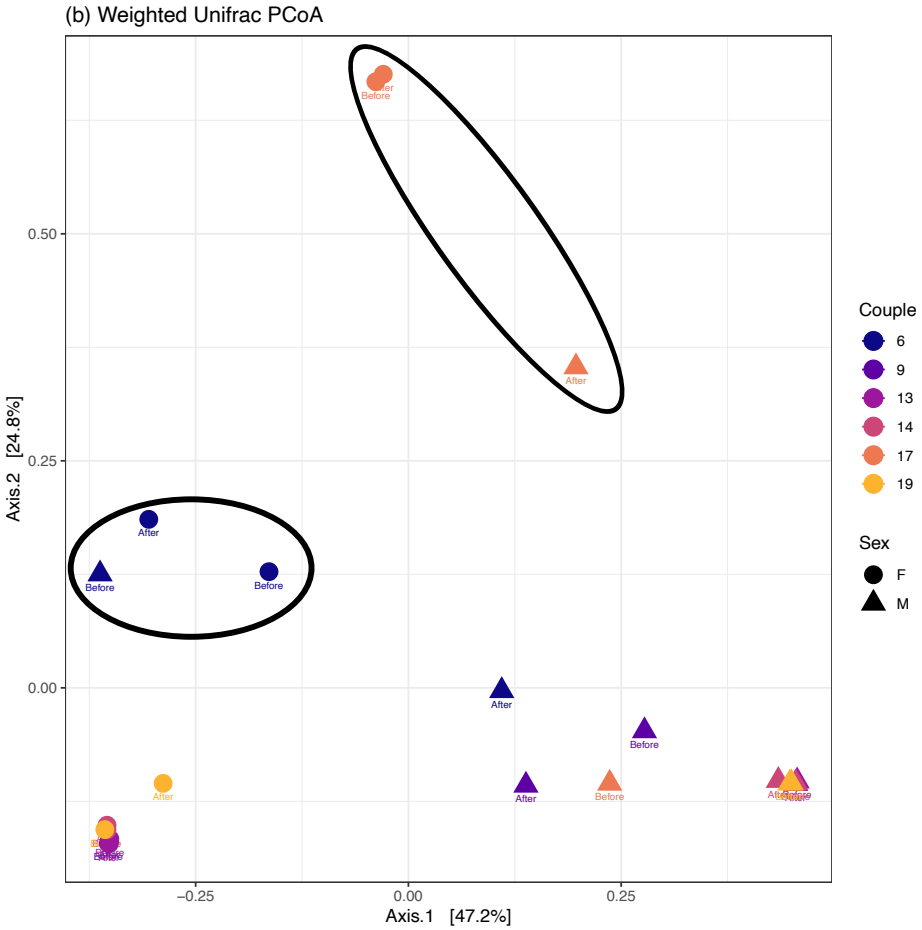


***Figure 3**: Beta diversity displayed through Principle Coordinate Analysis (PCoA) using Bray-Curtis (a) and UniFrac (b) distance measures. Clustering of couple 6 and 17 are circled in the figure.*

3.2.5. Shared Microbiota

The overlap of bacterial taxa (ASVs) *before* and *after* intercourse within male-female pairings was observed and displayed using Venn diagrams [Fig 4]. Couples 6, 9, 17 and 19 displayed the most notable change in bacterial taxa. Within couple 6 at baseline, there were 64 shared ASVs. Post-coitus, there were 68 shared ASVs indicating an increase of four when compared with baseline sampling. Within couple 9 at baseline, there were zero shared ASVs. Post-coitus, there were 31 shared ASVs indicating an increase of 31 when compared with baseline sampling. Within couple 13 at baseline, there were two shared ASVs. Post-coitus, there was no change to the total shared ASVs. Within couple 14 at baseline, there were three shared ASVs. Post-coitus, there was a decrease of one ASV compared with baseline

sampling. Within couple 17 at baseline, there were 19 shared ASVs. Post-coitus, there were 26 shared ASVs indicating an increase of seven when compared with baseline sampling. Within couple 19 at baseline, there were zero shared ASVs. Post-coitus, there were three shared ASVs.. Within couple 17 at baseline there were 19 shared ASVs. Post-coitus there were 26 shared ASVs indicating an increase of seven when compared with baseline sampling. Within couple 19 at baseline there were zero shared ASVs. Post-coitus there were three shared ASVs.

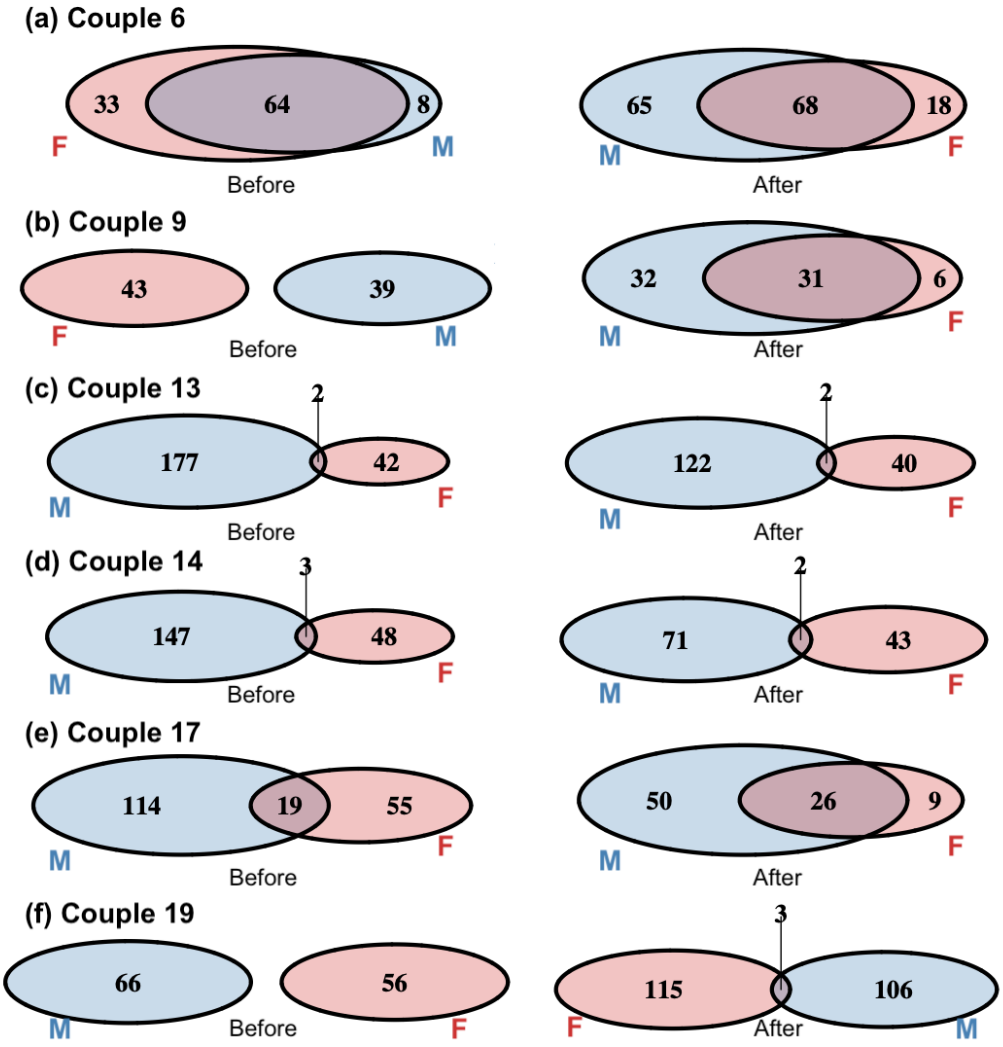


***Figure 4**: Venn diagrams indicating overlap of taxa (at ASV level) in male and female pairings both before and after coitus. All couples are displayed. Couples 6, 9, 17 and 19 indicate an increase in the amount of shared ASVs post-coitus.*

Changes in shared ASVs between couples 6, 9, 17 and 19 post-coitus were observed [Table 3]. ASVs identified as belonging to the *Lactobacillus* genera made up the majority of shared taxa, which is expected due to the high abundance of *Lactobacillus* across male and female samples. Only two shared ASVs post-coitus, ASV248 (*Porphyromonas*) and ASV104 (*Finegoldia*), identified across all four of these couples were not detected in either male or female sample at baseline. Only two shared ASVs, ASV214 (*Staphylococcus*) and ASV146 (*Finegoldia*), were identified in the male baseline sample, with the remainder detected in female baseline samples.

***Table 3**: ASVs that were not identified as shared at baseline for couples 6, 9, 17 and 19. This does not include taxa that were identified as shared at both time points. Text in bold indicates ASVs that were present in the female-only at baseline. Underlined text indicates ASVs that were present in the male-only at baseline. Regular text indicates ASVs that were not detected in either male or female samples at baseline. Couples 13 and 14 were excluded from this table as there was an insignificant change in shared ASVs. Genus level identification is listed in square brackets.*

Couple 6	Couple 9	Couple 17	Couple 19
ASV104 [*Finegoldia*]	**ASV1 [*Lactobacillus*]**	**ASV145 [*Gardnerella*]**	**ASV3 [*Lactobacillus*]**
ASV112 [*Finegoldia*]	**ASV10 [*Lactobacillus*]**	ASV146 [*Finegoldia*]	ASV248 [*Porphyromonas*]
ASV115 [*Peptoniphilus*]	**ASV101 [*Lactobacillus*]**	**ASV201 [*Gardnerella*]**	**ASV58 [*Lactobacillus*]**
ASV79 [*Finegoldia*]	**ASV11 [*Lactobacillus*]**	**ASV22 [*Lactobacillus*]**	
	ASV12 [*Lactobacillus*]	**ASV28 [*Lactobacillus*]**	
	ASV13 [*Lactobacillus*]	**ASV30 [*Lactobacillus*]**	
	ASV14 [*Lactobacillus*]	**ASV33 [*Lactobacillus*]**	
	ASV15 [*Lactobacillus*]	**ASV37 [*Lactobacillus*]**	
	ASV16 [*Lactobacillus*]	**ASV38 [*Lactobacillus*]**	
	ASV17 [*Lactobacillus*]	**ASV40 [*Lactobacillus*]**	
	ASV18 [*Lactobacillus*]	**ASV43 [*Lactobacillus*]**	
	ASV19 [*Lactobacillus*]	**ASV47 [*Lactobacillus*]**	
	ASV2 [*Lactobacillus*]	**ASV48 [*Lactobacillus*]**	
	ASV20 [*Lactobacillus*]	**ASV50 [*Lactobacillus*]**	
	ASV21 [*Lactobacillus*]	**ASV52 [*Lactobacillus*]**	
	ASV214 [*Staphylococcus*]	**ASV53 [*Lactobacillus*]**	
	ASV23 [*Lactobacillus*]	**ASV54 [*Lactobacillus*]**	
	ASV24 [*Lactobacillus*]	**ASV55 [*Lactobacillus*]**	
	ASV25 [*Lactobacillus*]	**ASV62 [*Lactobacillus*]**	
	ASV26 [*Lactobacillus*]	**ASV64 [*Lactobacillus*]**	
	ASV27 [*Lactobacillus*]		
	ASV3 [*Lactobacillus*]		
	ASV31 [*Lactobacillus*]		
	ASV32 [*Lactobacillus*]		
	ASV4 [*Lactobacillus*]		
	ASV5 [*Lactobacillus*]		

	ASV6 [*Lactobacillus*] **ASV7 [*Lactobacillus*]** **ASV8 [*Lactobacillus*]** **ASV89 [*Lactobacillus*]** **ASV9 [*Lactobacillus*]**		

The Venn diagrams [Fig 4] indicated an increase of seven shared taxa in couple 17 post-coitus. Further investigation indicated taxa identified as shared at baseline were not detected in post-coitus sampling. Only six of the shared ASVs at baseline were detected as shared post-coitus. With the exception of ASV146, the remainder of the shared ASVs in couple 17 identified post-coitus were new and previously detected in the baseline female sample. For all other couples, shared ASVs detected at baseline remained post-coitus.

4. Discussion

The female genital microbiome has been extensively studied due the relationship between the vaginal microbiome and fetal health during gestation (29). The microbiome of penile skin has been significantly less studied by comparison (30). This study is the first to examine pre and post-coitus penile skin and vaginal microbiomes using a short-read analysis of the 16S rRNA gene. Strict study parameters were set to simulate a once-off penetrative sexual encounter as best as possible within ethical and practicable bounds.

4.1. Sample Diversity in Male and Female Pairs

Bacterial amplicon sequencing showed that limited shared taxa were observed in male and female samples at baseline. The microbiome of the male penile skin samples revealed highly diverse environments with a low abundance of taxa. Female vaginal samples were significantly less diverse, and taxa appeared in a higher abundance. The diversity of male and female samples observed in this study is consistent with current literature (30). The diversity of female samples observed in this study is consistent with current literature which states that *lactobacillus* species are highly abundant in the vaginal microbiome (30). While the penile skin microbiome is not well documented, current literature describe the environment as being

highly diverse, with Corynebacterium, Peptoniphilus, Anaerococcus, Porphyromonas being the most abundant genera (31). The results observed in this study further complement current literature stating the highly diverse environment of the penile skin, however, the most abundant genera identified differs. In this study we identified Lactobacillus (24.84%) as the dominant genera in penile samples followed by Staphylococcus (12.73%) and Finegoldia (12.38%).

Taxonomic analysis found that in both male and female samples, there was an increase in the total genera observed post-coitus. A heatmap displaying mean relative abundance of taxa grouping at a genus level indicated the most notable change in abundance post-coitus was in male samples and related to an increase in the dominant female taxa, *Lactobacillus*. Few changes were observed in female samples; however, a disruption to the microbial composition was noted. For forensic applications, a microbial disruption could be useful for ascertaining if sexual contact has occurred.

Beta diversity analysis, which assesses the change in diversity across samples, was carried out using PCoA with Bray-Curtis and weighted UniFrac distance measures. The data indicated that samples generally clustered based on gender rather than by couple. Clustering of male and female samples was observed in couple 6, with *after* samples more closely clustered than *before*, indicating that microbial diversity in this couple became more homogenous (or “shared”) post-coitus. The use of beta diversity analysis for microbiome studies has evolved from microbial ecology research. When assessing the similarity of samples within different communities (e.g. male and female), PCoA plots display clustering for samples which are similar in their microbial diversity. A comparative study of the human gastrointestinal microbiome employed PCoA with unweighted UniFrac distance measures to show the clustering of sample diversity was driven by body mass index (32). Similarities are seen in this study regarding vaginal and penile microbiomes with the clustering of samples indicating a similarity in diversity.

The data from the Bray-Curtis PCoA indicated close clustering of male *before* and *after* samples for couples 13, 14 and 19. The corresponding female participants' samples indicate an increase in the distance between *before* and *after* samples, with the female *after* sample of couple 19 being the most prominent division. It may be inferred from this that there is a disruption to the microbial composition in these three female samples post-coitus and that the driving factor is their male partner. For couple 9, the inverse is occurring. The female *before* and *after* samples are closely clustered, while the male samples have a greater distance between them. This indicates a disruption to the microbial diversity in the male sample post-coitus. These couples, where disruption to the microbial composition has been observed, all used a barrier contraceptive. Without knowing what behaviours occur during their sexual encounter, it can be difficult to hypothesise a relationship between microbial diversity and the effect of intercourse. However, at a most basic level, this demonstrates that both male and female genital microbiomes are susceptible to disruption by the opposite sex and may allow for forensic exploitation for casework situations.

The most notable change in diversity was noted in couple 17. Clustering of the female *before* and *after* samples was observed on PCoA, along with the male *after* sample. The male *before* sample was clustered with other male samples from this cohort. This indicates a change in the microbial diversity of the male sample post-coitus, and the composition *after* is closely related to the female partner. Venn diagrams displaying the change in shared ASVs post-coitus show an increase in the shared ASVs for couples 6, 9, 17 and 19. The increase in shared ASVs confirms the relationship in couples 6 and 17 as clustering based on similarity post-coitus.

The clustering of couple 6 in beta diversity analysis indicates that baseline and post-coitus samples for both male and female participants are similar in their composition. During alpha diversity analysis, male and female samples were identified as being significantly different in their diversity. Based on this information, if microbial transfer were to occur during intercourse, samples would be expected to exhibit clustering of the post-coitus samples and

not baseline samples as observed in couple 6. Further investigation of relative abundance in couple 6 indicated that the male baseline sample exhibited characteristics typical of female vaginal diversity. The most abundant genera for the male participant was *Lactobacillus* (99.34%), *Peptostreptococcus* (0.16%), *Ureaplasma* (0.13%), *Finegoldia* (0.11%), and *Streptococcus* (0.11%). The percentage distribution is more similar to that of the female samples and these genera are more commonly associated with the vaginal microbiome (33).

Similarly, the female participant of this couple exhibited higher diversity during baseline sampling than the female participants of other couples. Initially, it was theorised that the unexpected diversity in these samples could be due to a sampling or human error. However, the post-coitus male sample in couple 6 detected the same genera present at baseline, and newly detected post-coitus genera were also detected in the female sample at baseline. If an error had caused the diversity observed in the male baseline sample, it would be expected that post-coitus diversity would be distinctively different. This data provides evidence that the microbial diversity in both male and female samples is most likely a true reflection of diversity and not due to a human or sampling error. This illustrates how the variation between individual microbiomes can be exploited for forensic use.

Venn diagrams displaying shared microbiome during baseline and post-coitus sampling, a decrease in the overall observed ASVs post-coitus was noted in couples 9, 13, 14 and 17. A well-documented theory regarding gut microbiome stability is that it undergoes a reduce then restore cycle. A study by Maifeld et al. (34) reported that the diversity of the gut microbiome followed a trend of reduced then restored diversity following fasting. If the genital microbiome follows a similar cycle, it would not be unusual to see a decrease in the overall ASVs count data between two different time points. Couples with a reduced ASV count post-coitus may have been in a restorative phase during baseline sampling and a reducing phase during post-coitus sampling. Inversely, couples 6 and 19 had an increase in the total ASVs detected during post-coitus sampling. As previously explained, it may be inferred that during baseline sample collection, the microbiomes of these couples could have been in a reducing

phase, with the restorative phase occurring during post-coitus sampling. Further studies of the human sexome should assess additional timepoints to glean insight into the restorative timeline for microbiota.

4.2. Microbial Transfer Post-Coitus

Couples 9, 13, 14 and 19 used a barrier contraceptive during intercourse. These couples exhibited the least amount of clustering during beta diversity analysis. Couples 9 and 19, however, did display a disruption to the diversity in male *after* and female *after* samples, respectively. Couples 13 and 14 did not display a significant disruption to microbial diversity, which was expected in couples that used a barrier contraceptive. While couple 6 did use a condom, the survey information they provided indicates that sexual intercourse began without one. The initial penile-vaginal contact in this couple explains why beta diversity analysis displayed clustering. Finally, as identified previously, couple 17 displayed clustering in PCoA. The male participant in this couple indicated that they had used no barrier contraceptive during intercourse. Further heatmap visualisation at an ASV level indicated that three ASVs of the top 30 were unique to the female sample of couple 17. Following intercourse, the male of this couple displayed an increase in the abundance of these three ASVs identified as unique to their female partner. Through in-depth analysis of this couple, we observed bacterial transfer of unique taxa from the female to the male during intercourse detected through post-coitus sampling. With more research this may prove useful and semi-diagnostic within a forensic context in detecting sexual intercourse.

During the investigation of microbial taxonomy, several genera were observed in post-coitus samples that were not present in either male or female baseline samples. These genera included *Prevotella, Escherichia/Shigella, Ezakiella, Lawsonella* and *Megasphaera.* However, they are still commonly observed in the male and female genital microbiome, so their presence was not unusual. Aside from *Prevotella* and *Ezakiella*, taxa appeared in low relative abundance (<2%), so it can be inferred that they were potentially missed or represented in

such low coverage that they were overlooked during baseline sampling. Interestingly, two other genera were identified post-coitus that are not commonly found in the genital microbiome. These were *Haemophilus* and *Streptococcus.*

Blast analysis against the NCBI database revealed ASVs assigned in the *Haemophilus* genera matched *Haemophilus parainfluenzae* (99.35% identity), *Haemophilus influenzae* (99.35% identity) and uncultured bacterium clones (99.57% identity). Most species within the *Haemophilus* genera are pathogenic in nature; however, species of this bacterium can be commonly found in the upper respiratory tract of healthy individuals (35, 36). Blast analysis of ASVs assigned in the *Streptococcus* genera matched *Streptococcus anginosus* (99.14% - 99.35% identity) and uncultured bacterium clones (99.78% identity). *Streptococcus anginosus* typically reside in the upper respiratory tract and have been reported in the oral microbiome (37). Based on the absence of these bacterial genera at baseline in female and male samples, along with their reported existence in the upper respiratory tract of healthy individuals, it may be inferred that these bacterium have transferred to the genital region during oral intercourse or through contamination at the time of collection by participants. Data on the health of participants at the time of sampling was not collected and may be a useful addition for further studies.

4.3. Forensic Microbiome Studies

Several studies have aimed to identify microbiome transfer in a forensic setting. A study by Neckovic et al. (38) investigated the transfer of the skin microbiome on the hands through direct and indirect contact. It was found that there was a disruption to the skin microbiome of individuals, and direct skin-to-skin contact was conducive to this contact. Another study by Williams et al. (15) investigated the microbiome of the pubic mound and assessed the ability to detect sexual contact. It was found that couples who indicated that they had been sexually active in the seven days prior to sample collection exhibited clustering on PCoA plots versus couples who had not been sexually active. The data from these studies

verify the data observed in this study, and similar relationships with microbiome transfer were identified.

4.4. Future Research Considerations

While ASVs were identified in this study as unique to the participant they were collected from, this is only valid given the small sample size. In order to identify if these taxa are, in fact, unique to an individual, an increased sample size must be observed and further genetic information on bacteria obtained. The survey information received for each participant did not clearly indicate how many days since menses baseline samples were collected. This could mean that cyclic variation influenced the results, and further research would need to gather more specific menses information. Interactions between hormone levels and microbial community have been demonstrated in the case of gut (39) and oral (40) bacteria microbiome. Future capture of participant metadata should aim to include questions relating to oral intercourse so that oral microbiome transfer can be investigated. Concerning taxa that still did not yield a species identification, further research investigating long-read sequencing of the 16S rRNA and shotgun sequencing should be considered. Both long-read and shotgun whole-genome sequencing will increase research costs moving forward. Nonetheless, as substantial bacterial transfer post-coitus has been identified, the cost would be justifiable. Emphasis should be put on identifying species and strain level classification of the *Lactobacillus* genera due to its high abundance and ubiquitous nature. It would also be advantageous to consider using traditional culture techniques to isolate and increase the presence of taxa that appear in low abundance. This will aid in investigating taxa that may be unique to an individual but lost during early bioinformatic processes and filtering.

5. Conclusion

The data generated in this project indicate that at a genus level, there is a disruption to the microbial composition of both penile and vaginal samples following intercourse. Upon

further investigation of microbial diversity at an ASV level, taxa were identified as unique to participants. It was observed that couples who used a barrier contraceptive had inhibition of bacterial transfer during intercourse. However, it was noted that sexual contact did disrupt microbial diversity. The data for a couple who began intercourse without a barrier contraceptive indicated some bacterial transfer which is theorised to have transferred during the initial penetrative contact. Minimal transfer was observed in this couple as a condom was used at some point during the sexual contact. The couple that did not use any form of barrier contraceptive displayed significant bacterial transfer. Investigation of shared bacterial taxa post-coitus indicated a transfer of the female vaginal microbiome to the male participant. Data in this study also suggests evidence of oral microbiome transfer theorised to have transferred during oral intercourse. This project acts as a proof of concept that bacterial transfer occurs between individuals during intercourse, and sexual contact can be detected through the analysis of unique bacterial signatures post-coitus, and shows promise as an alternative approach for the forensic investigation of sexual assaults.

6. Acknowledgements

I would like to acknowledge the volunteers who provided samples for this project; Brendan Chapman for the overall project support; Siobhon Egan for project, bioinformatic and statistical support; David Chandler and The Australian Genome Research Facility for their assistance with sequencing; Dr Matthew Payne and Dr Christopher Mullally for their assistance with library preparation; Dr Andrew Currie for providing assistance with funding for sequencing.

Word Count: 7,074

References

1. Graham LM, Treves-Kagan S, Magee EP, DeLong SM, Ashley OS, Macy RJ, et al. Sexual Assault Policies and Consent Definitions: A Nationally Representative Investigation of U.S. Colleges and Universities. Journal of school violence. 2017;16(3):243-58.

2. Welfare AIoHa. Sexual assault in Australia2020. Available from: https://www.aihw.gov.au/getmedia/0375553f-0395-46cc-9574-d54c74fa601a/aihw-fdv-5.pdf.aspx?inline=true.

3. Centre SAR. Procedure for Responding to a Recent Sexual Assault. Department of Health; 2017.

4. Casey DG, Domijan K, MacNeill S, Rizet D, O'Connell D, Ryan J. The Persistence of Sperm and the Development of Time Since Intercourse (TSI) Guidelines in Sexual Assault Cases at Forensic Science Ireland, Dublin, Ireland. Journal of forensic sciences. 2017;62(3):585-92.

5. Vuichard S, Borer U, Bottinelli M, Cossu C, Malik N, Meier V, et al. Differential DNA extraction of challenging simulated sexual-assault samples: A Swiss collaborative study. Investigative genetics. 2011;2(1):11-.

6. Inci F, Ozen MO, Saylan Y, Miansari M, Cimen D, Dhara R, et al. A Novel On-Chip Method for Differential Extraction of Sperm in Forensic Cases. Advanced science. 2018;5(9):1800121-n/a.

7. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA 'fingerprints'. Nature (London). 1985;318(6046):577-9.

8. Norris JV, Manning K, Linke SJ, Ferrance JP, Landers JP. Expedited, Chemically Enhanced Sperm Cell Recovery from Cotton Swabs for Rape Kit Analysis. Journal of forensic sciences. 2007;52(4):800-5.

9. Lounsbury JA, Nambiar SM, Karlsson A, Cunniffe H, Norris JV, Ferrance JP, et al. Enhanced recovery of spermatozoa and comprehensive lysis of epithelial cells from sexual assault samples having a low cell counts or aged up to one year. Forensic science international : genetics. 2013;8(1):84-9.

10. Burg A, Kahn R, Welch K. DNA testing of sexual assault evidence: The laboratory perspective. Journal of forensic nursing. 2011;7(3):145-52.

11. Bredemeyer S, Roewer L, Willuweit S. Next generation sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis. Forensic sciences research. 2021;ahead-of-print(ahead-of-print):1-6.

12. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial Community Variation in Human Body Habitats across Space and Time. Science (American Association for the Advancement of Science). 2009;326(5960):1694-7.

13. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. Journal of clinical microbiology. 2007;45(9):2761-4.

14. Van Der Pol WJ, Kumar R, Morrow CD, Blanchard EE, Taylor CM, Martin DH, et al. In Silico and Experimental Evaluation of Primer Sets for Species-Level Resolution of the Vaginal Microbiota Using 16S Ribosomal RNA Gene Sequencing. The Journal of infectious diseases. 2019;219(2):305-14.

15. Williams DW, Gibson G. Classification of individuals and the potential to detect sexual contact using the microbiome of the pubic region. Forensic science international : genetics. 2019;41:177-87.

16. Nye N. The sexome: identifying unique microbial signatures in male and female pairings: Murdoch University; 2020.

17. How to take your own Low Vaginal Swab [Internet]. Government of Western Australia. 2019. Available from: https://pathwest.health.wa.gov.au/~/media/PathWest/Documents/For-Patients/Vaginal-Swab-January-2019.pdf.

18. Hickey RJ, Zhou X, Pierson JD, Ravel J, Forney LJ. Understanding vaginal microbiome complexity from an ecological perspective. Translational research : the journal of laboratory and clinical medicine. 2012;160(4):267-82.

19. Invitrogen. PureLink™ Microbiome DNA Purification Kit User Guide (Buccal, Vaginal, or Skin Swab Samples). ThermoFisher Scientific; 2015. p. 3.

20. Pacific Biosciences of California I. Procedure & Checklist - Preparing SMRTbell® Libraries using PacBio® Barcoded Universal Primers for Multiplexing Amplicons. 2 ed. California, US2020.

21. DADA2: Fast and accurate sample inference from amplicon data with single-nucleotide resolution [Internet]. Available from: https://benjjneb.github.io/dada2/index.html.

22. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods. 2016;13(7):581-3.

23. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research. 2013;41(Database issue):D590-D6.

24. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50(D1):D20-d6.

25. Team R. RStudio: Integrated Development for R. Boston, MA RStudio, PBC; 2020.

26. Yaghjyan L, Mai V, Wang X, Ukhanova M, Tagliamonte M, Martinez YC, et al. Gut microbiome, body weight, and mammographic breast density in healthy postmenopausal women. Cancer causes & control. 2021;32(7):681-92.

27. Baselga A, Freckleton R. Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. Methods in ecology and evolution. 2013;4(6):552-7.

28. Lozupone CA, Knight R. The unifrac significance test is sensitive to tree topology. BMC Bioinformatics. 2015;16(1):211.

29. Jašarević E, Howard CD, Morrison K, Misic A, Weinkopff T, Scott P, et al. The maternal vaginal microbiome partially mediates the effects of prenatal stress on offspring gut and hypothalamus. Nature neuroscience. 2018;21(8):1061-71.

30. Ghemrawi M, Torres AR, Duncan G, Colwell R, Dadlani M, McCord B. The genital microbiome and its potential for detecting sexual assault. Forensic science international : genetics. 2021;51:102432-.

31. Zozaya M, Ferris MJ, Siren JD, Lillis R, Myers L, Nsuami MJ, et al. Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. Microbiome. 2016;4(1):16-.

32. Lin S-W, Freedman ND, Shi J, Gail MH, Vogtmann E, Yu G, et al. Beta-diversity metrics of the upper digestive tract microbiome are associated with body mass index: UGI Microbiome Beta-Diversity Associated with BMI. Obesity (Silver Spring, Md). 2015;23(4):862-9.

33. Fredricks DN, Marrazzo JM. Molecular methodology in determining vaginal flora in health and disease: its time has come. Current infectious disease reports. 2005;7(6):463-70.

34. Maifeld A, Bartolomaeus H, Löber U, Avery EG, Steckhan N, Markó L, et al. Fasting alters the gut microbiome reducing blood pressure and body weight in metabolic syndrome patients. Nature Communications. 2021;12(1):1970.

35. Edouard S, Million M, Bachar D, Dubourg G, Michelle C, Ninove L, et al. The nasopharyngeal microbiota in patients with viral respiratory tract infections is enriched in bacterial pathogens. European journal of clinical microbiology & infectious diseases. 2018;37(9):1725-33.

36. Yi H, Yong D, Lee K, Cho Y-J, Chun J. Profiling bacterial community in upper respiratory tracts. BMC Infectious Diseases. 2014;14(1):583.

37. Kwack KH, Lee J-H, Moon J-H. Whole genome and RNA sequencing of oral commensal bacterium Streptococcus anginosus subsp. anginosus with vancomycin tolerance. The journal of microbiology. 2022;60(2):167-76.

38. Neckovic A, van Oorschot RAH, Szkuta B, Durdle A. Investigation of direct and indirect transfer of microbiomes between individuals. Forensic science international : genetics. 2020;45:102212-.

39. Qi X, Yun C, Pang Y, Qiao J. The impact of the gut microbiota on the reproductive and metabolic endocrine system. Gut Microbes. 2021;13(1):1894070.

40. Bostanci N, Krog MC, Hugerth LW, Bashir Z, Fransson E, Boulund F, et al. Dysbiosis of the Human Oral Microbiome During the Menstrual Cycle and Vulnerability to the External Exposures of Smoking and Dietary Sugar. Frontiers in cellular and infection microbiology. 2021;11:625229.

Supplementary Information

PARTICIPANT INSTRUCTIONS (FEMALE)

PRIOR TO INTERCOURSE

1. **Wash your hands** thoroughly with soap and water and dry before taking a sample.
2. Find a comfortable position, either sitting or standing.
3. Remove the swab applicator labelled "BEFORE" from the packaging. **Avoid touching the cotton tip.**
    a. You are able to apply all five swabs at once; or complete one each time – five times
4. Part the labia and put the applicator end (cotton tip) about 2cm (length of one finger joint) inside your vagina.
5. Gently turn the swab around once, then leave for a count of 10 seconds.
6. Remove the swab, being careful not to touch any other skin.
7. Place the swab directly into its container.
8. Wash your hands.
9. Place into plastic bag provided.

AFTER INTERCOURSE

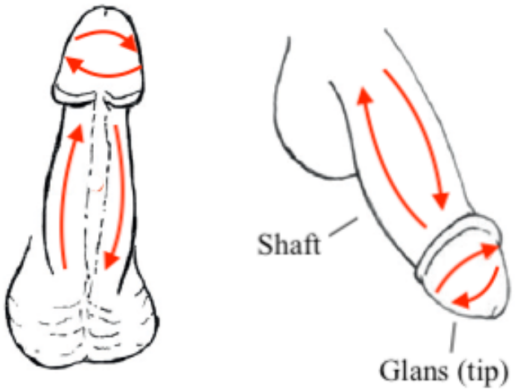
10. Repeat steps 1 – 9 for swabs labelled "AFTER".
11. Store in refrigerator until delivery to Murdoch University collection point B235 3.003 (Loneragan level 3). You may keep the sample at room temperature for up to 2hrs in order to transport it to campus.
12. If you at all feel uncomfortable with the storage conditions or collection point, you do not have to participate in the study and can dispose of the samples as you see fit.
13. For any withdrawals after your samples have been handed in, please contact Rhonda at R.Loxley@murdoch.edu.au or ethics@murdoch.edu.au
14. Once you have fully completed the sample collection process, you are welcomed to address any concerns or comments to the evaluation survey included in this instruction pack.

***Supplementary Figure S1**: Method for female self-sampling as per PathWest low vaginal swab technique.*

PARTICIPANT INSTRUCTIONS (MALE)

PRIOR TO INTERCOURSE

1. **Wash your hands** thoroughly with soap and water and dry before taking a sample.
2. Find a comfortable position, either sitting or standing.
3. Remove the swab applicator labelled "BEFORE" from the packaging**. Avoid touching the cotton tip**.
4. Moisten the swab with the distilled water provided.
    a. Please complete ONE swab at a time, moistening EACH swab with the distilled water provided.
    b. NOTE: Please use distilled water labelled "BEFORE". Discard "BEFORE" water after use.
5. As per the diagram below, swab up and down the penis shaft and around the glans (head) of the penis (if you are uncircumcised, you'll need to retract the foreskin prior to doing this). Repeat this action 10 times.
6. Place the swab directly into its container.
7. Repeat steps 1 – 6 for all five swabs.
8. Wash your hands.
9. Place in the plastic bag provided.



AFTER INTERCOURSE

10. Repeat steps 1 – 9 for swabs labelled "AFTER".
    a. NOTE: Please use distilled water labelled "AFTER'. Discard "AFTER" water after use.
11. Place into plastic bag provided.
12. Store in refrigerator until delivery to Murdoch University collection point at B235 3.003 (Loneragan level 3). You may keep the sample at room temperature for up to 2hrs in order to transport it to campus.
13. If you at all feel uncomfortable with the storage conditions or collection point, you do not have to participate in the study and can dispose of the samples as you see fit.
14. For any withdrawals after your sample has been handed in, please contact Rhonda at R.Loxley@murdoch.edu.au or ethics@murdoch.edu.au
15. Once you have fully completed the sample collection process, you are welcomed to address any concerns or comments to the evaluation survey included in this instruction pack.

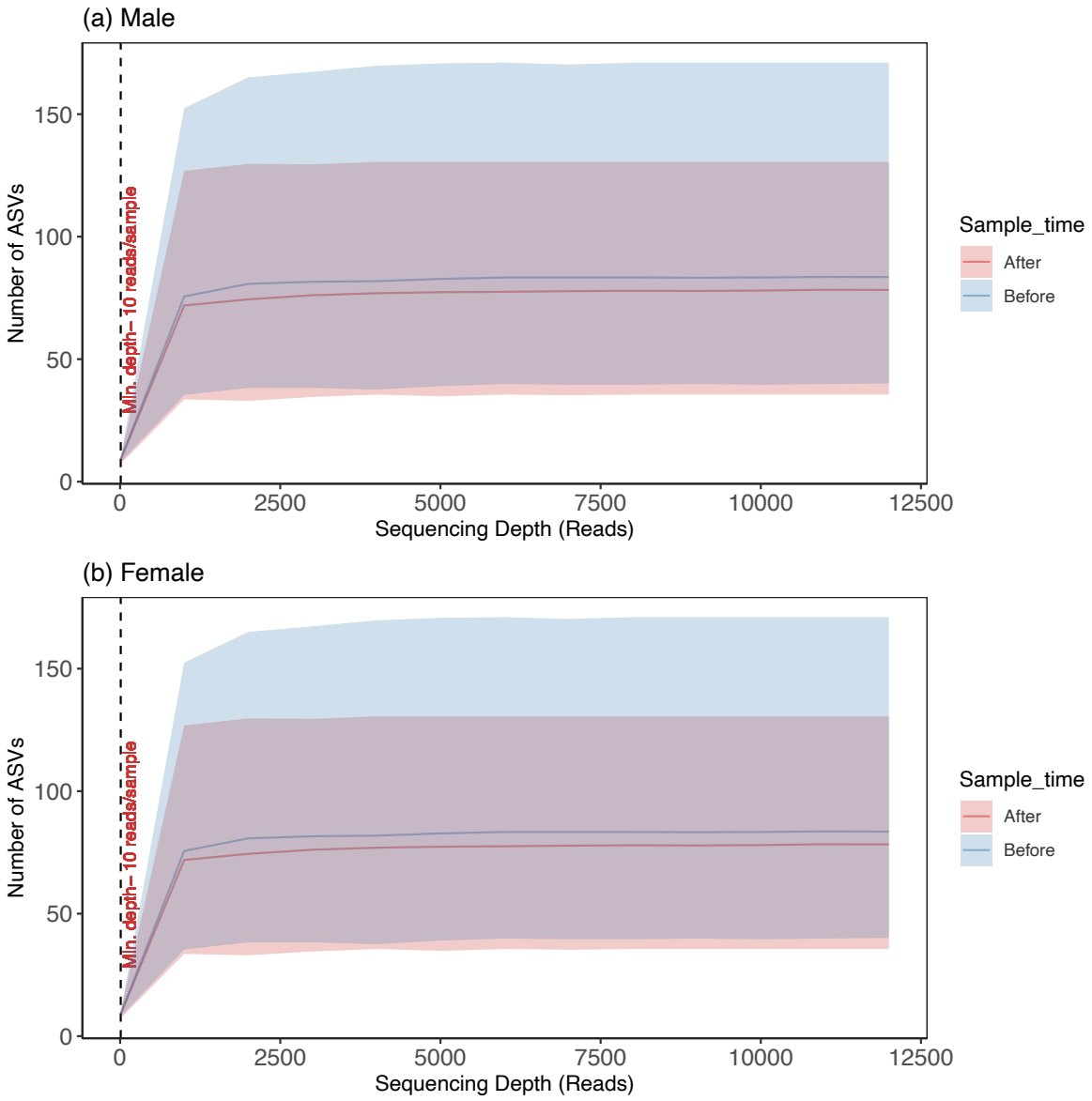
***Supplementary Figure S2**: Method for male self-sampling of penile skin*

Supplementary Table S1: *Detailed summary of participant survey results from 2021 cohort.*

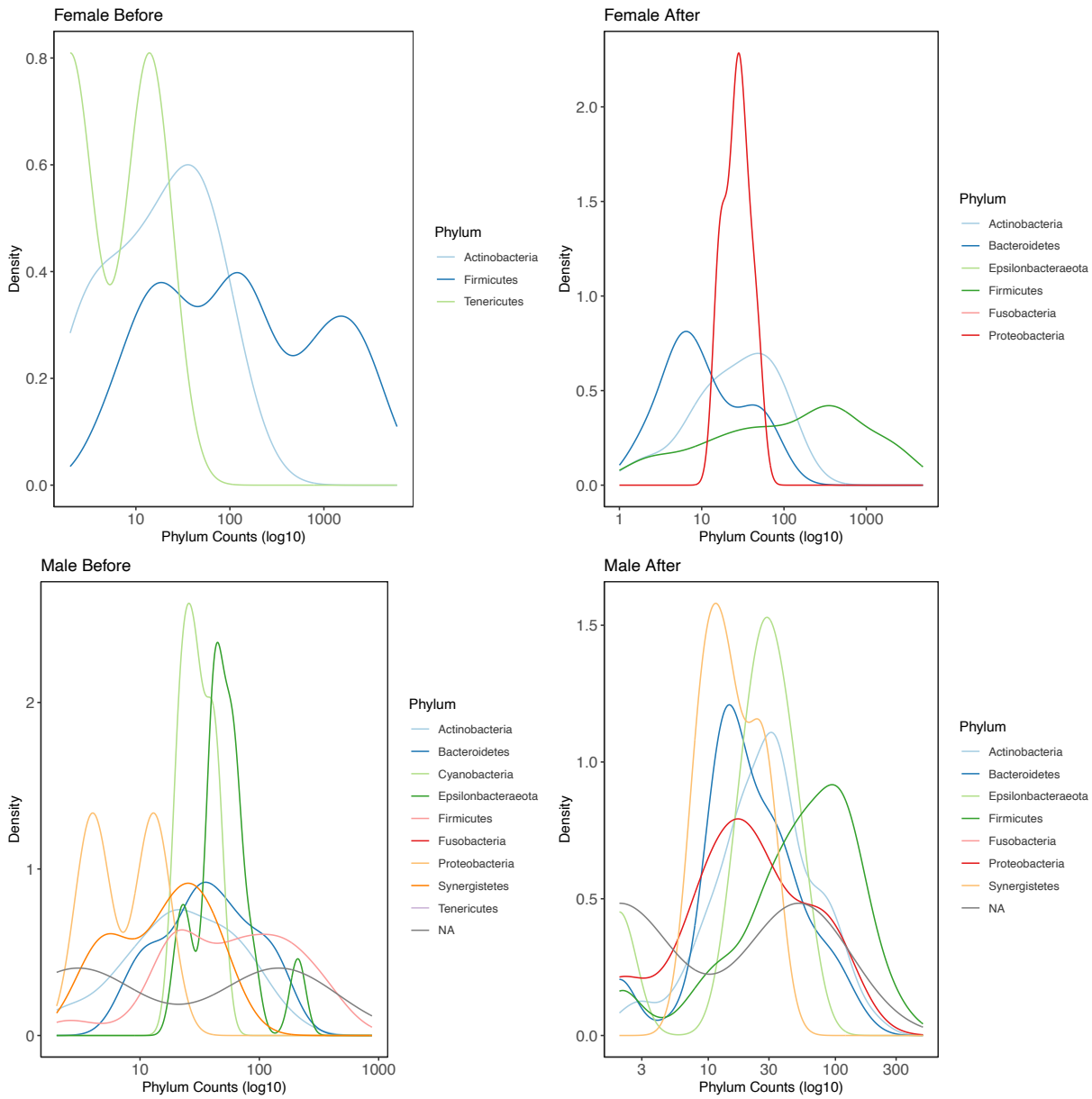
Couple	6		9		13		14		17		19	
Participant Information	Male = 30	Female = 28	Male = 23	Female = 26	Male = 22	Female = 22	Male = 24	Female = 22	Male = 29	Female = 27	Male = 28	Female = 25
Medical History	N/A	N/A	High Blood Pressure	N/A	Cancer	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Current Medications	N/A	N/A	Olmesartan (40mg) / Amlodipine (10mg)	Contraceptive Pill	N/A	Contraceptive Pill (Levlen)	Fluoxetine, Sodium Vlproatic, Dexamphetamine	N/A	N/A	N/A	N/A	N/A
Contraception	Condom (Started Without)		Condom		Contraceptive Pill/Condom		Condom		Vasectomy		Condom	
Time Abstinent Prior to Intercourse	10 days		5 days		6 days		3 days		*		7 days	
Lubrications Used	No		No		No		Yes		Yes		Yes	
Days Since Last Menstrual Period	N/A	6/3/20 - 10/3/20	N/A	26/5/20-30/5/20	N/A	2/2/20 – 6/2/20	N/A	7/8/20 – 12/8/20	*		N/A	28/8/20-3/9/20
Most recent underwear type	Cotton	Cotton	Cotton	Cotton	Neoprene	Nylon	Cotton	Polyester/Cotton	Cotton	Cotton	Cotton	Cotton
Circumsized (males only)	No	N/A	Yes	N/A	No	N/A	No	N/A	Yes	N/A	No	N/A
Public hair	Natural	Mostly Trimmed/No Hair	Minimally Trimmed	Minimally Trimmed	Natural	No Hair	Mostly Trimmed	Mostly Trimmed	Trimmed	Brazilian	Minimally Trimmed	No Hair
Hair Removal Method	N/A	N/A	N/A	N/A	N/A	Shaving	Razor	Razor/Shave	Trimmer	Wax	N/A	Shave

Supplementary Table S2: *Summary of sequences obtained from MPS of V3-4 regions of bacterial 16S rRNA gene using Illumina™ MiSeq™ platform.*

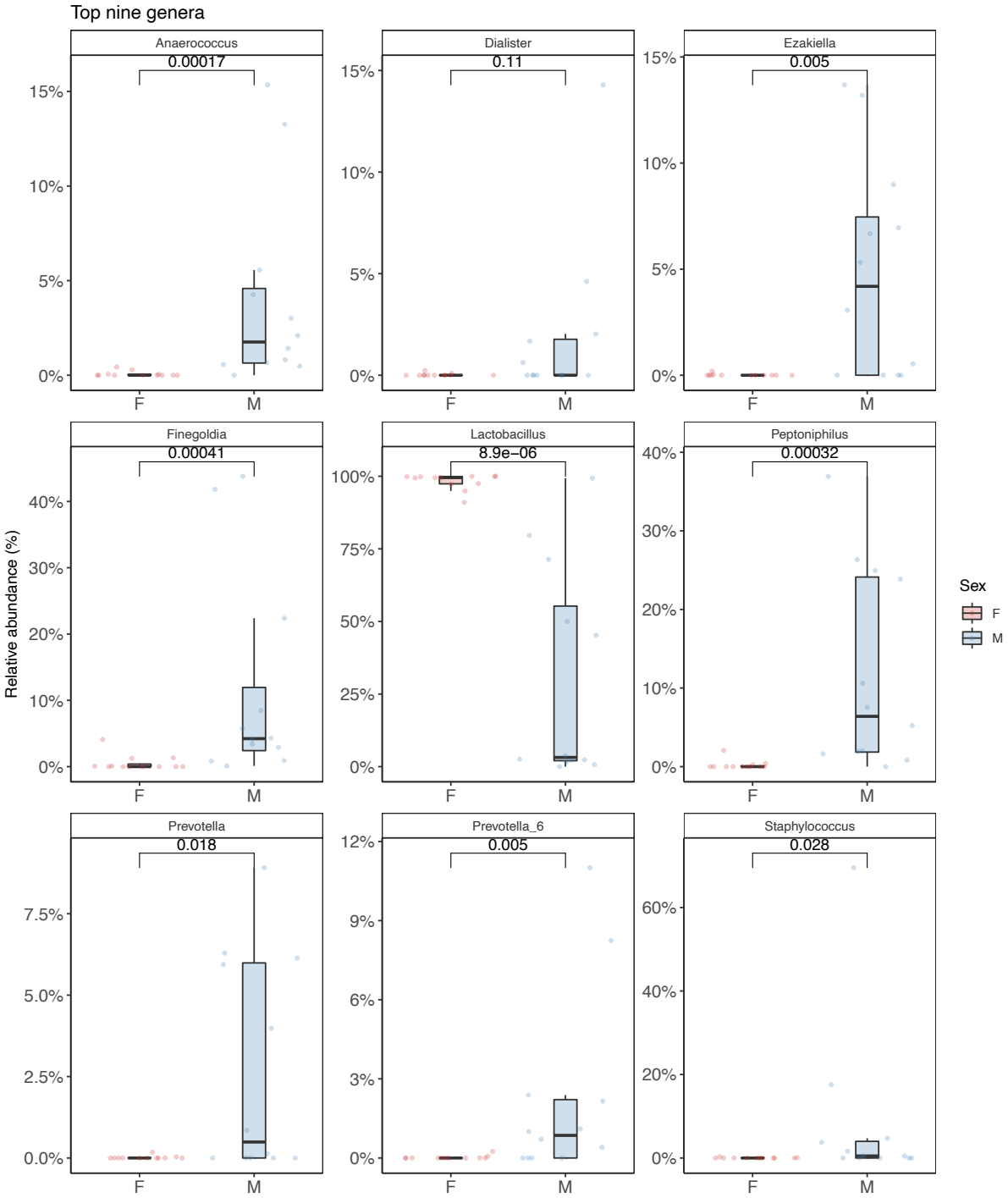
Bioinformatic Step	Sample (Totals)	Female		Male	
		Before	After	Before	After
Sum input	908914	313968	269464	197858	127624
Mean input	37871	52328	44911	32976	21271
Sum filtered	626643	249881	213401	103055	60306
Mean filtered	26110	41647	35567	17176	10051
Filtered %	69	80	79	52	47
Sum denoised	620581	249049	212046	101214	58272
Mean denoised	25858	41508	35341	16869	9712
Sum merged	609123	246339	208853	98452	55479
Mean merged	25380	41057	34809	16409	9247
Merged %	67	78	78	50	43
Sum non-chimeric	284053	116808	81276	44002	28486
Mean non-chimeric	11836	19468	16255	7334	4748
Input non-chimeric %	31	37	30	22	22



***Supplementary Figure S3**: Rarefaction curve of observed ASV reads in male and female samples. This is used to assess richness and sequencing depth. Curve plateaus at a depth of 2,200 reads.*



Supplementary Figure S4: *The taxonomic distribution of samples visualised using distribution plots for males and females at both time points (before and after intercourse). Grouped by phylum level identity.*



Supplementary Figure S5: Bar plots displaying differential abundance of top nine genera grouped by male and female samples. Statistical analysis was performed using Wilcoxon pair-wise significance test.