2-6-2018 10:00 AM

Characterization of Urinary Microbiome and Their Association with Health and Disease

Yige Bao, *The University of Western Ontario*

Supervisor: Burton, Jeremy P., *The University of Western Ontario*
Co-Supervisor: Reid, Gregor, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Microbiology and Immunology

Abstract

There has been a growing interest in human microbiome studies in the past decade, with the development of high-throughput sequencing techniques. These microorganisms interact and respond to the host as an entity, and are involved in various homeostatic functions including nutrition digestion, immune response, metabolism and endocrine regulation. The urinary microbiome, however, remains relatively under-investigated.

One of the technical challenges of urinary microbiome studies is the samples usually contain a large number of host cells and low microbial biomass. These samples with the high host, low microbial abundance ("high-low" samples) are associated with increased risk of compromised quality of 16s rRNA gene sequencing results. An analysis with mock samples showed that mechanisms of host materials interfering with microbiome analysis includes reducing microbial DNA extract yield by competitively binding to the filter of DNA extraction column, inhibiting PCR amplification of 16S rRNA gene regions as non-target DNA, and consuming sequencing depth by unspecific amplification from PCR. To counter these issues, a refined processing protocol and a quality checking tool were developed for handling "high-low" samples. With these methods, a combination of sequencing-based methods and enhanced culture-based methods showed evidence of bacteria in renal tissue samples.

On the other hand, the optimal urine sample collection and storage methods for microbiome study have not been reported. An optimisation experiment showed that urine samples with a volume higher than 20 mL and stored in centrifuged pellets generated the best sequencing results.

The urinary microbiome of healthy subjects and urinary stone patients were characterised using 16s rRNA gene sequencing and enhanced quantitative urine culture (EQUC) techniques. Although no clear distinction was observed of urinary microbiome profiles between healthy subjects and urinary stone patients, male and female individuals do have their unique urinary microbiome profiles. The urinary microbiome profile of an individual remained stable throughout three months.

Investigation of urine samples of metabolic stone patients before and after lithotripsy showed fluctuations in their urinary microbiome profiles, with newly-emerged microbes in sequencing results correlated with microbes cultured from stone samples. These results suggested bacteria liberated from metabolic stones during lithotripsy.

Keywords

Urinary microbiome, Urinary microbiota, 16s rRNA gene sequencing, enhanced quantitative urine culture, contamination, metagenomics, renal cell carcinoma, urinary stones, metabolic stones

Co-Authorship Statement

The experiment and analysis presented in this thesis were primarily conducted by Yige Bao, Jeremy Burton and Gregor Reid. Exceptions are noted below.

In chapter 2, DNA quantification with fluorometer and spectrophotometer were performed partly by a student volunteer, Jasper Wu, under supervision of Yige Bao. Microbiological culture of the tissue samples and Sanger sequencing of culture-derived colonies were partly performed by Jasper Wu and Seema Nair, a visiting scholar in the Burton lab. In chapter 3, processing and culture of collected urine samples were partly performed by Jasper Wu. In chapter 4, microbiological cultures of urine and stone samples collected in West China Hospital were partly performed by the microbiology clinicians from the clinical lab of the hospital.

Acknowledgments

I was with great fortune to have had my two supervisors, Drs. Gregor Reid and Jeremy Burton. It is never possible for me with my limited English vocabulary to express my full acknowledgement for their support, help and guidance on this journey. I had very poor background knowledge of microbiology and bioinformatics when entering the program but your patient supervision, thoughtful consideration and kind support make this thesis possible.

I would like to thank Dr. Jeremy Burton for his great mentorship and kind help. His unlimited patience is the greatest encouragement and comfort I could ever have. He is a great supervisor as he has great balance of support and pressure, a perfect dose for most motivation. Your care and guidance have gone way beyond as a supervisor and I felt very lucky to meet you in my life.

I also give my most sincere acknowledgement to Dr. Gregor Reid. His wisdom and foresight brought the inspiration of this project. From him I not only learned scientific knowledge, but also the dedication, enthusiasm and standards to be a great scientist.

I would also like to thank my advisory committee member, Drs. Greg Gloor and John Denstedt. Dr. Gloor provided solid bioinformatic enlightenment to me and Dr. Denstedt, without his support I would never had a chance to join this great program.

A big thank you to Dr. Jean Macklaim, Kyle McDonald, Jordan Bisanz, Camilla Urbaniak, Shannon Seney, Lesley Carmichael and all members from Burton, Reid and Gloor lab. Your kind help, support and encouragement are the strongest motivation for me.

I also thank Drs. Qiang Wei and Kunjie Wang from Department of Urology, West China Hospital who provided great understanding, trust and support for this program.

I would also personally thank my Mom and Dad, and my girlfriend Jie. Your love and support are strong, sweet and sincere, minding me of home far, far away.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALDEx	ANOVA- like differential expression
ANOVA	Analysis of variance
BLAST	Basic local alignment tool
bp	Base pair
°C	Degrees Celsius
CBA	Columbia blood agar
CFU	Colony forming unit
clr	Centered log-ratio
ddH2O	Double distilled water
DNA	Deoxyribonucleic acid
EQUC	Enhanced quantitative urine culture
g	Gram
h	Hour(s)
ISU	Individual sequence units
kb	Kilobase pair
Mb	megabases
min	Minute(s)
OTU	Operational taxonomy unit
PBS	Phosphate buffered saline
PCA	Principal components analysis
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RCT	Randomized control trials
RDP	Ribosomal database project
RNA	Ribonucleic acid
rpm	Rotations per minute
rRNA	Ribosomal ribonucleic acid
SVD	singular value decomposition

Chapter 1

1 General Introduction

This chapter has been adapted with permission from:

- Di Bella, J. M., Bao, Y., Gloor, G. B., Burton, J. P., & Reid, G. (2013). High throughput sequencing methods and analysis for microbiome research. Journal of microbiological methods, 95(3), 401-414.
- Bao, Y., Welk, B., Reid, G., & Burton, J. P. (2014). Role of the microbiome in recurrent urinary tract infection. Novel Insights into Urinary Tract Infections and their Management, 49.
- Bao, Y., Al, K. F., Chanyi, R. M., Whiteside, S., Dewar, M., Razvi, H., ... & Burton, J. P. (2017). Questions and challenges associated with studying the microbiome of the urinary tract. Annals of translational medicine, 5(2).

1.1 The human microbiome

The human microbiome is usually defined as all microbial populations (bacteria, fungi, archaea, and viruses) together with their genes and genomes that inhabit the human body (Turnbaugh *et al.*, 2007). Another related term is "microbiota", which refers to only the bacteria present at a particular site. Lots of these microbes are commensal, and their mutualistic relationship with the host has been reported. This relationship is more apparent upon study of the gut microbiome, which contains the highest abundance of microbes associated with any site in/on the human body. Within the gut, microorganisms uptake nutrients primarily from ingested food, and fulfill a number of homeostatic functions including the maintenance of barrier function, immune regulation, defence against pathogens (Round & Mazmanian, 2009) and digestion of complex substances from the human diet that are otherwise difficult to break down (Gill et al., 2006). Many of these microorganisms may be potential pathogens, but it is believed that these microbes are controlled and limited by the whole microbial ecosystem in the niche. The components and structure of such microbial ecosystems may change over time, but in most occasions, for a healthy subject, such a community maintains its relative stability (Ley *et al.*, 2006). Disruption of this stability could cause, or be the result of, a "dysbiotic" state which usually links to pathological conditions including inflammatory bowel disease (Garrett *et al.*, 2010), dental caries (Ling *et al.*, 2010), and obesity (Ley *et al.*, 2006).

1.2 High throughput, culture-independent sequencing methods for microbiome study

Our current understanding of the human microbiome has largely been due to the recent advancements in sequencing technologies allowing for high-throughput assessment of microbiome populations at relatively low cost. These studies have shown different body sites have distinct microbiological communities, and these communities differ between individuals (Turnbaugh *et al.*, 2007).

Despite the fact that microbial cells vastly outnumber eukaryotic cells in the human body, most of the microbes have not yet been cultivated (Amann *et al.*, 1995). Therefore, most uncultivated microorganisms remain to be fully characterized. With the recognition that DNA sequences, particularly the 16S ribosomal gene sequence, are useful phylogenetic markers and microbial identifiers, scientists are starting to gain an appreciation of the complexity of microbial ecosystems (Pace *et al.*, 1997 & 2009; Woese *et al.*, 1987). These developments, along with those of PCR and in situ hybridization, have led to a shift toward culture-independent studies of microbial communities.

1.2.1 Common strategies of sequencing-based methods for microbiome study

Three commonly used sequencing-based methods for investigating the microbiome are amplicon, metagenome, and meta-transcriptome. In amplicon sequencing, a certain gene or gene fragment is amplified, and the sequence is determined. This sequence is then used as a comparator to other similar sequences within other microorganisms. If a highly conserved genetic locus is chosen, such as segments of the 16S rRNA gene, the organisms are able to be taxonomically identified. By comparing the organisms present within multiple samples, researchers can identify how each differs within the environment.

While amplicon-style analysis only depends on a single gene, to obtain more detailed information regarding the particular sample, a metagenomic approach is used. Here, the aim is to determine the total genetic composition of all microorganisms present in order to make inferences about the functionalities or pathways present in these bacteria. With

sufficient sequencing depth, metagenomic sequencing captures higher resolution differentiating closely related organisms than 16s rRNA gene sequencing and is less susceptible to PCR-associated bias (Poretsky *et al*., 2014). This technique can also be utilized to explore the physiological capability of certain microbes in a sample with its ability to obtain information of certain functional gene pathways. The main limitation of metagenomic sequencing includes high cost and its dependence on gene library.

The previous two methods allow for the identification of the bacteria present in a sample, or their genetic potential. However, these methods do not identify what these microbes are functionally producing. For this, meta-transcriptome sequencing (RNA-Seq) is used. All mRNA transcripts in a sample are sequenced and then analyzed to determine relative or differential gene expression between different conditions or samples. All approaches rely on a high-throughput sequencing technology, but differ in their sample preparation, sequencing approach and bioinformatic analyses.

Studies of the microbiome require high-quality data to generate accurate results. Therefore, each step of the processing streamline requires rigorous quality control, including optimal sample collection methods, a sufficient number of samples, obtaining samples from biologically relevant sites, controlling for confounding factors, choosing appropriate sequencing platforms and using appropriate analytical tools. Each of these factors will be discussed in later sections.

1.2.1.1 Sample collection and metadata acquisition.

Optimal methods for sample collection and storage vary for different sample types. For microbe-rich samples such as stool, different collection and storage conditions yield reproducible sequencing results (Al et al., manuscript under review). However, for samples with low microbial biomass or samples in liquid or gaseous form, care should be taken during processing to avoid compromising quality. Sequencing results of samples with low microbial biomass are easily confounded by environmental or reagent contamination (Salter *et al.*, 2014), and in certain cases it is almost impossible to retrieve an accurate profiling of the microbial community in the target niche (Kennedy *et al.* 2014; D. Kim *et al.*, 2017; Salter *et al.*, 2014). Liquid samples may require optimization

for reliable storage. For example, when storing urine it is best to avoid cycles of freezing and thawing as this process damages the cellular structure of microbes, and the DNA is lost in the liquid phase of the sample. In such cases, concentrating the bacterial component of the sample and storing in pellet form may be required.

Acquisition and recording of metadata is another very important factor. Metadata is any information regarding the samples. The metadata can include but is not limited to where the sample came from, when it was collected, how it was collected, the estimated microbial abundance, risk of contamination of that environment, how that sample was processed and sequenced. The metadata that should be collected depends on the nature of the sample. With regards to a soil sample, it would be useful to record the latitude, longitude, and depth where sampling occurred. Whereas for a human oral swab it would be useful to record age, gender and all aspects of the health and/or disease status. Metadata recording is becoming increasingly important as more metagenomic and meta-transcriptomic sequences are submitted to databases; it is helpful to be able to search by a disease state, habitat or other characteristics. It is also useful to have information stored about the methods used to obtain the sequences, such as the sequencing platform used. The Genome Standards Consortium has proposed checklists for metadata such as the Minimum Information about any Sequence (MIxS), the Minimum Information about a Marker Sequence (MIMARKS), and the Minimum Information about a Metagenome Sequence (MIMS) checklists in order to standardize the metadata collected, as well as "environmental packages" for specific sample types (Yilmaz *et al.*, 2011). Beyond submission to databases, metadata are also useful for any study in which samples are being compared—the more metadata that are available, the more information a researcher has about how the samples differ and how these might affect microbiome findings.

1.2.1.2 Nucleic acid extraction and library preparation

Microbial genomic DNA extraction and purification serve as the first step of library preparation. Since most sequencing protocols require nanogram to microgram quantities of DNA, efficient DNA isolation and purification is critical to ensure proper representation of the microbial community present. Cell lysis and subsequent DNA extraction from certain microbes, especially those living in extreme environments, can be

difficult as these organisms have rigid cell wall structures and may also release stable nucleases upon lysis. However, since such microbes are an interesting source for bioprospecting new enzymes and other components, efforts are constantly made to perform those challenging DNA extractions. With the application of mechanical lysis (bead beating and sonicating), chemical lysis (Sodium dodecyl sulfate (SDS), phenol) and enzymatic lysis (proteinase K, mutanolysin, etc.), high-quality DNA has been retrieved from a variety of environments including soil (Hårdeman & Sjöling, 2006; Pathak *et al.*, 2009; Voget *et al.*, 2006; Waschkowitz *et al.*, 2009), marine picoplankton (Stein *et al.*, 1996), saline soil (Purohit *et al.*, 2009), contaminated subsurface sediments (Abulencia *et al.*, 2006), groundwater (Uchiyama *et al.*, 2005), hotsprings, mud holes in solfataric fields (Rhee *et al.*, 2005) and many other locations. To avoid wasting expensive reagents for library preparation and sequencing, the quantity and quality of the DNA need to be confirmed before amplification and sequencing. Many different technologies exist for measuring quality, such as the 2100 Bioanalyzer (Agilent Technologies), fluorimeters (Qubit, Life Technologies), spectrometers (Nanodrop) and quantitative PCR for DNA damage identification (Ginzinger *et al.*, 2000). After verification of quality and quantity, the DNA templates can go through further preparation steps.

1.2.1.3 High throughput sequencing methodologies

The first-generation DNA sequencing technology was developed by Sanger et al.(Sanger *et al.*, 1977) based on the selective incorporation of chain-terminating dideoxynucleotides and required manual determination of the sequence. The first automatic sequencing machine (AB370) was produced by Applied Biosystems in 1987 (L. Liu *et al.*, 2012). Using the Sanger technique, the first bacterial genome sequence was completed in 1995 (Fleischmann *et al.*, 1995), and constituted the main part of the Human Genome Project in 2001(Collins *et al.*, 2003). These technical advancements promoted further development of sequencing technology. With the launch of the Genome Sequencer 20 system by 454 Life Sciences in 2005, second-generation sequencing techniques came into recognition with massively parallel analysis, high-throughput and reduced cost. This significant advance reduced the difficulty of sequencing (Metzker *et al.*, 2005) and made it possible to analyze bacterial genomes in days rather than months or years (Loman *et*

al., 2012). While most second-generation sequencing techniques rely on a sequencing-by-synthesis design, the newly emerged third-generation sequencing techniques are performed on a single-molecule basis with no initial DNA amplification step. The features of each next-generation sequencing device are summarized in Table 1.1.

1.2.1.4 General considerations for high throughput sequencing

One of the first questions a scientist considering using high-throughput sequencing should ask is: Is this methodology required to answer the scientific question? While the sequencing is increasing in popularity, speed, reliability and decreasing in cost, it is not always the appropriate tool to be used. For some experiments, other methods such as microarrays may be appropriate and yield results that are more appropriate for the specific hypotheses being addressed. However, because of its the depth and relatively unbiased nature, the high-throughput sequencing technique is, in many cases, replacing microarray technologies.

If sequencing is the appropriate tool, the second question is which platform should be used. There are many different platforms on the market each with their distinct strengths and weaknesses. Some may generate the largest total throughput per run, while others may have better speed, read length or accuracy. The appropriate selection depends on the aims of the experiment. In some cases, a combination of techniques could provide higher coverage of the genome.

To increase the efficiency of sequencing, microbial samples are usually barcoded and multiplexed. Understanding the complexity of your samples is critical to determine whether multiplexing is needed and how many samples can be assayed in a single run. If using a multiplexing approach, the concentration of each sample should be similar so that equal amounts of data are achieved for each multiplexed sample.

Table 1 Comparison of next-generation sequencing techniques

Sequencing Device	Chemistry	Read Length (bp)	Run Time	Throughput per run	Reads per run
High-end instruments					
454 GS FLX + (Roche)	Pyrosequencing	700	23h	700 Mb	~1,000,000 shotgun, ~700,000 amplicon
HiSeq 2000/2500 (illumine)	Reversible terminator	2 x 150	High output: ~11 days Rapid run: ~27h	High output: 600 Gb Rapid run: 120 Gb	High output: 3 billion x 2 Rapid run: 600 million x 2
5500xl W SOLiD (Life Technologies)	Ligation	1x75 Frag, 2x50 MP	8 days	~320 Gb	1.4 billion x2
Bench-top devices					
454 GS Junior (Roche)	Pyrosequencing	400	8h	35 Mb	100,000 shotgun 70,000 amplicon
Ion PGM (Life Technologies)	Proteon detection	100 or 200	3h	100 Mb (314 chip) 1 Gb (316 chip) 2 Gb (318 chip)	400-550 thousand (314 chip) 2-3 million (316 chip) 4-5.5 million (318 chip)
MiSeq (Illumina)	Reversible terminator	2x250	27h	8.35 Gb	6.8 million (LRGC routinely getting > 15M)

1.2.2 Common applications of high-throughput sequencing and bioinformatic tools

High-throughput sequencing techniques produce massive amounts of data. To determine if and what conclusions can be made, it is necessary to analyze the information computationally. In this section, the most commonly used methods and tools for targeted amplicon sequencing analysis, shotgun metagenome analysis, and meta-transcriptome analysis will be discussed, as well as what technologies are emerging. Figure 1 details a summary of the steps for bioinformatic analyses for each of the different methodologies. Amplicon sequencing will be the major focus going forward as this technology is the primary one used for this thesis.

1.2.2.1 Targeted amplicon sequencing and analysis of bacterial diversity

Amplifying and sequencing variable regions of highly conserved bacterial genes is a common way to determine the taxonomic composition of a microbiome. By comparing them to existing databases, one can determine from which organisms the sequences are homologous to in order to make an inference into their identity, and thus determine the bacterial profiles and proportions within a sample.

1.2.2.1.1 Amplicon choice

Typically, the gene encoding the 16S rRNA subunit is used to analyze prokaryotic taxonomic composition in samples, as it is highly conserved in all prokaryotes. However, since the 16S rRNA gene is approximately 1550 base pairs in length, it is difficult to sequence the entirety of the gene using high-throughput sequencing methods without requiring an assembly step. While techniques such as expectation maximization iterative reconstruction of genes from the environment (EMIRGE) have been developed to assemble the full 16S gene for taxonomic studies (Miller, Baker, Thomas, Singer, & Banfield, 2011), it is not frequently used. Increasing the length of the sequence decreases the depth of coverage, making it more difficult to assemble the gene and resolve rare taxa.

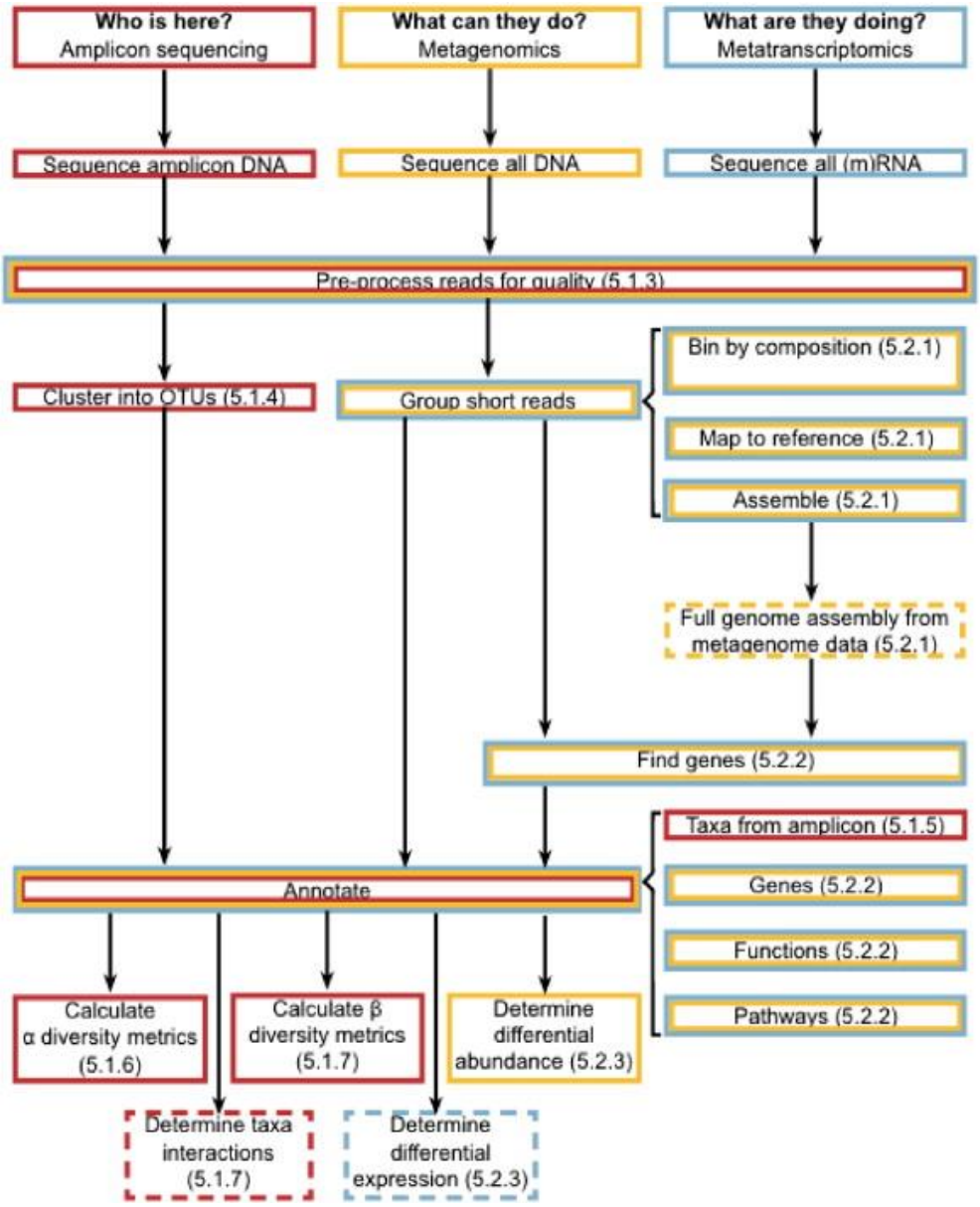

Figure 1 Flowchart detailing the steps in the bioinformatic analysis of sequence data. Boxes in red represent steps for amplicon sequence analysis, boxes in yellow represent steps for metagenome analysis, and boxes in blue represent steps for meta-transcriptome analysis

Instead, one or more of the nine variable (V) regions of the 16S rRNA gene are amplified and sequenced. However, the variable regions differ between species in different ways—some species can be distinguished in one variable region but not the other. Schloss et al. reported that the use of different variable regions influences the way OTUs cluster and the reported richness and evenness of communities. It was recommended to use caution in the analysis of these data and that the data should not be analyzed in the same way that one would analyze a full-length 16S rRNA gene sequence (Schloss *et al.*, 2011). Kim *et al*. reviewed and analyzed commonly used primers to determine how the use of different variable regions for analysis affects OTU clustering, annotation, and estimates of diversity when applied to sequences available from the Ribosomal Database Project (RDP) database. It was concluded that it is easiest to analyze the V1–V3 or V1–V4 regions, as these are more variable and are better represented in the RDP database (Kim *et al.*, 2011). Liu *et al*. also compared several designs of 16S rRNA gene amplicon sequencing and found that sequencing the V2–V3 regions produced the most accurate results, as determined by modeling short read sequence data based on full 16S rRNA gene sequences (Z. Liu *et al.*, 2008). It was also suggested that several primers in combination could be used (Z. Liu *et al.*, 2008). Youssef *et al.* compared the effects of the choice in the variable region on the amount of OTUs detected and found that V4, V5–V6, and V6–V7 sequencing produces the best estimates for species richness in a sample (Youssef *et al.*, 2009). The Earth Microbiome Project (Gilbert *et al.*, 2010) is standardized to amplify the V4 region since the primers for that region are complimentary homologous to most bacteria and archaea (www.earthmicrobiome.org).

Similar to the sequencing platform used, the choice of which variable region to amplify depends on a variety of factors specific to the sample and experiment. This includes the bacterial types likely present, level of taxonomic resolution (species, genus, or higher taxonomic level), and the read length that is afforded by the sequencer. Further studies relating to the effects of study design on the results of amplicon sequencing studies can be found in the several publications (Mizrahi-Man *et al.*, 2013; Shah *et al.*, 2011; Soergel *et al.* 2012).

While the 16S rRNA gene is by far the most frequently used gene for studies of community membership and structures, its use has limitations. One main criticism of the technique is that many bacteria have more than one copy of the rrn operon, which contains the 16S rRNA gene (Acinas *et al.*, 2004). This is problematic for abundance studies based on these sequences since bacteria with more 16S rRNA copies will be over-represented in a sample. Also, there can be high levels of sequence divergence between the multiple copies of the 16S rRNA gene present within the genome of a single bacterium, in particular when examining extremophiles (Acinas *et al.*, 2004). These divergent sequences inflate the diversity estimates as their different rRNA genes make it appear that multiple bacterial variants are present when in reality it is a single bacterium. Lastly, many of the 16S rRNA gene primers are considered "universal" for all eubacterial sequences despite the fact they may preferentially bind to sequences from some taxa over others. This will also over-represent some taxa while under-representing others.

To combat these problems, other highly conserved genes have been used as an alternative to the 16S rRNA gene. The most commonly used alternative is cpn60 that encodes the type I chaperonin proteins. Previous research has shown that cpn60 sequencing has a better resolution to distinguish species compared to the16S rRNA gene (Hill *et al.*, 2010; Schellenberg *et al.*, 2009). However, in organisms with multiple copies of cpn60, they tend to be even more divergent than 16S rRNA genes and can potentially inflate taxonomic differences similarly to the case for multiple rRNA genes (Links *et al.*, 2012). For other highly conserved genes, such as rpoB, no universal primers exist making these genes difficult to work with for studying the whole microbiome without bias. Another major disadvantage of using alternatives to the 16S rRNA gene is that databases for these genes are less well developed and thus annotation of the reads is more challenging (see Section 1.2.2.1.5 for details on annotation).

1.2.2.1.2 Commonly used software packages and pipelines

Amplicon sequencing-based studies are becoming increasingly popular, and so a variety of pipelines and tools have been developed to facilitate analyses. Most of the tools described below can be found in these packages, allowing analysis of sequence data without having to learn many different programs. The most popular pipelines are

Quantitative Insights Into Microbial Ecology (QIIME)(Caporaso *et al.*, 2010) and mothur (Schloss *et al.*, 2009). Each has the most common tools for analyses of amplicon sequencing data.

For less common tools, the code is often available separately. One useful tool widely used is the software environment known as R. This is an open source programming language that is often used for statistical analyses and visualization of data (www.R-project.org). Many software packages that are useful for analysis of microbiome data have been developed for use in R, such as vegan, used for assessing diversity (http://cran.r-project.org.proxy1.lib.uwo.ca/package=vegan), and phyloseq, used to analyze operational taxonomic unit abundance data (McMurdie & Holmes, 2013).

1.2.2.1.3 Pre-processing of raw amplicon reads

All sequencing methods have some error. Amplicon sequencing analysis is particularly vulnerable to erroneous conclusions caused by incorrectly sequenced reads, as errors in raw sequence reads can lead to erroneously high estimates of bacterial diversity (Kunin *et al.*, 2010). Clustering reads based on sequence similarity counteracts some but not all of these effects (Kunin *et al.*, 2010). Because of this, pre-processing is often used before further analysis.

Quality-based pre-processing methods are commonly used to filter out low-quality reads, reducing the number of erroneous reads. Previous studies have filtered out reads with ambiguous bases, mismatches to primers, low-quality scores, or minimal or no sequence match to the amplified sequence (Huse *et al.*, 2010). Another method that has been used is to use quality scores to trim lengths of low-quality sequence from the reads (Kunin *et al.*, 2010), or trimming a fixed number of bases from the ends of reads. This is used because sequence quality decreases with length and therefore this removes the lowest quality data from each read.

For pyrosequencing methods such as Roche 454, additional tools exist to remove errors due to these techniques' tendency to have insertions and/or deletions in their sequences at homopolymeric tracts. Tools such as AmpliconNoise (Quince *et al.*, 2011) and Denoiser

(Reeder & Knight, 2010) analyze the flowgrams from the sequencers to detect potentially erroneous reads.

Another source of sequence artifacts is chimeric sequences created during the PCR amplification step. These forms when an aborted PCR product from one sequence binds to and acts as the primer for the extension of another sequence, thus creating a hybrid DNA molecule that was made from two templates (Hess *et al.*, 2011). Like errors introduced by sequencing, chimeric sequences can be falsely interpreted as novel amplicons and thus incorrectly increase estimates of bacterial diversity (Huse *et al.*, 2010). Although the risk of chimeras forming is lower when sequencing variable regions, it is still worthwhile to check for them. A number of tools exist to detect and remove chimeric reads, including ChimeraSlayer that uses a reference set of amplicons (Hess *et al.*, 2011), Perseus, that acts de novo (Quince *et al.*, 2011), and UCHIME, that can either use a reference database or act de novo to detect chimeras (Edgar *et al.*, 2011).

1.2.2.1.4 Operational taxonomic unit (OTU) clustering

It is incorrect to assume that each unique read represents a different species: mutations and sequencing errors can lead to slightly different amplicon sequences within a species. To prevent this from interfering, reads are aligned, and those that are similar are clustered together forming operational taxonomic units (OTUs). The OTUs are defined by clustering together sequences that are below a particular percentage of divergence from each other. As this percentage increases, the number of reads clustered together increases, as the percentage decreases the number of OTUs must also decrease. Clustering of sequences in an OTU is a compromise between detecting more species than are actually present (interpreting mutations and sequence errors as separate species), and detecting fewer species than are actually present (clustering similar reads from different species into separate OTUs). Generally, a 97% identity thresh-hold is used as an approximation of species-level resolution. Meaning, all reads in a single OTU do not have greater than 3% sequence difference from each other (Stackebrandt & Goebel, 1994).

Since determining which OTU a sequence belongs to requires comparison of all input sequences with each other, this procedure can take a large amount of computational

power, and thus, a lot of time to complete. To make this task easier, pre-clustering is performed. This involves collapsing all identical reads into one category; however, it is important that the abundance information for these reads is stored for use in subsequent quantitative analyses. Once the reads are pre-clustered, existing tools can be used to determine OTUs.

There are two main independent approaches when assigning OTUs. The most commonly used is de novo OTU picking, this technique compares all sequences to each other and clusters them by similarity with no reference to outside sources of data. Cd-hit (W. Li & Godzik, 2006), UCLUST(Edgar *et al.*, 2010), ESPRIT-Tree (Cai & Sun, 2011), CROP (Hao *et al.*, 2011), BeBAC(Cheng *et al.*, 2012), and M-pick (Wang *et al*. 2013) are all tools that can choose OTUs independent of taxonomy. The other approach to OTU picking is taxonomy-based where tools such as BLAST(Altschul *et al.*, 1997) compare sequences to a database and cluster OTUs based on their sequence similarity to known sequences in the database. This approach is less likely to include erroneous sequences in the analysis and immediately annotates the clustered reads; however, it discards reads with sequences that do not match those in the database. This eliminates potentially novel species. A third option is to combine the two approaches by using taxonomy-based OTU picking first, and then using de novo clustering on the remaining reads (Bik *et al.*, 2012).

1.2.2.1.5 Annotation

Once OTUs are determined, it is necessary to find out from what species they came. The database used is dependent on the sequence amplified. For 16S rRNA gene sequences several databases exist. The SILVA database has sequences for both the small and large ribosomal subunits in prokaryotes as well as eukaryotes. Taxonomic classification of rRNA sequences can be done using the SINA tool, available on its website(Pruesse et al., 2007). The Ribosomal Database Project (RDP) contains prokaryotic 16S rRNA sequences and has an RDP Classifier tool for taxonomic classification of rRNA gene sequences, as well as several other tools for 16S rRNA amplicon processing and analysis(Cole *et al.*, 2008). GreenGenes also contains prokaryotic 16S rRNA sequences along with a tool to compare sequences to the database (DeSantis *et al.*, 2006). There are fewer databases dedicated to other prokaryotic marker genes. For cpn60 sequences, the

cpnDB exists to aid annotation (Hill *et al.*, 2004). For other markers, there are few or no dedicated databases. In these cases, it is thought to be best to search against a general purpose protein database.

1.2.2.1.6 Describing richness and evenness within samples

Richness is a description of the number of species present in a sample, while diversity is a measure that combines richness with how evenly different species make up the sample's microbiome. Richness may be determined by the number of OTUs present, but this can be influenced by the sequencing depth, so metrics such as Chao1(Chao *et al.*, 1984) and ACE (Chao & Lee, 1992) are used to estimate the minimum number of species present in the sample. Typically, the diversity within a sample is referred to as alpha-diversity. It is measured using the Simpson diversity index and the Shannon diversity index; a detailed description of alpha-diversity indices and application in microbiome research can be found by Li et al. (Li *et al.*, 2012).

1.2.2.1.7 Comparing samples

Diversity between samples is often referred to as beta-diversity. At its simplest, samples can be compared by the OTUs shared, or by visualizing the relative abundance of different OTUs in a heatmap. A more detailed approach can be used that considers differences in species presence or absence, abundance, or phylogeny. This requires the use of multivariate analyses. Commonly used tools and some emerging techniques are described below. For a more detailed description of the statistical background of many commonly used tools, the review by Remette et al. is helpful (Ramette *et al.*, 2007).

To begin to compare samples, a distance matrix is made that quantifies the differences between samples. This analyzes which taxa are present, and may or may not consider phylogeny or abundance of different species. Phylogeny-independent metrics treat all sequences equally regardless of their similarity. Sørensen's similarity matrices and the Jaccard index measure differences qualitatively by determining the distance between samples based on the presence or absence of OTUs. The Bray–Curtis and Morisita similarity indices measure differences quantitatively based on the abundance of OTUs in each sample (Chao *et al.*, 2006). These metrics do not consider differences in sequencing

depth, and so sequences must be normalized, or alternative methods that consider unobserved species must be used (Chao *et al.*, 2006; Engen *et al.*, 2011). Conversely, phylogeny-dependent metrics consider samples with more similar sequences to be more similar. This is done by first putting the reads on a phylogenetic tree. This can be constructed de novo by aligning and comparing samples, or by using a previously constructed reference tree. A good reference tree may more accurately represent the phylogenetic difference between species; however, a de novo tree will allow analysis of sequences that do not match anything on the reference tree. The most commonly used phylogeny-dependent similarity metric is the UniFrac distance. The unweighted UniFrac distance considers the similarity between sequences in the samples, but does not compare the sequences' abundances between samples, while the weighted UniFrac distance considers both the similarity between sequences and the sequences' abundances (Lozupone, Lladser, Knights, Stombaugh, & Knight, 2011).

Once the distances between the samples can be determined, they can be compared by their similarities. One approach is clustering the data by similarity. Methods such as the neighbor-joining method or the unweighted pair-group method with arithmetic mean (UPGMA) can be used to arrange the distance data into clusters of samples, that can then be visualized as a dendrogram to show the samples that are most related. Squash clustering is a novel method that uses the Kantorovich–Rubenstein distance (in microbiome analyses this functions in a similar fashion as a weighted UniFrac distance) to cluster samples by their similarity, taking into account similarity and phylogeny; this method is available in the software package “guppy” (Matsen IV & Evans, 2013). Unlike other methods, in a dendrogram constructed by squash clustering the distance drawn between internal nodes (points that connect multiple samples) corresponds to the difference between them.

Another approach to comparing samples is the use of principal coordinates analysis (PCoA). Here, the points (*i.e.*, the samples) are plotted in multidimensional space so that the distance between them is as close as possible to that in the distance matrix calculated earlier. The direction in this multidimensional space that separates the points the most is the first principal coordinate, the direction that separates the points second best is the

second principal coordinate, and so on. Each principal coordinate has a percentage value associated with it that represents the percent of the variation explained by that dimension. One can plot the points where the axes represent the first and second principal coordinates (and for a three-dimensional plot, the third principal coordinate). This allows visualization of the distance data by reducing its dimensionality—points that cluster, in this case, have more similar microbiomes as determined by the distance matrix determined earlier. Points can be color-coded by metadata to qualitatively examine if different types of samples cluster or form gradients along any components. Biplots can then be used to show the abundance of different taxa and in which samples these taxa are found.

A variation of this method was published recently, called edge principal components analysis (edge PCA) (Matsen IV & Evans, 2013). Edge PCA uses a phylogeny-dependent distance metric that considers the relatedness of taxa, and the resulting principal components correspond to differences in particular clades of bacteria. This method produces an edge PCA plot that can be interpreted in a similar way to a PCoA plot. It also produces a phylogenetic tree of the organisms in the sample in which organisms that drive a given principal component have their edges color-coded (by whether they influence the principal component in the positive or negative direction) as well as having different thicknesses of the edges depending on how much they drive that principal component. This method is useful if differences in samples are driven by closely related taxa. Edge PCA can be performed using the "guppy" software package.

The above methods are mainly exploratory. For more specific analysis, nonparametric hypothesis testing tools are implemented. These include permutational multivariate analysis of variance (PerMANOVA, also referred to as nonparametric multivariate analysis of variance, NPMANOVA) (Anderson *et al.*, 2001), analysis of similarity (ANOSIM) (Claerke *et al.*, 1993), and the Mantel test (Mantel *et al.*, 1967). These can be used with any distance metric to determine if the microbiome differs significantly between groups of samples. Recently, parametric tools have been developed to test hypotheses related to the microbiome, such as those found in the HMP R package. These

tools are more powerful than the nonparametric tests but generally require pooling rare taxa (La Rosa *et al.*, 2012).

1.2.2.1.8 Describing interactions between bacteria

Amplicon sequencing and analysis can estimate the composition of different samples. Therefore, one can theoretically determine interactions between bacteria, for example, which bacteria co-occur and which co-exclude. However, this analysis is deceptively challenging. Since abundances of bacteria determined by these methods are proportional data, they are not independent; if one goes up, one or more others must go down. Standard measures of correlation assume that the data are independent, thus it is not valid to apply these to the proportional data, and doing so may result in spurious correlations (Aitchison *et al.*, 1986). Because of this, specialized tools must be made for these analyses.

There is currently no tool to analyze co-occurrence or co-exclusion that is frequently used in microbiome analysis, although some tools are under development. Faust et al. used a novel analytic technique to determine co-occurrence and co-exclusion and analyze the relationships' significance for human microbiome data; however, a computational tool to analyze data in the same way was not developed (Faust *et al.*, 2012). Friedman and Alm produced a tool, SparCC (Friedman & Alm, 2012), that applies the log-ratio transformation (Egozcue *et al.*, 2003) to transform proportional data into independent data and then approximates correlations to find co-occurring and co-excluding bacterial pairs. The problem of determining co-occurrence and co-exclusion based on abundance data is by no means resolved, and developing statistically sound methods that accurately represent and interpret these data is an active area of research.

1.2.3 Future perspectives of high-throughput sequencing

High-throughput sequencing technologies have improved in output and quality, and have become an indispensable tool for an increasingly wide variety of experiments, including in phylogenetic, diagnostic, and ecological contexts. Through these tools, scientists can gain insight into the composition, activities, and dynamics of a wide variety of

microbiomes. This will help to elucidate how bacteria interact with each other and how they interact with their environment.

Updates to the laboratory and computational tools are ongoing. In the near future, this will result in sequencers that require less handling (*e.g.*, removal of the amplification step) in order to gain more automation and less error, while still maintaining high speeds and throughputs with low costs and error rates. On the computational side, tools are becoming faster and more specialized for multi-species samples, while also being able to investigate new questions such as differential expression from high-throughput RNA sequence data, genome assembly from metagenomic sequences, and interactions between bacteria from amplicon sequence data. Many of these tools have a steep learning curve for a biologist who is not computationally inclined, although specialized workflows are being developed to allow analysis of these data without extensive computational experience. As tools and statistical methodologies are refined, metagenomics and meta-transcriptomics will be able to shift the focus of microbiome research from "Who is here?" to "What are they doing?". This will allow deeper insight into how bacteria interact with their environment and each other. Incorporation of metabolomes, nutrients, host genomes, and other metadata will allow the production of an incredibly detailed picture of interactions between the microbiome and its environment, helping us understand the role of prokaryotes in all environments in which they live.

1.3 Functional anatomy of the urinary tract

The urinary system dynamically exerts a cleaning and waste disposal function of the human body. It is composed of the upper urinary tract (kidney and ureter) and lower urinary tract (bladder and urethra). The two kidney resides bilaterally in the retroperitoneal space, surrounded by renal capsule, perirenal fat, Gerota's fascia and pararenal fat. Being a bean-shaped organ, a kidney possesses a recessed area on the concave border called renal hilum, where the renal vasculatures and ureter enters. The kidney has high blood throughput, as it receives about 20% of the cardiac output, and filters approximately 150 liters of blood per day, the condensed output contributing about 1.5 liters of urine after renal reabsorption, secretion and excretion. The urine is then

emptied into the minor calyces by renal papilla, collected into the major calyces and renal pelvis, and flows into the ureter.

The two ureters are long, tubular structures that connect the kidney to the urinary bladder. The smooth muscle fibers in its wall form a peristaltic pressure wave that pushes the urine moving forward. The urine flow rate at the ureter is about 0.5 to 5 ml per minute. The ureteral peristalsis, together with the anti-reflux structure at ureterovesical junction warrants the unidirectional flow of urine. This also serves as one of the defensive mechanisms of the urinary system against bacterial ascension to the kidney.

The urinary bladder is a hollow organ located in the pelvis. It serves to contain urine, making periodic voiding by alternative functioning of its detrusor and sphincter muscles. Since the bladder is in close proximity to the exterior environment and urine in the bladder remains relatively stationary, the bladder is the most common site for adult urinary tract infections (UTIs). Because of this, the bladder possesses a vast catalogue of defensive mechanisms that resist bacterial adhesion and colonization. Commonly recognized lower urinary tract antibacterial defense mechanisms are listed in Table 2.

The urethra is the final pathway before urine exits the human body. The length of the male urethra is approximately 20 cm while the female one is about 4 cm. Although natural defenses against bacterial infection are present, the shorter urethral length in females is a major contributing factor to their increased vulnerability to UTIs. Structural abnormality of the urinary tract remarkably compromises its natural resistance to bacterial invasion. Typically, urinary stasis is the direct result of the structural abnormality, and is the core cause of UTI. Common causes of urinary obstruction are listed in Table 3. Obstruction compromises the urine flow, and urinary stasis is a crucial risk factor or infection. Urinary stones are a common cause of urinary stasis. Hormonal and mechanical changes in the urinary system of pregnant women could lead to physiological urinary stasis, making them among highest risk groups in females. For males, the ones with congenital abnormalities and bladder outlet obstruction due to benign prostate hypertrophy are usually prone to obstruction to the free flow of urine.

Table 2. Antibacterial host defense mechanisms in the urinary tract (Jack D. Sobel, 2015)

Urine (osmolality, pH, organic acids)
Urine flow and micturition
Urinary tract mucosa (bactericidal activity, peptides, cytokines) - innate immunity Inflammatory response
Polymorphonuclear neutrophils (PMNs)
Tamm-Horsfall protein
Cytokines
Low-molecular-weight oligosaccharides
Secretory immunoglobulin A (SIgA)
Lactoferrin
Peptides
Adaptive immune system, humoral immunity
Cell-mediated immunity
Miscellaneous prostatic secretions

Vesicoureteral reflux is another anomaly-associated risk factor for UTI. Urine reflux contributes to upper urinary tract infection by promoting bacterial invasion via the ascending route and perpetuates infections by inducing postvoid residual urine.

Table 3. Common cause of urinary tract obstruction (Jack D. Sobel, 2015)

Extrarenal obstruction
Congenital anomalies of the ureter or urethra
Valves
Stenosis
Bands
Calculi
Extrinsic ureteral compression
Benign prostatic hypertrophy
Intrarenal obstruction
Nephrocalcinosis
Uric acid nephropathy
Analgesic nephropathy
Polycystic kidney disease
Hypokalemic nephropathy
Renal lesions of sickle cell trait/ disease

1.4 Role of microbiota in recurrent urinary tract infections (rUTI)

Urinary tract infections (UTIs) are the second most common bacterial infection in humans. Of these, 20-25% of patients will experience a recurrence of their UTI (rUTI) (Bsrsv Foxman, 1990). This can lead to considerable morbidity and health care costs. Young, sexually active females are the most common sufferers of UTIs. About 50% of women will experience a UTI at least once throughout their lifetime, of these, 20-25% will experience rUTI and 3% will have this recurrence within six months following treatment (Foxman *et al.*, 2000). Known risk factors for rUTI include hormone deficiency, diabetes, problems associated with urination, kidney stones and foreign objects within urinary tracts (*e.g*. ureteral stents). Another less widely known factor is the disruption of the vaginal microbiota, but it is believed this may also extend to other potential reservoir sites, such as the rectum (Bergeron *et al.*, 2012).

1.4.1 Etiology of rUTI

The cause and etiology of rUTI may be quite different between patient populations. For a healthy young woman, the most common cause of rUTI is uropathogenic E.coli (UPEC) of rectal origin colonization in the vagina, periurethral and distal urethra (Stamey *et al.*, 1971). Pediatric rUTIs are usually associated with conditions such as vesicoureteral reflux (VUR) (Conway *et al.*, 2007). Male rUTIs are relatively rare and may be associated with urinary stasis following urinary tract obstruction, dysfunction and instrumentation. Bacterial colonization in prostate also contributes to male UTI recurrence (Carter, Levy, & Zeitlin, 2012).

1.4.2 Reservoirs for rUTI

Pathogens of rUTI typically originate from the rectum, sometimes replacing the normal microbiota in the vagina. This occurs due to a state of microbiome flux, such as during bacterial vaginosis, which allows these pathogens to colonize the periurethral area, urethra and ascend into the bladder. Women with rUTI had a higher incidence of uropathogens colonizing in the vagina with longer duration and higher severity compared to women without rUTI, suggesting the role of the vagina in harboring uropathogenic

strains that constantly invades urinary tract (Brumfitt *et al.*, 1987). In fact, as high as 68% of recurrences are caused by the same bacterial strain that caused the initial infection (Foxman *et al.*, 2000). It is documented that this same strain could take as long as three years after the initial UTI to cause recurrence (Gillespie *et al.*, 2000). In a normal vagina, bacteria are present to prevent pathogens from causing infection. *Lactobacillus crispatus* and *Lactobacillus iners* are the predominant species (Burton & Reid, 2002). It is known these bacteria can prevent pathogens from multiplying to a symptomatic level by releasing lactic acid, hydrogen peroxide (H_2O_2), and neutralizing uropathogens via coaggregation effects (Reid *et al.*, 1990). To cause initial infection, several daily activities may facilitate bacteria migrating from the rectum into the vagina; wiping from the anus towards the vagina after a bowel movement, retaining urine before voiding (*i.e.*, not urinating when the initial urge is present), sexual intercourse and using a diaphragm or spermicide for birth control. Women who experience rUTI often have specific vaginal microbiota pattern shift. There is an increased proportion of E. coli and depletion of normally predominant H2O2-producing lactobacilli (Gregor Reid et al., 2011). Although it is not known when this shift occurs, it is possible it is not a result of the infection but may predispose women to rUTI (Dwyer & O'reilly, 2002).

1.4.3 Pathogens of rUTI

Bacteria are the most common cause of UTIs, with UPEC causing the vast majority (up to 90%) of all diagnosed cases (Schmiemann *et al.*, 2010). When the first infection is caused predominantly by *E. coli*, secondary UTIs may develop within six months due to other organisms (Scholes *et al.*, 2000). Other common uropathogens include *Staphylococcus saprophyticus, Enterococcus faecalis, Klebsiella pneumoniae,* and *Proteus mirabilis* (Grabe *et al.*, 2009).

1.4.4 Pathophysiology of rUTI

1.4.4.1 Bacterial adhesion

Bacterial adhesion to host tissues serves as the first critical step in UTI pathogenesis. UPECs express a wide variety of virulence factors, including adhesion factors, hemolysin, siderophores and certain O antigens to exert their pathogenesis (Cadieux *et*

al., 2009). UPEC can express two types of pili that adhere to specific receptors on the uroepithelium. Interestingly, lactobacilli by-products have been shown to down-regulate genes associated with these functions in UPEC strains (McMillan *et al.*, 2012). The type 1 pilus adheres to mannosylated uroplakin proteins on the luminal surface of the bladder (Wright & Hultgren, 2006), while P pili bind to globoseries glycolipids on the kidney epithelial surfaces (Lane & Mobley, 2007). Some people are more susceptible to UTIs based on a variety of factors. For example, individuals with an A, B, or AB blood type are known as "non-secretors" as they do not secrete the normal antigens for that blood type in bodily fluids. Non-secretors are more prone to recurrent UTIs because the cells lining the vagina and urethra allow bacteria to attach more readily (Kinane *et al.*, 1982).

1.4.4.2 Internalization and intracellular persistence of uropathogens

Traditionally, bacteria responsible for UTIs were thought to exert their pathogenesis on an extracellular basis. However, evidence now suggests UPEC strains encode mechanisms to invade or be internalized by the host bladder epithelial and kidney epithelial cells (Bishop *et al.*, 2007). Similar invasive activities are observed in other uropathogens, including S. saprophyticus (Szabados *et al.*, 2008), K. pneumoniae (Rosen *et al.*, 2008), and S. enterica. Although the exact mechanism of invasion remains unclear and might differ between pathogens, bacteria also take advantage endocytosis by the epithelial cells (Bishop *et al.*, 2007; Rosen *et al.*, 2008). The pathogens avoid being expelled by exocytosis and escape the endocytic compartments to enter the cytoplasm (Eto *et al.*, 2006). Once in the cytoplasm, they are able to multiply. Protected from the host immune responses, pathogens can invade further into the underlying transitional epithelium to establish a quiescent intracellular reservoir (QIR) that can serve as the source of rUTIs (Hunstad & Justice, 2010). Studies based on murine models have reported that the bacteria within a QIR persist for months after initial UTI (Schilling *et al.*, 2002), avoid host immune attack and systemic antibiotic therapy, and serve as the source for rUTI without re-inoculation into the bladder (Agace *et al.*, 1993).

1.4.4.3 Immune evasion strategies of the pathogen

Unlike many other organs, the immune response of the bladder is characterized by increased levels of urinary cytokines and neutrophil influx (Agace et al., 1993). This is dependent upon the innate immune responses (Nielubowicz & Mobley, 2010) rather than adapted responses (Chan, John, & Abraham, 2013). This, to some extent, explains why chronic and recurrent UTI is common. The bladders' innate immune response is initiated by the activation of host pattern recognition receptors of the Toll-like receptor pathway, including Toll-like receptor 4 (TLR4) (Schilling, Martin, Hung, Lorenz, & Hultgren, 2003). When urothelial cells expressing TLR4 recognizes lipopolysaccharide (LPS) from Gram-negative bacteria, the proinflammatory and prosurvival NF-κB pathway is activated, and chemokines/cytokines including IL-6 and IL-8 are secreted (Akira *et al.*, 2006). These inflammatory chemokines accumulate in the bladder mucosa and urine and further induce an influx of neutrophils into the bladder and in turn phagocytosis and clearance of bacteria. Though UPEC was reported to induce an adaptive immune response in murine models (Thumbikat *et al.*, 2006), pathogen-specific antibodies cannot be detected in the serum and urine of patients with bladder-restricted infection (Thumbikat *et al.*, 2006).

Uropathogens can suppress inflammatory responses activated via TLR4 and other receptors on the epithelial and bladder-resident immune effector cells. UPEC is able to stabilize IκB and suppress NF-κB activity, thus hampering the increased apoptosis of bladder epithelial cell upon exposure to UPEC (Klumpp *et al.*, 2001), lowering expression levels of IL-6 and IL-8, and suppressing secretion of these cytokines (Klumpp *et al.*, 2001). UPEC was also reported to downregulate the neutrophil gene related with inflammatory signaling and in neutrophil chemotaxis, adhesion, and migration, thus delaying neutrophil influx during very early stages of infection, facilitating bacterial internalization (Klumpp *et al.*, 2001). Interestingly, clinically isolated UPEC strains are less susceptible to neutrophil killing in vitro and more resistant to reactive oxygen species (ROS) and antibacterial peptides released by polymorphonuclear leukocytes and other host cells in response to UPEC infection (Klumpp *et al.*, 2001).

1.4.5 Role of microbe in management for rUTI

1.4.5.1 Lactobacillus probiotics for rUTI

Since depletion of normal vaginal H2O2-producing lactobacilli is associated with increased risks of UTI and recurrence, it seems to be a reasonable strategy to replete the normal vaginal microbiota with lactobacilli to reduce the risk of UTI. Certain *Lactobacillus* and avirulent *E. coli* species have been studied for their probiotic potential (Barrons *et al.*, 2007). Various clinical trials have reported to restore the protective vaginal microbiota via oral (Beerepoot *et al.*, 2012), intravaginal (Baerheim *et al.*, 1994; Reid *et al.*, 1992; Reid *et al.*, 1990) and direct intravesical delivery for rUTIs (Hagberg *et al.*, 1989). When studies using ineffective strains or testing for safety were excluded, most studies have confirmed a reduced recurrence rate in patients receiving *Lactobacillus*-containing probiotics. Individuals receiving intravaginal probiotic therapy seemed to have a lower recurrence rate, and those who did recur did so on a longer interval between episodes (Stapleton *et al.*, 2011). These studies showed the importance of maintaining good vaginal ecology for preventing rUTI. Although probiotic treatment might generally have a slightly inferior therapeutic effect (Beerepoot *et al.*, 2012) when compared to traditional long-term antibiotic prophylaxis, this antimicrobial-sparing, well-tolerated strategy could possibly reduce the UTI-associated costs and morbidity, including time lost from work, health care costs, antimicrobial use, and the ultimate development of antimicrobial resistance (Russo & Johnson, 2003). More evidence with larger sample size and validated dosing strategies are needed to confirm the efficacy of this approach.

1.4.5.2 Future probiotics and microbial therapies

Currently, a dilemma exists over antibiotic usage in rUTI management and the development of alternative therapies, including probiotics. This is because many of these alternative therapies are susceptible to antibiotic use. Besides the probiotic agents mentioned above, several strategies have been undertaken to develop new probiotics including developing avirulent E.coli strains, new lactobacilli and other strains. Avirulent *E.coli* strains (*i.e.*, strains few virulence markers of adhesion, specific O and K serotypes,

hemolysin production, and resistance to serum effects) could serve as competitive colonizers of the bladder (Hagberg *et al.*, 1989). When instilled into the bladder, they colonized the mucosa and may interfere with the onset of symptomatic infection. Instillation of avirulent E. coli 83972 into the bladders successfully created a long-term, antibiotic-resistant colonization that reduced symptomatic infection (Hull *et al.*, 2000). Another strain, *E.coli* Nissle 1917, has been verified by previous studies for its ability to colonize the bowel, displacing gut pathogens as well as boosting the innate immune response of the gastrointestinal tract (Henker *et al.*, 2008). Recent in vitro studies have observed the ability of *E. coli* Nissle 1917 to eradicate or out-grow ~75% of known uropathogens, suggesting its potential of being used as a urinary probiotic.

The horizon of *Lactobacillus*-containing probiotics is continuously being widened. Besides its abilities to restore the normal vaginal microbiota and prevent UTI, recent studies reported that *L.rhamnosus* GR-1 was able to strengthen the immune response of the urinary system (Karlsson *et al.*, 2012), and that oral application of Lactobacillus mixture on pregnant mice would reduce the risk of UTI on the newborn (Lee *et al.*, 2013). It is also quite possible that new strains of Lactobacillus, both natural and genetically-modified, together with other bacterial species, could be applied for rUTI prevention.

Even if there is a predisposition to having UPEC at sites beyond the urogenital tract, such as the colon and rectum, probiotics and other microbiota modifying therapies may reduce microbial carriage and subsequent reinfection risk. While extreme therapies such as fecal transplant are becoming more common for chronic intestinal infections, such as recurrent *Clostridium difficile*, such therapies will not be adopted quickly for UTI (Ettinger *et al.*, 2013). If more was understood about the microorganisms that naturally occupy the bladder, their role in maintaining a homeostatic environment, immunomodulatory effects and direct antagonistic effects on pathogens, a probiotic or “synthetic urine” treatment to restore the “healthy” bladder microbiota might exist.

1.5 The urinary microbiota: the uncultivable side

Urine was once considered sterile and only contained bacteria during infection. For most clinicians, the presence of bacteria in urine is a pathological condition with its severity varying from a mild asymptomatic bacteriuria to a lethal urosepsis. Such ideas are formed due to the fact that culture-based techniques remain as the "gold standard" for microbial pathogen detection for clinical practice. However, recent studies have shown that although commonly used methods for clinical sample culture are effective at detecting fast-growing pathogenic bacteria, it will neglect some slow-growing, fastidious commensal microbial species for which their role in disease prevention and homeostasis has not yet been determined. With the introduction of high-throughput DNA sequence-based techniques, several studies have identified evidence of commensal microbial communities in the urinary bladder, especially in female populations. Wolfe et al. was the first reported evidence of uncultivated bacteria in the urinary bladder (Wolfe *et al.*, 2012). Since then, multiple studies have investigated the urinary microbial communities using sequencing-based techniques among various populations, as demonstrated in Table 4. Guided by the sequencing results, enhanced culture methods based on increased sample volume, extended culture time, a wide spectrum of media and atmospheric conditions were developed that could isolate and identify many urinary microorganisms that traditional culture does not (Hilt *et al.*, 2014). Although still preliminary, these studies have shown that the urinary microbiota significantly differs between male and females, with the female microbiota containing higher abundance and greater diversity, which was undoubtedly associated with microbial ascension from the lower genital tract (Wolfe & Brubaker, 2015). These discoveries have drawn interest in exploring the association of urinary microbiota with several urinary conditions

Table 4. published studies characterizing the urine microbiota in human populations

Study	Patient Demographics	Sample Collection Method
Nelson *et al.*, 2010	Males with STI (n=10) Males negative for STI (n=9)	First-void urine
Dong *et al.*, 2011	Males with STI (n=10) Males negative for STI (n=22)	First-void urine
Siddiqui *et al.*, 2011 (	Healthy females (n=8)	Clean-catch midstream urine
Fouts *et al.*, 2012	Healthy controls (n=26; females=58%) Neurogenic bladder patients (n=27, female=48%)	Midstream urine, Intermittent Catheterization, Foley Catheter
Nelson *et al.*, 2012	Healthy adolescent males (n=18)	First-void urine
Siddiqui *et al.*, 2012	Females with IC (n=8)	Clean-catch midstream urine
Wolfe *et al.*, 2012	Healthy females (n=12) Females with pelvic organ	Clean-catch midstream urine, suprapubic aspirate, transurethral catheter

	prolapse or urinary incontinence (n=11)	
Lewis *et al.*, 2013	Healthy males (n=6) Healthy females (n=10)	Clean-catch midstream urine
Fricke *et al.*, 2014	First renal transplant patients (n=60; females=37%)	Not described
Hilt *et al.*, 2014	Healthy females (n=24) Females with overactive bladder (n=41)	Transurethral catheterization
Pearce *et al.*, 2014	Healthy females (n=58) Females with urgency urinary incontinence (n=60)	Transurethral catheterization
Willner *et al.*, 2014	Acute uncomplicated UTI patients (n=50; females=76%)	Midstream urine
Horwitz *et al.*, 2015	Aged adults with chronic urethral catheter (n=11, male = 82%)	Urethral catheter section
Thomas-White *et al.*, 2015	Healthy females (n=60) Females with urinary urgency incontinence (n = 74)	Transurethral catheterization

Shoskes *et al.*, 2016	Chronic Prostatitis/Chronic Pelvic Pain Syndrome patients (n=25) Healthy males (n=25)	Midstream urine
Karstens *et al.*, 2016	Female urinary urgency incontinence patients (n=10) Healthy females (n=10)	Transurethral catheterization
Modena *et al.*, 2017	Renal transplantation patients (n=25) Healthy individuals (n=20, males = 50%)	Midstream urine
Groah *et al.*, 2016	Neuropathic bladder patients (n=24) Healthy controls (n=23)	Midstream urine
Rani *et al.*, 2016	Renal transplantation patients (n=21) Healthy controls (n=8)	Midstream urine
Gottschick *et al.*, 2017	Women with BV/receiving treatment (n=109/43)	Midstream urine

	Healthy controls (n=80 male=39%)	
Shrestha *et al*., 2017	Prostate cancer patients (n=65)	Midstream urine
	Males with benign prostate (n=65)	

1.6 Common urinary conditions and their association with the urinary microbiota

Multiple studies have suggested that the human urinary microbiota may contribute to symptoms or pathogenesis of many urinary conditions. As initially reported by Wolfe *et al.*, the commensal urinary microbiota can be detected by both sequencing-based and enhanced culture methods (Hilt *et al.*, 2014, Wolfe *et al.*, 2012), and are clearly different form contaminating microorganisms and pathogens causing UTI (Nienhouse *et al.*, 2014, Siddiqui *et al.*, 2011). Pearce *et al.* examined the urinary microbiota of females with urgency urinary incontinence and found several bacterial species more common in the diseased state than in healthy controls, including *Actinobaculum, Actinomyces*, *Aerococcus*, *Arthrobacter, Corynebacterium, Gardnerella*, *Oligella*, *Staphylococcus*, *Streptococcus* and *L.gasseri* (Pearce et al., 2014). Siddiqui *et al*. used 454 pyrosequencing to investigate urine samples from women with interstitial cystitis and found a remarkable increase in Lactobacillus together with decreased microbial diversity (Siddiqui *et al*., 2012). Fouts *et al.* compared urinary microbiota pattern of patients with spinal cord injury and found *Lactobacillales, Enterobacteriales, Actinomycetales*, *Bacillales*, *Clostridiales*, *Bacteroidales*, *Burkholderiales*, *Pseudomonadales*, *Bifidobacteriales,* and *Coriobacteriales* to be the most abundant bacterial taxa (Fouts *et al.*, 2012).

1.7 Questions and challenges associated with studying the microbiota of the urinary tract

1.7.1 Common questions of urinary microbiota studies

1.7.1.1 What do they do in the urinary tract?

The urinary microbiota shares some similarities with microbiota of the lower genitalia, but with a lower abundance. Questions have been raised about the function of these commensal urinary microbes. In the lower urogenital tract, such as the vagina, it is known that certain bacteria act as sentinels and protectors against potentially dangerous microorganisms that may ascend further and cause an infection, such as UTI (Beerepoot *et al.*, 2012; Beerepoot *et al*., 2016). It is not known if or how the microbiota may play a protective role in the bladder, but it is assumed that the lower urogenital tract is the gateway to this region. Studies demonstrating probiotic treatment to the urogenital tract reduces UTI recurrence rates supports this idea (Beerepoot *et al*., 2016). The numbers of commensal bacteria in the bladder are thought to be considerably less than the lower genital tract, and several remain to be cultured(Whiteside *et al*., 2015). Even though bacterial numbers are lower in this region, they still have the potential to out-compete pathogens for necessary nutrients, produce antimicrobial substances and stimulate the immune system.

Where bacteria colonize other human surfaces, they play a key role in the homeostasis of the mucosal environment and numerous studies detail these on skin, buccal cavity, nasal, small intestine, bowel, lung and vaginal interactions (Cosseau *et al.*, 2008; Dickson *et al*., 2015; Hummelen *et al.*, 2011; Lopetuso *et al*., 2014; Pfefferle & Renz, 2014; SanMiguel & Grice, 2015). It makes sense to presume that even while there may be fewer bacteria in the urinary system than many other colonized sites, these microorganisms still play a significant role in the maintenance of the uroepithelium.

1.7.2 Is the urinary microbiota associated with urinary malignancy?

The theory of inflammation causing bladder cancer dates back to Virchow's report in 1860 (Virchow & Chance, 1860) and presently remains under controversy. Multiple epidemiological studies have looked at the association between UTI and bladder cancer

yielding conflicting results (Abol-Enein *et al*., 2008; Hartge *et al.*, 1990; Kantor *et al.*, 1984; Kjaer *et al*., 1989; Pelucchi *et al*., 2006; Pisani *et al*., 1997). It is understandable since most studies have relatively small sample size, and had a cross-sectional or retrospective design. Vermeulen *et al.* investigated 1809 bladder cancer patients and compared their data to 4370 control participants. The results showed that regular cystitis is a contributing risk factor for bladder cancer, and this risk may be alleviated by receiving antibiotics for UTI treatment (Vermeulen *et al*., 2014). It is with the hope that the introduction of meta-omics techniques could shed further light on this mystery.

1.7.2.1 How far up the urinary tract do bacteria go and how commonly are they found in the kidney?

While urine passes from the kidney through the urological system to exit in the external environment, the kidney is not typically considered to be routinely occupied by bacteria. Under pathological conditions such a severe UTI or urine reflux, bacteria may invade the kidney by an ascending route causing pyelonephritis, pyonephrosis, and urosepsis. Urinary stasis caused by obstructions of the upper urinary tract also predisposes the kidney to bacterial infection. However, in case of upper urinary obstruction such a stone in the upper ureter, there are no factors promoting bacterial ascending, making it difficult to explain the source of bacteria in the kidney with current theories. In preliminary studies looking at the microbiota within kidney stones, it appears that bacteria can be detected both by culture and non-culture microbiological techniques (Barr-Beare *et al.*, 2015). Interestingly, preliminary data of five pediatric subjects suggests that bacteria may be detected as far as the kidneys and may be present within the kidney stones themselves. The nidus for kidney stone formation and precipitation from urinary supersaturation is thought to be inflammation. The potential for bacteria to cause this is one possibility and in murine studies, it was shown that the presence of calcium oxalate deposits increases pyelonephritis risk, likely due to preferential aggregation of bacteria on and around calcium oxalate crystals (Barr-Beare *et al.*, 2015).

1.8 Microbiota investigations of human tissue with low microbial biomass

While the urological system has been shown to have bacteria within it, these bacteria are in relatively small numbers compared to other sites occupied by bacteria in the human body (Lewis *et al.*, 2013). The processing and handling of samples with low microbial biomass are much more challenging than conventional microbe-rich samples as the sequencing results can be overshadowed by "noise" reducing the reliability when studying a microbiota of interest. Reagent contamination is a known issue for molecular biology experiments (Champlot *et al.*, 2010; Shen *et al*., 2006), it is also recognized in metagenomics studies (Glassing *et al*., 2016). In 2014, Salter *et al*. showed a serially diluted pure culture of *Salmonella bongori* samples then subjected to DNA extraction, 16S rRNA gene amplification and sequencing yields growing the proportion of background noise in sequencing reads as the dilution proceeds. The number of PCR cycles also affected sequencing quality. For samples with low microbial biomass, 20 cycles were too low to generate sufficient amplicons while 40 cycles generated too much noise (Salter *et al.*, 2014). The optimal PCR cycle number may need to be optimized for each sample type or batch of samples. On the other hand, Kennedy and colleagues showed that the template concentration in samples, *i.e.*, the input microbial biomass, also affected sequencing quality, weakening the microbial vs. background ratio (Kennedy *et al.*, 2014). Since then, multiple studies have confirmed the existence of inherent bacterial DNA in reagents and environment, contributing contaminating microbes to a total number of 181 genera (Glassing *et al.*, 2016; Shen *et al.*, 2006; Weiss *et al.*, 2014). The term "kitome" or "contaminome" was then introduced to describe these contaminating reagent microbial DNA (Dickson *et al.*, 2017; Kverka *et al*., 2017). The kitome differs by manufacturer and model, kit batch, and even in different tubes or lots of one product (Glassing *et al.*, 2016; Weiss *et al.*, 2014). Therefore, it is crucial to process all samples in a project in a short period, preferably using one kit if possible. Recording the type of kit used, together with its batch number is strongly recommended. In case multiple kits were required to be used, it should be taken into the potential confounding factors for statistical analysis.

In this thesis, another confounding factor was analyzed. The host DNA that is present in the sample may affect sequencing results, especially in case of samples with high host-, low microbial biomass ("high-low samples"). When quality-checking samples, the host DNA abundance should also be taken into consideration, rather than only evaluate the microbial vs. background/reagent ratio as most previously published article has done. This analysis will be further discussed as a major component of the second chapter.

1.9 Scope and objectives of the thesis

In this thesis, the microbiota of the human urinary tract was investigated. Multiple trials were conducted to collect various clinical sample including urine and surgical tissue samples from the urinary system. There are a limited number of studies looking at the urinary microbiota, and at the onset, none have analyzed their sample handling techniques and potential confounding factors. To our knowledge, there is no published data of microbiota of renal tissue, and urine of stone and bladder cancer patients. The general hypothesis is that a dynamic microbiota exists in the urinary tract and these microbial shifts exhibit different phenotype across various urinary conditions. Since the method of studying the urinary microbiota has yet to be standardized, this was the first step to develop and optimize algorithms for urinary tissue microbiota and urine microbiota. This methodology was used in Chapter 2, 3, 4 to handle "high-low" samples. Samples with low microbial biomass are recently reported to be vulnerable to reagent contaminations, and in Chapter 2 this was further examined to show that the presence of host DNA could exert negative effects on sequencing results on multiple levels. A refined protocol was developed and utilized in renal microbiota profiling in a later section of Chapter 2. In chapter 3, we focused on characterizing the urinary microbiota of a healthy individual and urinary stone patients and investigated the stability of urinary microbiota by analyzing several temporally discrete samples from the same donor. In Chapter 4, we observed fluctuations of urinary microbiota during lithotripsy surgeries in urinary stone patients to examine the possible existence of bacteria in metabolic urinary stones.

Overall, these studies constitute a comprehensive picture of the urinary microbiota in various conditions. The following studies have generated multiple methods and techniques for microbiota research in sites without a rich microbial community. The

importance of understanding the potential pitfalls of microbiota research is underscored, and this thesis will provide further insight into future techniques of carrying out such studies.

1.10 References

Abol-Enein, H. (2008). Infection: is it a cause of bladder cancer? *Scandinavian Journal of Urology and Nephrology*, 42(sup218), 79-84.

Abulencia, C. B., Wyborski, D. L., Garcia, J. A., Podar, M., Chen, W., Chang, S. H., ... & Keller, M. (2006). Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Applied and environmental microbiology*, 72(5), 3291-3301.

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., & Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *Journal of bacteriology*, 186(9), 2629-2635.

Agace, W. W., Hedges, S. R., Ceska, M., & Svanborg, C. (1993). Interleukin-8 and the neutrophil response to mucosal gram-negative infection. *Journal of Clinical Investigation*, *92*(2), 780.

Aitchison, J. (1986). The statistical analysis of compositional data.

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell*, 124(4), 783-801.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), 3389-3402.

Amann, R. I., Ludwig, W. & Schleifer, K.-H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, 59(1), 143–169.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology*, 26(1), 32–46.

Baerheim, A., Larsen, E. & Digranes, A. (1994). Vaginal application of lactobacilli in the prophylaxis of recurrent lower urinary tract infection in women. *Scandinavian journal of primary health care*, 12(4), 239–243.

Barr-Beare, E., Saxena, V., Hilt, E. E., Thomas-White, K., Schober, M., Li, B., ... & Schwaderer, A. L. (2015). The interaction between Enterobacteriaceae and calcium oxalate deposits. *PloS one*, 10(10), e0139575.

Barrons, R. & Tassone, D. (2008). Use of Lactobacillus probiotics for bacterial genitourinary infections in women: a review. *Clinical therapeutics*, 30(3), 453–468.

Beerepoot, M. & Geerlings, S. (2016). Non-antibiotic prophylaxis for urinary tract infections. *Pathogens*, 5(2), 36.

Beerepoot, M., Ter Riet, G., Nys, S. (2012). Lactobacilli versus antibiotics to prevent urinary tract infections. A randomized double-blind non-inferiority trial in postmenopausal women. *Archives of internal medicine*, 172(9), 704–712.

Bergeron, C. R., Prussing, C., Boerlin, P., Daignault, D., Dutil, L., Reid-Smith, R. J., Zhanel, G. G., et al. (2012). Chicken as reservoir for extraintestinal pathogenic Escherichia coli in humans, Canada. *Emerging infectious diseases*, 18(3), 415.

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R. & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in ecology & evolution*, 27(4), 233–243.

Bishop, B. L., Duncan, M. J., Song, J., Li, G., Zaas, D. & Abraham, S. N. (2007). Cyclic AMP-regulated exocytosis of Escherichia coli from infected bladder epithelial cells. *Nature medicine*, 13(5), 625–630.

Brumfitt, W., Gargan, R. & Hamilton-Miller, J. (1987). Periurethral enterobacterial carriage preceding urinary infection. *The Lancet*, 329(8537), 824–826.

Burton, J. P. & Reid, G. (2002). Evaluation of the bacterial vaginal flora of 20 postmenopausal women by direct (Nugent score) and molecular (polymerase chain reaction and denaturing gradient gel electrophoresis) techniques. *The Journal of infectious diseases*, 186(12), 1770–1780.

Cadieux, P., Burton, J., Devillard, E. & Reid, G. (2009). Lactobacillus by-products inhibit the growth and virulence of uropathogenic Escherichia coli. *J Physiol Pharmacol*, 60(Suppl 6), 13–18.

Cai, Y. & Sun, Y. (2011). ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. *Nucleic acids research*, 39(14), e95–e95.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., ... & Huttley, G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5), 335-336.

Carter, S. C., Levy, M. E. & Zeitlin, S. I. (2012). Recurrent Urinary Tract Infections in Men. Urological Men's Health (pp. 39–53).

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E. A., Grange, T. & Geigl, E. M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PloS One*, 5(9). e13042

Chan, C. Y., John, A. L. S. & Abraham, S. N. (2013). Mast cell interleukin-10 drives localized tolerance in chronic bladder infection. *Immunity*, 38(2), 349–359.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of statistics*, 265–270.

Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T. J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics*, 62(2), 361–371.

Chao, A. & Lee, S.-M. (1992). Estimating the number of classes via sample coverage. *Journal of the American statistical Association*, 87(417), 210–217.

Cheng, L., Walker, A. W. & Corander, J. (2012). Bayesian estimation of bacterial community composition from 454 sequencing data. *Nucleic acids research*, 40(12), 5240–5249.

CLARKE, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, 18(1), 117–143.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., ... & Tiedje, J. M. (2008). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic acids research*, 37(suppl_1), D141-D145.

Collins, F. S., Morgan, M. & Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science*, 300(5617), 286–290.

Conway, P. H., Cnaan, A., Zaoutis, T., Henry, B. V., Grundmeier, R. W. & Keren, R. (2007). Recurrent urinary tract infections in children: risk factors and association with prophylactic antimicrobials. *JAMA*, 298(2), 179–186.

Cosseau, C., Devine, D. A., Dullaghan, E., Gardy, J. L., Chikatamarla, A., Gellatly, S., ... & Hancock, R. E. (2008). The commensal Streptococcus salivarius K12 downregulates the innate immune responses of human epithelial cells and promotes host-microbe homeostasis. *Infection and immunity*, 76(9), 4163-4175.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, 72(7), 5069–5072.

Dickson, R. P., Erb-Downward, J. R., Freeman, C. M., McCloskey, L., Falkowski, N. R., Huffnagle, G. B. & Curtis, J. L. (2017). Bacterial topography of the healthy human lower respiratory tract. *MBio*, 8(1), e02287–16.

Dickson, R. P., Erb-Downward, J. R. & Huffnagle, G. B. (2015). Homeostasis and its disruption in the lung microbiome. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 309(10), L1047–L1055.

Dong, Q., Nelson, D. E., Toh, E., Diao, L., Gao, X., Fortenberry, J. D. & Van Der Pol, B. (2011). The microbial communities in male first catch urine are highly similar to those in paired urethral swab specimens. *PloS one*, 6(5), e19709.

Dwyer, P. L. & O'reilly, M. (2002). Recurrent urinary tract infection in the female. *Current Opinion in Obstetrics and Gynecology*, 14(5), 537–543.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16), 2194–2200.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3), 279–300.

Engen, S., Grøtan, V. & Sæther, B.-E. (2011). Estimating similarity of communities: a parametric approach to spatio-temporal analysis of species diversity. *Ecography*, 34(2), 220–231.

Eto, D. S., Sundsbak, J. L. & Mulvey, M. A. (2006). Actin-gated intracellular growth and resurgence of uropathogenic Escherichia coli. *Cellular microbiology*, 8(4), 704–717.

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J. & Huttenhower, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS computational biology*, 8(7), e1002606.

Felis, G. E. & Dellaglio, F. (2007). Taxonomy of lactobacilli and bifidobacteria. *Current issues in intestinal microbiology*, 8(2), 44.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. science, 496–512.

Fouts, D. E., Pieper, R., Szpakowski, S., Pohl, H., Knoblach, S., Suh, M.-J., Huang, S.-T., et al. (2012). Integrated next-generation sequencing of 16S rDNA and metaproteomics differentiate the healthy urine microbiome from asymptomatic bacteriuria in neuropathic bladder associated with spinal cord injury. *Journal of translational medicine*, 10(1), 174.

Foxman, B. (1990). Recurring urinary tract infection: incidence and risk factors. *American journal of public health*, 80(3), 331–333.

Foxman, B., Barlow, R., D'Arcy, H., Gillespie, B. & Sobel, J. D. (2000). Urinary tract infection: self-reported incidence and associated costs. *Annals of epidemiology*, 10(8), 509–515.

Foxman, B., Gillespie, B., Koopman, J., Zhang, L., Palin, K., Tallman, P., Marsh, J. V., et al. (2000). Risk factors for second urinary tract infection among college women. *American journal of epidemiology*, 151(12), 1194–1205.

Fricke, W., Maddox, C., Song, Y. & Bromberg, J. (2014). Human microbiome characterization in the course of renal transplantation. *American Journal of Transplantation*, 14(2), 416–427.

Friedman, J. & Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS computational biology*, 8(9), e1002687.

Garrett, W. S., Gallini, C. A., Yatsunenko, T., Michaud, M., DuBois, A., Delaney, M. L., Punit, S., et al. (2010). Enterobacteriaceae act in concert with the gut microbiome to induce spontaneous and maternally transmitted colitis. *Cell host & microbe*, 8(3), 292–300.

Gilbert, J. A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C. T., Brown, C. T., Desai, N., et al. (2010). Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. *Standards in genomic sciences*, 3(3), 243.

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., et al. (2006). *Metagenomic analysis of the human distal gut microbiome. science*, 312(5778), 1355–1359.

Ginzinger, D. G., Godfrey, T. E., Nigro, J., Moore, D. H., Suzuki, S., Pallavicini, M. G., Gray, J. W., et al. (2000). *Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. Cancer research*, 60(19), 5405–5409.

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B. & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiome in low bacterial biomass samples. *Gut pathogens*, 8(1), 24.

Gottschick, C., Deng, Z.-L., Vital, M., Masur, C., Abels, C., Pieper, D. H. & Wagner-Döbler, I. (2017). The urinary microbiome of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. *Microbiome*, 5(1), 99.

Grabe, M., Bjerklund-Johansen, T., Botto, H. & others. (2009). Guidelines on urological infections. Retrieved from *http://www.uroweb.org/professional-resources/guidelines/online.*

Groah, S. L., Pérez-Losada, M., Caldovic, L., Ljungberg, I. H., Sprague, B. M., Castro-Nallar, E., ... & Pohl, H. G. (2016). Redefining healthy urine: a cross-sectional exploratory metagenomic study of people with and without bladder dysfunction. *The Journal of urology*, 196(2), 579-587.

Hagberg, L., Bruce, A. W., Reid, G., Svanborg Eden, C., Lincoln, K., & Lidin-Janson, G. (1989). Colonization of the urinary tract with live bacteria from the normal fecal and urethral flora in patients with recurrent symptomatic urinary tract infections. *Host-parasite interactions in urinary tract infections*, 194-197.

Hao, X., Jiang, R. & Chen, T. (2011). Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. *Bioinformatics*, 27(5), 611–618.

Hårdeman, F. & Sjöling, S. (2006). Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS microbiology ecology*, 59(2), 524–534.

Hartge, P., Harvey, E. B., Linehan, W. M., Silverman, D. T., Sullivan, J. W., Hoover, R. N. & Fraumeni, J. F. (1990). Unexplained excess risk of bladder cancer in men. *JNCI: Journal of the National Cancer Institute*, 82(20), 1636–1640.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331(6016), 463–467.

Hill, J. E., Fernando, W. U., Zello, G. A., Tyler, R. T., Dahl, W. J. & Van Kessel, A. G. (2010). Improvement of the representation of bifidobacteria in fecal microbiome metagenomic libraries by application of the cpn60 universal primer cocktail. *Applied and environmental microbiology*, 76(13), 4550–4552.

Hill, J. E., Penny, S. L., Crowell, K. G., Goh, S. H. & Hemmingsen, S. M. (2004). cpnDB: a chaperonin sequence database. *Genome Research*, 14(8), 1669–1675.

Hilt, E. E., McKinley, K., Pearce, M. M., Rosenfeld, A. B., Zilliox, M. J., Mueller, E. R., ... & Schreckenberger, P. C (2014). Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *Journal of clinical microbiology*, 52(3), 871–876.

Horwitz, D., McCue, T., Mapes, A. C., Ajami, N. J., Petrosino, J. F., Ramig, R. F., & Trautner, B. W. (2015). Decreased microbiome diversity associated with urinary tract infection in a trial of bacterial interference. *Journal of Infection*, 71(3), 358–367.

Hummelen, R., Macklaim, J. M., Bisanz, J. E., Hammond, J.-A., McMillan, A., Vongsa, R., Koenig, D., et al. (2011). Vaginal microbiome and epithelial gene array in post-menopausal women with moderate to severe dryness. *PLoS One*, 6(11), e26602.

Hunstad, D. A. & Justice, S. S. (2010). Intracellular lifestyles and immune evasion strategies of uropathogenic Escherichia coli. *Annual review of microbiology*, 64, 203–21.

Huse, S. M., Welch, D. M., Morrison, H. G. & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental microbiology*, 12(7), 1889–1898.

Jack D. Sobel, D. K. (2015). Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, eighth Edition. In Bennett, John E.; Dolin, Raphael; Blaser, Martin J. (Ed.), (pp. 886–913).

Kantor, A. F., Hartge, P., Hoover, R. N., Narayana, A. S., Sullivan, J. & Fraumeni, J. F. (1984). Urinary tract infection and risk of bladder cancer. *American journal of epidemiology*, 119(4), 510–515.

Karstens, L., Asquith, M., Davin, S., Stauffer, P., Fair, D., Gregory, W. T., Rosenbaum, J. T., et al. (2016). Does the urinary microbiome play a role in urgency urinary incontinence and its severity? *Frontiers in cellular and infection microbiology*, 6.

Kennedy, K., Hall, M. W., Lynch, M. D., Moreno-Hagelsieb, G. & Neufeld, J. D. (2014). Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Applied and environmental microbiology*, 80(18), 5717–5722.

Kim, D., Hofstaedter, C. E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., ... & Conrad, M. (2017). Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*, 5(1), 52.

Kim, M., Morrison, M. & Yu, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of microbiological methods*, 84(1), 81–7.

Kinane, D., Blackwell, C. C., Brettle, R., Weir, D., Winstanley, F. & Elton, R. (1982). ABO blood group, secretor state, and susceptibility to recurrent urinary tract infection in women. *Br Med J (Clin Res Ed)*, 285(6334), 7–9.

Kjaer, S. K., Knudsen, J. B., Sørensen, B. L. & Møller Jensen, O. (1989). The Copenhagen case-control study of bladder cancer. V. Review of the role of urinary-tract infection. *Acta oncologica (Stockholm, Sweden)*, 28(5), 631–6.

Klumpp, D. J., Weiser, A. C., Sengupta, S., Forrestal, S. G., Batler, R. A. & Schaeffer, A. J. (2001). Uropathogenic Escherichia coliPotentiates Type 1 Pilus-Induced Apoptosis by Suppressing NF-κB. *Infection and immunity*, 69(11), 6689–6695.

Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental microbiology*, 12(1), 118–123.

Kverka, M. & Tlaskalová-Hogenová, H. (2017). Intestinal Microbiome: Facts and Fiction. *Digestive Diseases*, 35(1-2), 139–147.

Lane, M. & Mobley, H. (2007). Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic Escherichia coli (UPEC) in the mammalian kidney. *Kidney international*, 72(1), 19–25.

Lewis, D. A., Brown, R., Williams, J., White, P., Jacobson, S. K., Marchesi, J. R. & Drake, M. J. (2013). The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. *Frontiers in cellular and infection microbiology*, 3.

Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature*, 444(7122), 1022–1023.

Li, K., Bihan, M., Yooseph, S. & Methe, B. A. (2012). Analyses of the microbial diversity across the human microbiome. *PloS one*, 7(6), e32118.

Li, W. & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659.

Ling, Z., Kong, J., Jia, P., Wei, C., Wang, Y., Pan, Z., Huang, W., et al. (2010). Analysis of oral microbiome in children with dental caries by PCR-DGGE and barcoded pyrosequencing. *Microbial ecology*, 60(3), 677–690.

Links, M. G., Dumonceaux, T. J., Hemmingsen, S. M. & Hill, J. E. (2012). The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. *PloS one*, 7(11), e49755.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., et al. (2012). Comparison of next-generation sequencing systems. *BioMed Research International*, 2012.

Liu, Z., DeSantis, T. Z., Andersen, G. L. & Knight, R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic acids research*, 36(18), e120–e120.

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology*, 30(5), 434–439.

Lopetuso, L. R., Scaldaferri, F., Franceschi, F. & Gasbarrini, A. (2014). The gastrointestinal microbiome-functional interference between stomach and intestine. *Best practice & research Clinical gastroenterology*, 28(6), 995–1002.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *The ISME journal*, 5(2), 169.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1), 209–220.

Matsen IV, F. A. & Evans, S. N. (2013). Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. *PLoS One*, 8(3), e56859.

McMillan, A., Macklaim, J. M., Burton, J. P. & Reid, G. (2012). Adhesion of Lactobacillus iners AB-1 to Human Fibronectin A Key Mediator for Persistence in the Vagina? *Reproductive Sciences*, 1933719112466306.

McMurdie, P. J. & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, 8(4), e61217.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome research*, 15(12), 1767–1776.

Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W. & Banfield, J. F. (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome biology*, 12(5), R44.

Vermeulen, S. H., Hanum, N., Grotenhuis, A. J., Castano-Vinyals, G., Van Der Heijden, A. G., Aben, K. K., ... & Kiemeney, L. A. (2015). Recurrent urinary tract infection and risk of bladder cancer in the Nijmegen bladder cancer study. *British journal of cancer*, 112(3):594.

Mizrahi-Man, O., Davenport, E. R. & Gilad, Y. (2013). Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PloS one*, 8(1), e53608.

Modena, B. D., Milam, R., Harrison, F., Cheeseman, J. A., Abecassis, M. M., Friedewald, J. J., ... & Salomon, D. R. (2017). Changes in urinary microbiome populations correlate in kidney transplants with interstitial fibrosis and tubular atrophy documented in early surveillance biopsies. American Journal of Transplantation, 17(3), 712–723.

Nelson, D. E., Van Der Pol, B., Dong, Q., Revanna, K. V., Fan, B., Easwaran, S., ... & Fortenberry, J. D. (2010). Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. *PloS one*, 5(11), e14116.

Nelson, D. E., Dong, Q., Van Der Pol, B., Toh, E., Fan, B., Katz, B. P., ... & Fortenberry, J. D. (2012). Bacterial communities of the coronal sulcus and distal urethra of adolescent males. *PloS one*, 7(5), e36298.

Nielubowicz, G. R. & Mobley, H. L. (2010). Host-pathogen interactions in urinary tract infection. *Nature Reviews Urology*, 7(8), 430–441.

Nienhouse, V., Gao, X., Dong, Q., Nelson, D. E., Toh, E., McKinley, K., ... & Brubaker, L. (2014). Interplay between bladder microbiome and urinary antimicrobial peptides: mechanisms for human urinary tract infection risk and symptom severity. *PLoS One*, 9(12), e114185.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science*, 276(5313), 734–740.

Pace, N. R. (2009). Mapping the tree of life: progress and prospects. *Microbiology and Molecular Biology Reviews*, 73(4), 565–576.

Pathak, G., Ehrenreich, A., Losi, A., Streit, W. & Gärtner, W. (2009). Novel blue light-sensitive proteins from a metagenomic approach. *Environmental microbiology*, 11(9), 2388–2399.

Pearce, M. M., Hilt, E. E., Rosenfeld, A. B., Zilliox, M. J., Thomas-White, K., Fok, C., Kliethermes, S., et al. (2014). The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *MBio*, 5(4), e01283–14.

Pelucchi, C., Talamini, R., Negri, E., Franceschi, S. & La Vecchia, C. (2006). Genital and urinary tract diseases and prostate cancer risk. *European journal of cancer prevention*, 15(3), 254–7.

Pfefferle, P. I. & Renz, H. (2014). The mucosal microbiome in shaping health and disease. *F1000prime reports*. 6

Pisani, P., Parkin, D. M., Muñoz, N., & Ferlay, J. (1997). Cancer and infection: estimates of the attributable fraction in 1990. *Cancer Epidemiology and Prevention Biomarkers*, 6(6), 387-400.

Poretsky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D. & Konstantinidis, K. T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One*, 9(4), e93827.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research*, 35(21), 7188–7196.

Purohit, M. K., & Singh, S. P. (2009). Assessment of various methods for extraction of metagenomic DNA from saline habitats of coastal Gujarat (India) to explore molecular diversity. *Letters in applied microbiology*, 49(3), 338-344.

Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, 12(1), 38.

Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS microbiology ecology*, 62(2), 142–160.

Rani, A., Ranjan, R., McGee, H. S., Andropolis, K. E., Panchal, D. V., Hajjiri, Z., Brennan, D. C., et al. (2017). Urinary microbiome of kidney transplant patients reveals dysbiosis with potential for antibiotic resistance. *Translational Research*, 181, 59–70.

Reeder, J. & Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nature methods*, 7(9), 668–669.

Reid, G., Bruce, A. & Taylor, M. (1992). Influence of three-day antimicrobial therapy and lactobacillus vaginal suppositories on recurrence of urinary tract infections. *Clinical therapeutics*, 14(1), 11–16.

Reid, G., Bruce, A. W., McGroarty, J. A., Cheng, K. J. & Costerton, J. W. (1990). Is there a role for lactobacilli in prevention of urogenital and intestinal infections? *Clinical microbiology reviews*, 3(4), 335–44.

Reid, G., Younes, J. A., Van der Mei, H. C., Gloor, G. B., Knight, R. & Busscher, H. J. (2011). Microbiome restoration: natural and supplemented recovery of human microbial communities. *Nature Reviews Microbiology*, 9(1), 27–38.

Rhee, J.-K., Ahn, D.-G., Kim, Y.-G. & Oh, J.-W. (2005). New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Applied and environmental microbiology*, 71(2), 817–825.

La Rosa, P. S., Brooks, J. P., Deych, E., Boone, E. L., Edwards, D. J., Wang, Q., ... & Shannon, W. D. (2012). Hypothesis testing and power calculations for taxonomic-based human microbiome data. *PloS one*, 7(12), e52078.

Rosen, D. A., Pinkner, J. S., Jones, J. M., Walker, J. N., Clegg, S. & Hultgren, S. J. (2008). Utilization of an intracellular bacterial community pathway in Klebsiella pneumoniae urinary tract infection and the effects of FimK on type 1 pilus expression. *Infection and immunity*, 76(7), 3337–3345.

Round, J. L. & Mazmanian, S. K. (2009). The gut microbiome shapes intestinal immune responses during health and disease. *Nature Reviews Immunology*, 9(5), 313–323.

Russo, T. A. & Johnson, J. R. (2003). Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. *Microbes and Infection*, 5(5), 449–456.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology*, 12(1), 87.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., ... & Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, 265(5596), 687-695.

SanMiguel, A. & Grice, E. A. (2015). Interactions between host factors and the skin microbiome. *Cellular and molecular life sciences*, 72(8), 1499–1515.

Schellenberg, J., Links, M. G., Hill, J. E., Dumonceaux, T. J., Peters, G. A., Tyler, S., Ball, T. B., et al. (2009). Pyrosequencing of the chaperonin-60 universal target as a tool

for determining microbial community composition. *Applied and environmental microbiology*, 75(9), 2889–2898.

Schilling, J. D., Lorenz, R. G. & Hultgren, S. J. (2002). Effect of trimethoprim-sulfamethoxazole on recurrent bacteriuria and bacterial persistence in mice infected with uropathogenic Escherichia coli. *Infection and immunity*, 70(12), 7042–7049.

Schilling, J. D., Martin, S. M., Hung, C. S., Lorenz, R. G. & Hultgren, S. J. (2003). Toll-like receptor 4 on stromal and hematopoietic cells mediates innate resistance to uropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences*, 100(7), 4203–4208.

Schloss, P. D., Gevers, D. & Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS one*, 6(12), e27310.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23), 7537–7541.

Schmiemann, G., Kniehl, E., Gebhardt, K., Matejczyk, M. M. & Hummers-Pradier, E. (2010). The diagnosis of urinary tract infection: a systematic review. *Deutsches Ärzteblatt International*, 107(21), 361.

Scholes, D., Hooton, T. M., Roberts, P. L., Stapleton, A. E., Gupta, K. & Stamm, W. E. (2000). Risk factors for recurrent urinary tract infection in young women. *The Journal of infectious diseases*, 182(4), 1177–1182.

Shah, N., Tang, H., Doak, T. G. & Ye, Y. (2011). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Biocomputing* 2011 (pp. 165–176).

Shen, H., Rogelj, S. & Kieft, T. L. (2006). Sensitive, real-time PCR detects low-levels of contamination by Legionella pneumophila in commercial reagents. *Molecular and cellular probes*, 20(3), 147–153.

Shoskes, D. A., Altemus, J., Polackwich, A. S., Tucky, B., Wang, H. & Eng, C. (2016). The urinary microbiome differs significantly between patients with chronic prostatitis/chronic pelvic pain syndrome and controls as well as between patients with different clinical phenotypes. *Urology*, 92, 26–32.

Shrestha, E., White, J. R., Yu, S.-H., Kulac, I., Ertunc, O., De Marzo, A. M., Yegnasubramanian, S., et al. (2017). Profiling the Urinary Microbiome in Men with Positive versus Negative Biopsies for Prostate Cancer. *The Journal of Urology*.

Siddiqui, H., Lagesen, K., Nederbragt, A. J., Jeansson, S. L. & Jakobsen, K. S. (2012). Alterations of microbiome in urine from women with interstitial cystitis. *BMC microbiology*, 12(1), 205.

Siddiqui, H., Nederbragt, A. J., Lagesen, K., Jeansson, S. L. & Jakobsen, K. S. (2011). Assessing diversity of the female urine microbiome by high throughput sequencing of 16S rDNA amplicons. *BMC microbiology*, 11(1), 244.

Soergel, D. A. W., Dey, N., Knight, R. & Brenner, S. E. (2012). Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *The ISME journal*, 6(7), 1440–4.

Stackebrandt, E. & Goebel, B. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, 44(4), 846–849.

Stamey, T., Timothy, M., Millar, M. & Mihara, G. (1971). Recurrent urinary infections in adult women. The role of introital enterobacteria. *California medicine*, 115(1), 1.

Stapleton, A. E., Au-Yeung, M., Hooton, T. M., Fredricks, D. N., Roberts, P. L., Czaja, C. A., Yarova-Yarovaya, Y., et al. (2011). Randomized, placebo-controlled phase 2 trial

of a Lactobacillus crispatus probiotic given intravaginally for prevention of recurrent urinary tract infection. *Clinical infectious diseases*, 52(10), 1212–1217.

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of bacteriology*, 178(3), 591–599.

Szabados, F., Kleine, B., Anders, A., Kaase, M., Sak\inç, T., Schmitz, I. & Gatermann, S. (2008). Staphylococcus saprophyticus ATCC 15305 is internalized into human urinary bladder carcinoma cell line 5637. *FEMS microbiology letters*, 285(2), 163–169.

Thomas-White, K. J., Hilt, E. E., Fok, C., Pearce, M. M., Mueller, E. R., Kliethermes, S., Jacobs, K., et al. (2016). Incontinence medication response relates to the female urinary microbiome. *International urogynecology journal*, 27(5), 723–733.

Thumbikat, P., Waltenbaugh, C., Schaeffer, A. J. & Klumpp, D. J. (2006). Antigen-specific responses accelerate bacterial clearance in the bladder. *The Journal of Immunology*, 176(5), 3080–3086.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C., Knight, R. & Gordon, J. I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature*, 449(7164), 804.

Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature biotechnology*, 23(1), 88.

Virchow, R. & Chance, F. (1860). *Cellular Pathology as based upon physiological and pathological histology. Twenty lectures delivered in... 1858. Translated from the second edition of the original by F. Chance. With notes and numerous emendations principally from MS. notes of the author, and illustrated by... engravings on wood.*

Voget, S., Steele, H. & Streit, W. (2006). Characterization of a metagenome-derived halotolerant cellulase. *Journal of biotechnology*, 126(1), 26–36.

Wang, X., Yao, J., Sun, Y. & Mai, V. (2013). M-pick, a modularity-based method for OTU picking of 16S rRNA sequences. *BMC bioinformatics*, 14(1), 43.

Waschkowitz, T., Rockstroh, S. & Daniel, R. (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Applied and environmental microbiology*, 75(8), 2506–2516.

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J. & Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology*, 15(12), 564.

Whiteside, S. A., Razvi, H., Dave, S., Reid, G. & Burton, J. P. (2015). The microbiome of the urinary tract [mdash] a role beyond infection. *Nature Reviews Urology*, 12(2), 81–90.

Willner, D., Low, S., Steen, J. A., George, N., Nimmo, G. R., Schembri, M. A. & Hugenholtz, P. (2014). Single clinical isolates from acute uncomplicated urinary tract infections are representative of dominant in situ populations. *MBio*, 5(2), e01064–13.

Woese, C. R. (1987). Bacterial evolution. *Microbiological reviews*, 51(2), 221.

Wolfe, A. J., & Brubaker, L. (2015). “Sterile urine” and the presence of bacteria. *European urology*, 68(2), 173.

Wolfe, A. J., Toh, E., Shibata, N., Rong, R., Kenton, K., FitzGerald, M., Mueller, E. R., et al. (2012). Evidence of uncultivated bacteria in the adult female bladder. *Journal of clinical microbiology*, 50(4), 1376–1383.

Wright, K. J. & Hultgren, S. J. (2006). Sticky fibers and uropathogenesis: bacterial adhesins in the urinary tract. *Future microbiology*, 1(1), 75–87.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., et al. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature biotechnology*, 29(5), 415–420.

Youssef, N., Sheik, C. S., Krumholz, L. R., Najar, F. Z., Roe, B. A. & Elshahed, M. S. (2009). Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Applied and environmental microbiology*, 75(16), 5227–5236.

Chapter 2

2 Optimization of methodology of handling samples of high host and low microbial biomass reveals the renal microbiota

2.1 Introduction

Through culture-independent microbial identification protocols, microbiota studies have demonstrated that a rich and diverse microbial community inhabits in the human body, varying in its structure, abundance and function depending on body sites (Costello *et al*., 2009; Ding & Schloss, 2014; Segal & Blaser, 2014). The microbes inhabiting these sites have been found to play critical roles in maintaining the normal function of host metabolism (Bäckhed *et al.*, 2004), hormone regulation (Markle *et al.*, 2013) and even psychological status (Hsiao *et al.*, 2013). Additionally, certain microbiota profiles have also been associated with numerous conditions including diabetes mellitus, inflammatory bowel diseases, and multiple sclerosis. (Berer *et al.*, 2011).

As the exploration of microbiota-associated diseases continues, a growing number of studies have been carried out to investigate the microbiota in sites with comparatively low microbial abundance. These include the brain (Branton *et al.*, 2013), breast (Urbaniak *et al.*, 2014), atherosclerotic plaques (Koren *et al.*, 2011), placenta (Aagaard *et al.*, 2014), blood (Potgieter *et al.*, 2015) and from an important urological perspective, the bladder (Wolfe *et al.*, 2012). Detection of bacterial DNA sequences in these sites questions the concept of sterility, and the role of bacteria within the body. However, before drawing such conclusions, sampling bias, DNA contamination, and accuracy of microbiota data analysis must be addressed.

The first definitive description of the bladder microbiota was only undertaken relatively recently, by the direct collection of urine aspirated supra-pubically in an attempt to discount urogenital contamination of specimens (Wolfe *et al.*, 2012). This discovery challenged the dogma that the urinary tract is sterile. Several other studies also confirmed an existant microbiota in the bladder and associated it with various conditions including

incontinence, interstitial cystitis and neurogenic bladder (Pearce *et al.*, 2014; Siddiqui *et al.*, 2012; Groah *et al.*, 2016). Although the exact association of the urinary microbiota of these conditions remains to be elucidated, it has gradually come to people's awareness that a viable urinary microbiota, although unable to be detect by routine culture methods, exist in the bladder.

However, whether microorganisms exist in the upper urinary tract remains unclear. To date, there's no study looking specifically at urinary microbiota of the upper urinary tract in asymptomatic individuals without infective conditions. However, the ascension of bacteria through the urinary tract beyond the bladder causing infections such as pyelonephritis have also been commonly reported (Andersen & Jackson, 1961), suggesting the possibility of bladder microbiota ascending to the upper urinary tract. On the other hand, asymptomatic infections do occur in the upper urinary tract (Dontas *et al.*, 1981; Kass *et al.*, 1972), which could be originated from the microbiota of the upper urinary tract among asymptomatic individuals.

Applying the novel sequencing-based techniques to the study of urinary microbiota is not without difficulties. One of the technical challenges of studying the microbiota of the upper urinary tract is sample collection. It is difficult to obtain kidney samples for microbiota analysis due to the invasiveness of procedures. This is especially true for obtaining samples from healthy individuals. The use of renal tumour and "healthy" adjacent tissue samples offers the opportunity to examine the microbiota without added morbidity to the patient. Samples of tissue are excised during nephrectomy. This setting also allows for comparison of the microbiota of tumour and healthy renal tissue.

Another technical challenge is the sample handling. Given that renal tissue samples have much lower microbial abundance, sample processing and analysis protocols taken directly from gut microbiota studies have not been properly validated. Problems not considered are relating to factors which are associated with the high proportion of host DNA that is present in the sample must be taken into consideration.

In this study, protocols were developed for analyzing low microbial abundance samples in tissue. These protocols were optimized and compared to standard protocols. The aim

was to determine whether renal tissues had a detectable microbiota and if it was different within tumours compared to healthy tissue sites.

2.2 Methods

2.2.1 Cell lines and bacterial strains

Transitional cell carcinoma-derived human 5637 cells were obtained from the American Type Culture Collection (ATCC® HTB-9™, Manassas, VA, USA). Cells were cultured in RPMI-1640 with 10% fetal bovine serum plus 2 mM glutamine. *E. coli* DH5α cells were maintained on Luria-Bertani (LB) medium containing 1.5% agar (LB agar). When required, a single isolated colony was transferred into LB medium and grown overnight at 37℃ on LB with continuous shaking. *Gardnerella vaginalis* (ATCC® 14018™) was maintained on Columbia blood agar (CBA) containing 5% sheep blood. When required, a single isolated colony was transferred into Brain Heart Infusion (BHI) broth and was grown overnight at 37℃ with continuous shaking.

2.2.2 Sample collection

Biological materials, including paired tumour and control "healthy-adjacent" tissue samples from fifty-six donors, were provided by the Ontario Tumour Bank, which is funded by the Ontario Institute for Cancer Research. Ethical approval for the study was granted by the University of Western Ontario Human Ethics Committee and Lawson Health Research Institute (protocol number HSREB 104309). The samples included tumour tissue and a control specimen consisting of an adjacent non-tumorous tissue, approximately 4 cm away from the tumour. Tumor samples represent various grades and pathologies, as shown in the appendix. Sample size varied from approximately 0.2 to 1 cm^3. These were aseptically transferred into cryogenic vials and stored at -80°C until further analysis.

To collect bacterial samples from the exterior of the tumour tissue, the entire tissue mass was transferred to a 2 ml microcentrifuge tube containing 1 ml DNA-free phosphate-buffered saline (PBS, pH 8.0). The samples were placed in a sonicating water bath

(Branson® Ultrasonic Bath Model 1200) at 60 Hz for 5 minutes then vigorously for 1 minute. The remaining liquid was enriched for any bacteria present on the surface. These samples underwent the standard DNA extraction protocol developed for the other tissue samples.

2.2.3 Tissue surface decontamination and fragmentation

Specimens were aseptically transferred to 20 ml of DNA-free PBS and vortexed for 20 sec. The PBS was removed, and 20 ml of DNA-free 75% ethanol was added for 3 minutes, followed by 20 ml of 5% bleach for 1 min, 75% ethanol for 2 min and finally five wash cycles of DNA-free PBS. The samples were then centrifuged, discarding the supernatant and then homogenised using a DNA-free microtube and pestle (Kimble-Chase Kontes™ Pellet Pestle™, Fisher Scientific). PBS was then added a final concentration of 0.4g tissue/ml. The tissue lysate, together with all the surface control samples, were used for routine culture, EQUC and DNA extraction for 16S rRNA gene sequencing. Five reagent controls were taken during the experiment (including PBS, water, DNA-extraction and PCR amplification reagents). Samples were stored at -80°C until further analysis.

2.2.4 DNA extraction

When required, tissue homogenates were thawed on ice, and the DNA was extracted. Tissue was homogenized using the manufacteurer's protocol specific to the MOBIO PowerSoil® 96 well Kit and Qiagen QIAamp kit. For the refined protocol, ethanol precipitation was applied instead of the conventional filter binding protocol.

2.2.5 gDNA copy number estimation

In pure culture samples, the gDNA copy number was estimated by Qbitz (Qiagen, Mississauga, ON, Canada) and nanodrop (Thermo Scientific). For mixed samples the E.coli DH5a gDNA copy number was estimated by qPCR quantification. To ensure the accuracy of qPCR quantification NEBnext® Microbiome DNA Enrichment Kit (New England Biolabs, Beverly, MA, USA)was used to remove host DNA and enrich E.coli

DNA. The DNA samples were then serially diluted to confirm an optimal Ct value to avoid potential inhibition.of amplification

2.2.6 PCR amplification

Extracted DNA from the clinical samples were PCR-amplified with barcoded primers as described by Gloor *et al*. (2010) shown in Appendix A. The PCR was performed in a 42-µl reaction system, with details listed in Appendix B. Following PCR, the amplicons were quantified with the Qubit 2.0 fluorometer (Invitrogen, CA, USA). Equimolar amounts of each amplicon were aliquoted into a single tube purified with the QIAquick purification kit (Qiagen, Mississauga, ON, Canada). Each product was then pooled and sequenced on the Illumina Mi-Seq platform in a paired-end run using a 600 cycle kit with a paired-end 220-bp run at the London Regional Genomics Center (Robarts Research Institute, London, ON, Canada), following standard operating procedures.

2.2.7 Sequence processing and taxonomic assignment.

Description of the processing algorithm has been previously published (Gloor *et al.*, 2010). The SOP and related software are accessible at http://github.com/ggloor/miseq_bin. The sequences were processed following algorithm as stated in Appendix C. Paired-end reads were overlapped with Pandaseq V2.5 (Masella *et al.*, 2012), allowing no ambiguous positions in the overlapped reads. Reads were then demultiplexed and tagged with the name corresponding to the sample identifier and barcode. After demultiplexing, reads from all samples were merged into one single file and were clustered into individual sequence units (ISUs), which were then further grouped into operational taxonomic units (OTUs) with open reference OTU picking method, using Uclust algorithm of USEARCH version 7, which has a *de novo* chimera filtering step built into it (Edgar *et al.*, 2010; Edgar *et al.*, 2015). The identity threshold was 97%. Singleton OTUs and ones with abundance < 1% were discarded. Taxonomical assignment was performed with mothur 1.3.4 to annotate the OTU sequences against the Silva database (v119).

2.2.8 Data analysis

For data exploration, OTUs with a zero value was replaced by a small numerical value with the count zero multiplicative methods in the zCompositions package (Palarea-Albaladejo *et al.,* 2015). Custom R codes were generated to perform centered log ratio (clr) transform of the zero-replaced data set. Then a singular value decomposition (SVD) was applied to the data. The output was used for PCA plots exploration. 128 Dirichlet Monte-Carlo (DMC) instances from the data with clr-transform were used for quantitative data analysis. ALDEx 2 V.1.6.0 package were used for differential abundance test. Significance was based on the Benjamini-Hochberg corrected P value of the Wilcoxon rank test (significance threshold $P < 0.1$).

2.3 Results

2.3.1 High host cell abundance is associated with poor microbial DNA yield and microbial community representation in column-extracted DNA samples

DNA extraction kits typically use silica gel membranes to bind sample DNA. The MO BIO Powersoil® kit, one of the most widely used DNA extraction kits for microbiota analysis, was tested for the maximal DNA yield of *E. coli* DH5α cells in the presence of human bladder 5637 cells. The Powersoil® kit and subsequent detection of bacterial DNA via qPCR allowed for more than 10^9 copies of *E.coli* gDNA to be detected when 10^{10} copies of *E.coli* gDNA was inputted. However, 10^{10} copies of bladder 5637 cells would only generate around 10^6 copies gDNA in the product. (Figure 2a). A similar phenomenon was observed when Qiagen QIAamp® DNA Stool Mini Kit, another common kit used for bacterial genomic extraction, and where DNA was inputted instead of cells for both extraction kits (Figure 3a, 4a, and 5a). When a mock sample of mixed *E. coli* DH5α cells and bladder 5637 cells went through DNA extraction, the microbial gDNA yield of the mock sample started to drop when bladder 5637 cell number exceeded 10^4, (Figure 2b, 3b, 4b, and 5b). However, when the final kit collection steps involving the silicon-embedded filters were substituted for ethanol precipitation, the “column

saturation effect" of the 5637 cell DNA yield was reduced (Figure 2c, 3c, 4c and 5c), and a higher microbial DNA detection/ recovery was observed from the mixed mock samples (Figure 2d, 3d, 4d and 5d). It appears that column-based DNA extraction methods have poorer microbial DNA yields in the presence of highly abundant host DNA reducing the sensitivity of its detection.

To investigate if the microbial DNA recovery rate is associated with microbial sampling bias and poorer microbiota representation, "mock" or simulated community samples with six bacterial strains were combined with increasing numbers of bladder 5637 cells. These samples were then DNA-extracted with the Powersoil® kit. Extracted DNA was then amplified with primers specific for the V4 region of the 16S rRNA gene and sequenced. Analysis demonstrated that in mock samples starting with equal numbers of each bacterial type, when the host gDNA present exceeds 10^5 copy/sample, the bacterial community representation was severely compromised (Figure 6). In summary, silicon columns-based DNA extraction methods have a limited DNA yield and may compromise microbial DNA yield with abundant host DNA in the sample.

2.3.2 Effect of background DNA on microbial DNA amplification and sequencing analysis

Background DNA from host tissue could be another factor which potentially hampers the PCR amplification and quality of downstream analysis. To look at the impact of this, a series of extractions containing increasing concentrations of bladder 5637 gDNA with a constant amount of *E. coli* DH5α gDNA was created. The bacterial gDNA copy was then estimated with qPCR. The results indicated that the amplification efficiency started to decrease when 10^3 copies of 5637 bladder cell gDNA were present and became severely inhibited above 3.5 x10^5 copies (Figure 7a). Dilution of sample concentration could help mitigate the inhibitory effect, as shown in Figure 7b, where when a mock sample with 10^9 copies of *E. coli* gDNA and 10^5 copies of bladder 5637 gDNA receive a hundred-fold dilution, the Ct difference to the mock sample with same microbial biomass but no bladder gDNA dropped from 3.12 (12.57 vs 9.45) to 0.06 (15.72 to 15.66). Similarly, a 100-fold dilution of a mock sample with 10^7 copies of *E. coli* gDNA and 10^5 copies of bladder 5637 gDNA reduces the Ct difference from 4.46 (20.12 vs 15.66) to 0.64 (23.49

vs 22.85) (Figure 7b). When observing the 16S rRNA gene sequencing results with mock microbial samples combined with 5637 bladder cell gDNA, the microbial composition becomes proportionally distorted with higher the background DNA present. This distortion occurred when DNA abundance was under 10^5 copies/ ml in the template, or when host gDNA concentration was 1000 times higher in copy numbers than microbial gDNA (Figure 7c and 7d). To mitigate inhibitory effects of host DNA during the PCR. We tried to use a commercial host DNA removal kit (NEBnext® Microbiome DNA enrichment Kit) to remove host DNA. However, the host DNA remained, especially in conditions where a high amount of host DNA was present (Figure 8 and 9a).

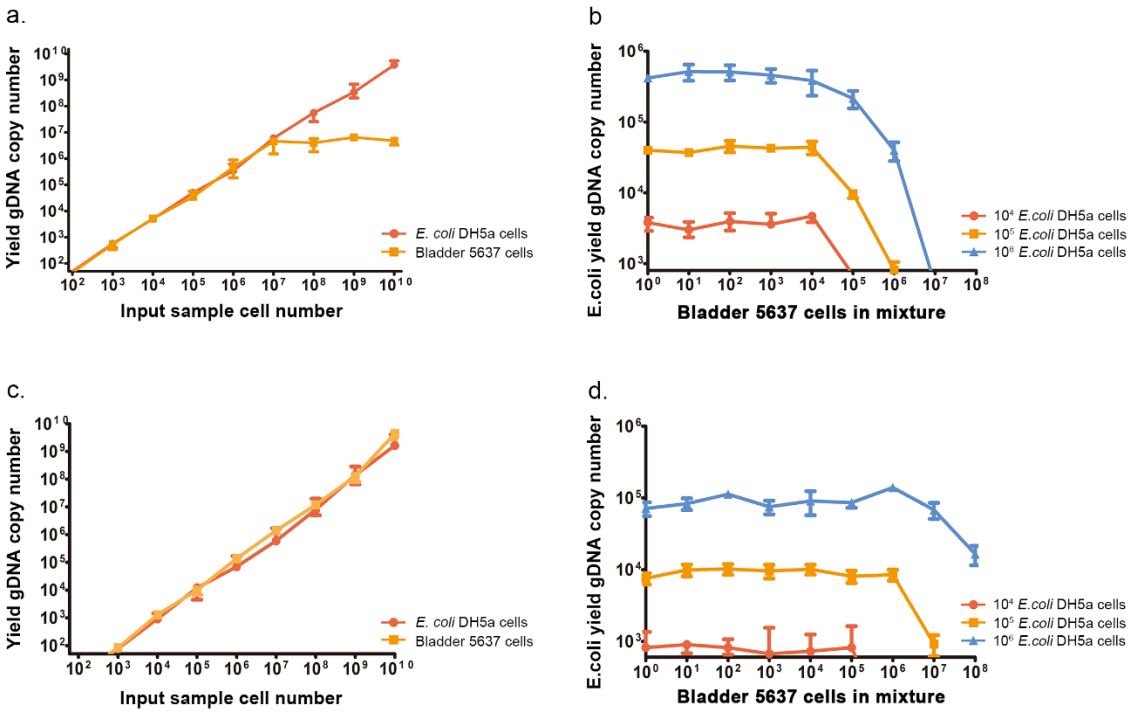

Figure 2 Effect of host cells in the sample on microbial DNA yield. In this plot, *E.coli* DH5a cells or bladder 5637 cells (for a and c) or their mixtures (for b and d) at various amounts received DNA extraction with MO BIO Powersoil® kit standard protocol (for a and b) or adapted protocol, with the last collection steps involving the silicon-embedded filters were substituted by ethanol precipitation (for c and d). X-axis shows sample cell number, Y-axis shows estimated DNA copy number after DNA extraction with qPCR (for *E.coli* DH5a cells) or DNA fluorometer and spectrophotometer (for bladder 5637 cells). (a) DNA yield of *E.coli* DH5a cells (Red) and bladder 5637 cells (Yellow) in

estimated copy numbers, with standard extraction protocol; (b) *E.coli* DH5a DNA yield in estimated copy numbers, with standard DNA extraction protocol, from mixture samples of *E.coli* DH5a cells or bladder 5637 cells; (c) DNA yield of *E.coli* DH5a cells (Red) and bladder 5637 cells (Yellow) in estimated copy numbers, with adapted extraction protocol; (d) *E.coli* DH5a DNA yield in estimated copy numbers, with adapted DNA extraction protocol, from mixture samples of *E.coli* DH5a cells or bladder 5637 cells.

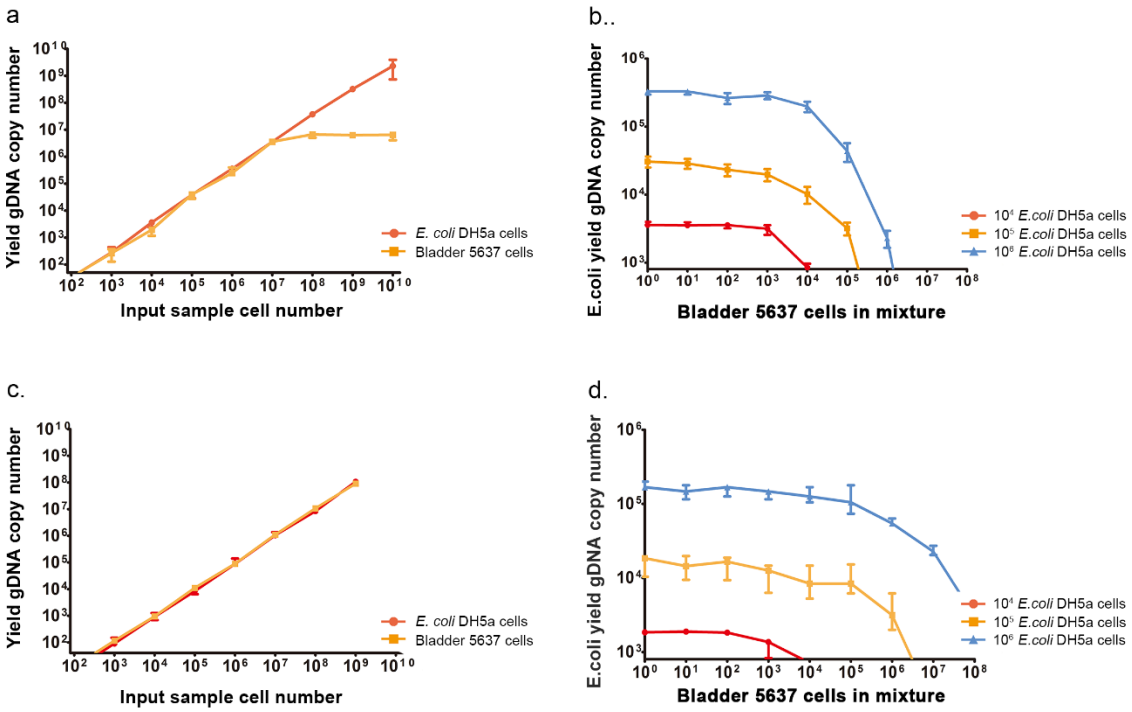

Figure 3 Effect of host cells in the sample on microbial DNA yield. In this plot, pure cultures of *E.coli* DH5a cells or bladder 5637 cells (for a and c) or their mixtures (for b and d) at various amounts received DNA extraction with Qiagen QIAamp® DNA Stool Mini Kit standard protocol (for a and b) or adapted protocol, with the last collection steps involving the silicon-embedded filters were substituted by ethanol precipitation (for c and d). The x-axis shows sample cell number; Y-axis shows estimated DNA copy number after DNA extraction with qPCR (for *E.coli* DH5a cells) or DNA fluorometer and spectrophotometer (for bladder 5637 cells). (a) DNA yield of *E.coli* DH5a cells (Red) and bladder 5637 cells (Yellow) in estimated copy numbers, with standard extraction protocol; (b) *E.coli* DH5a DNA yield in estimated copy numbers, with standard DNA

extraction protocol, from mixture samples of *E.coli* DH5a cells or bladder 5637 cells; (c) DNA yield of *E.coli* DH5a cells (Red) and bladder 5637 cells (Yellow) in estimated copy numbers, with adapted extraction protocol; (d) *E.coli* DH5a DNA yield in estimated copy numbers, with adapted DNA extraction protocol, from mixture samples of *E.coli* DH5a cells or bladder 5637 cells.

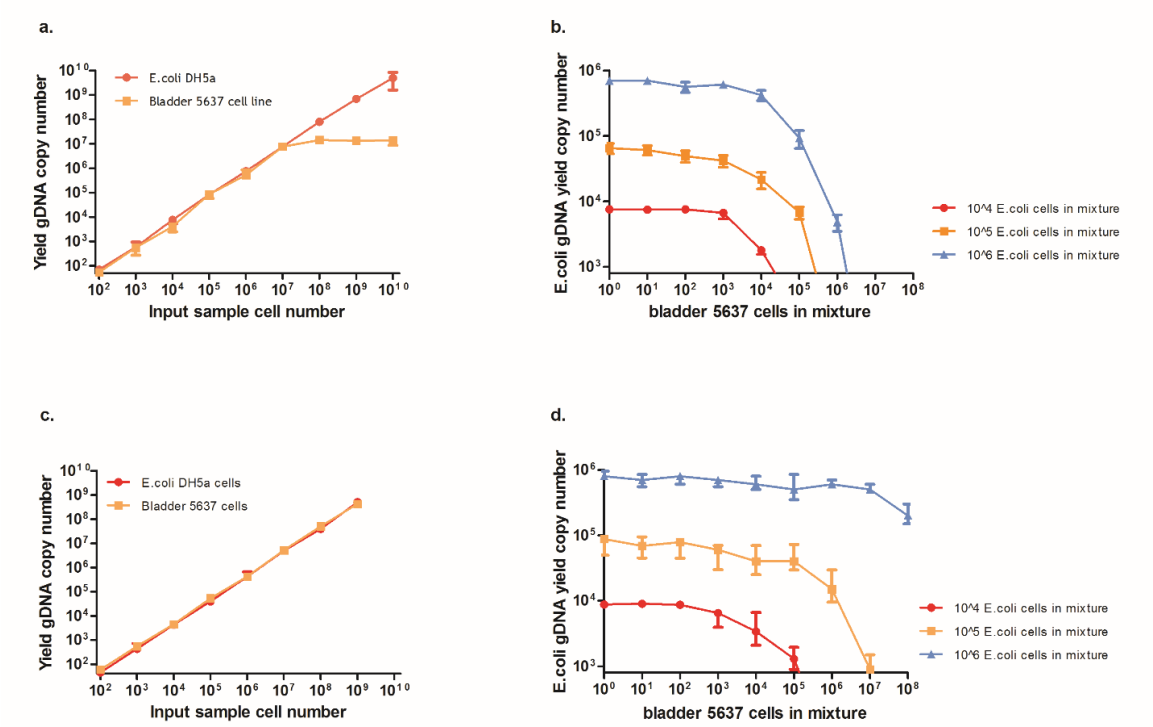

Figure 4 Effect of host DNA in the sample on microbial DNA yield. In this plot, *E.coli* DH5a DNA or bladder 5637 DNA samples (for a and c) or their mixtures (for b and d) at various copies where input into the DNA-binding column of MO BIO Powersoil® kit and went through its standard protocol (for a and b) or adapted protocol, with the last collection steps involving the silicon-embedded filters were substituted by ethanol precipitation (for c&d). X-axis shows sample cell number, Y-axis shows estimated DNA copy number after DNA extraction with qPCR (for *E.coli* DH5a cells) or DNA fluorometer and spectrophotometer (for bladder 5637 cells). (a) DNA yield of *E.coli* DH5a DNA (Red) and bladder 5637 DNA (Yellow) in estimated copy numbers, with standard extraction protocol; (b) *E.coli* DH5a DNA yield in estimated copy numbers, with standard DNA extraction protocol, from mixture samples of *E.coli* DH5a DNA or

bladder 5637 DNA; (c) DNA yield of *E.coli* DH5a DNA (Red) and bladder 5637 DNA (Yellow) in estimated copy numbers, with adapted extraction protocol; (d) *E.coli* DH5a DNA yield in estimated copy numbers, with adapted DNA extraction protocol, from mixture samples of *E.coli* DH5a DNA or bladder 5637 DNA.

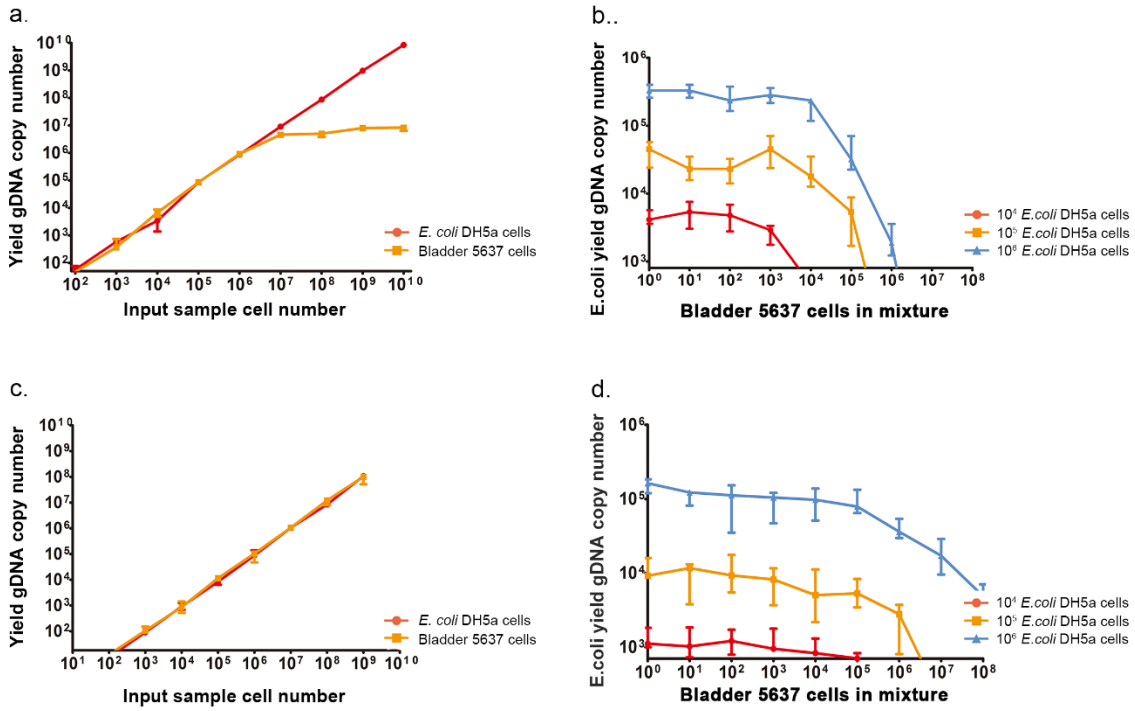

Figure 5 Effect of host DNA in the sample on microbial DNA yield. In this plot, *E.coli* DH5a DNA or bladder 5637 DNA samples (for a and c) or their mixtures (for b and d) at various copies where input into the DNA-binding column of Qiagen QIAamp® DNA Stool Mini Kit and went through its standard protocol (for a and b) or adapted protocol, with the last collection steps involving the silicon-embedded filters substituted by ethanol precipitation (for c&d). The x-axis shows sample cell number; Y-axis shows estimated DNA copy number after DNA extraction with qPCR (for *E.coli* DH5a cells) or DNA fluorometer and spectrophotometer (for bladder 5637 cells). (a) DNA yield of *E.coli* DH5a DNA (Red) and bladder 5637 DNA (Yellow) in estimated copy numbers, with standard extraction protocol; (b) *E.coli* DH5a DNA yield in estimated copy numbers, with standard DNA extraction protocol, from mixture samples of *E.coli* DH5a DNA or bladder 5637 DNA; (c) DNA yield of *E.coli* DH5a DNA (Red) and bladder 5637 DNA

(Yellow) in estimated copy numbers, with adapted extraction protocol; (d) *E.coli* DH5a DNA yield in estimated copy numbers, with adapted DNA extraction protocol, from mixture samples of *E.coli* DH5a DNA or bladder 5637 DNA.

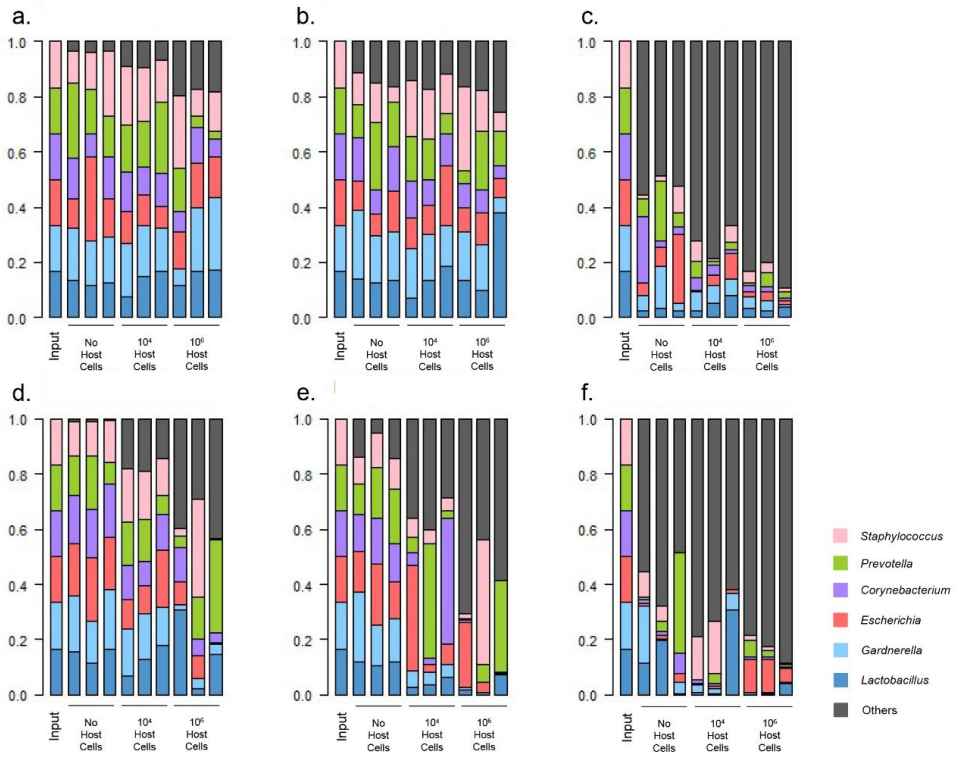

Figure 6 Effect of host DNA on 16s rRNA gene sequencing. Stacked bar plot showed percentage relative abundance of microbial species (n = 6) in even mock samples from 16s rRNA gene sequencing. Each column stands for a mock sample. Mock samples are made with mixtures of six bacterial culture at even amount. Samples are sent in triplicates for sequencing. For each plot, the first column on the left demonstrates the composition of the input microbial mock sample. The next nine columns showed sequencing results of three sets of samples in triplicates, with each set of microbial mock sample spiked with no bladder 5637 cells, 10^3 5637 cells, 10^5 5637 cells respectively. Samples in this plot were DNA-extracted with MO BIO Powersoil® kit standard protocol (for d,e, and f) or adapted protocol, with the final kit collection steps involving the silicon-embedded filters

were substituted by ethanol precipitation. (a) Mock samples with the estimated cell number of 10^9 and DNA-extracted with the adapted protocol. (b) Mock sample with the estimated cell number of 10^7 and DNA-extracted with the adapted protocol. (c) Mock sample with the estimated cell number of 10^5 and DNA-extracted with the adapted protocol. (d) Mock sample with the estimated cell number of 10^9 and DNA-extracted with the standard protocol. (e) Mock sample with the estimated cell number of 10^7 and DNA-extracted with the standard protocol. (f) Mock sample with the estimated cell number of 10^5 and DNA-extracted with the standard protocol.

Another issue associated with high host DNA abundance during PCR is the amplification of non-specific DNA by the 16S rRNA gene-targeted primers. Several primer sets based within conservative regions but designed to amplify adjacent V1-V3, V3-V4, V4 and V8-V9 variable regions have been reported to non-specifically amplify human DNA. In our study, isolated gDNA of renal tissue samples was amplified with V4 primers, agarose gel electrophoresis of the amplification product demonstrated two bands at 385 kb and 302 kb, which confirmed non-specific amplification (Figure 9a and 10). Non-specific amplification products can reach high abundance and potentially consume a significant proportion of a next generation sequencing run (Figure 9b). A commercial microbial DNA enrichment kit tested, was not able to fully mitigate the non-targeted amplification when human DNA was present in abundance (Figure 9c and 9d). In our experience, different primers have different amplification specificity on different samples types (Figure 11). Thus, a pilot study to test the amplification performance of a given primer-sample combination maybe necessary before the main study starts.

2.3.3 Investigating renal tumor tissue microbiota: simple protocol vs. refined protocol

Using the refined protocol for human tissue microbiota investigation, we extracted and analyzed 56 renal cell carcinoma tissue samples and their adjacent pairs in non-tumour tissue. The mass and qPCR-estimated microbial abundance is presented in supplement file. DNA samples were PCR-amplified using the V4 16S rRNA gene region primers and sent for high throughput sequencing. An exploratory biplot (Figure 15a) shows that the “simple” protocol, as typically used for other microbiota studies (*e.g.*, faecal material), utilizing column-based DNA extraction kits resembles that of the reagent control (no significant OTUs between the two group, using ALDEx2), suggesting sparse microbial

existence. However, when the samples were estimated of their microbial and host gDNA abundance with qPCR, DNA fluorometer and spectrophotometer, triaged according to their quality and processed through the refined algorithm,

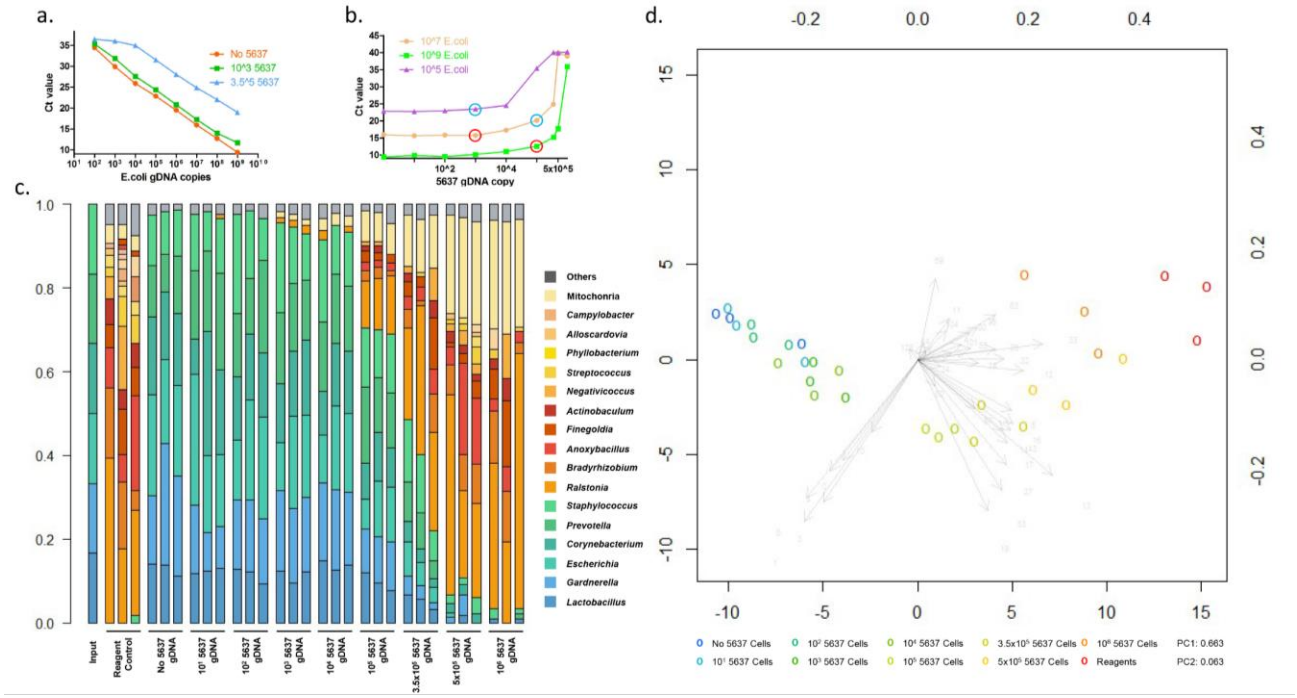

Figure 7. Influence of non-target DNA upon the performance of PCR amplification and subsequent sequencing results. (a) Ct values of quantitative PCR to detect *E.coli* DH5a DNA were shown in the line plot. Mock samples were made of serially diluted *E.coli* DH5a DNA solution spiked with none, 10^3 copy and a $3.5x10^5$ copy of bladder 5637 cell DNA. (b) Ct values of quantitative PCR to detect *E.coli* DH5a DNA were shown in the line plot. Mock samples were made of serially diluted bladder 5637 cell DNA solution spiked with 10^5 copy, 10^7copy and 10^9copy of *E.coli* DH5a DNA. Colored circles show changes of ct value of a 100-fold dilution 10^9 to 10^7 copy of *E.coli* DH5a DNA (red circles) or 10^7 to10^5 copy of *E.coli* DH5a DNA (blue circles). The inhibiting effect is alleviated after dilution. (c) Stacked bar plot showing results of percentage relative abundance of microbial species (n = 6) in even mock samples from 16s rRNA gene sequencing. Each column stands for a mock sample. Mock samples are made with mixtures of six bacterial culture with estimated 10^6 gDNA copy number. Samples are sent in triplicates for sequencing. The left most bars demonstrated the composition of the input microbial mock sample. The next three columns were sequencing results of reagent control. Each set of mock samples were spiked with bladder 5637 cell DNA at various concentration. (d) Compositional biplot of the same set of mock samples as described in (c). Sample sets were colored to make them distinguishable. The arrows indicate the amount and direction of variation of the ratio of each OTU to all others in the data set. The top and right axes indicate the values for variable loadings, and the bottom and left axes indicate the unit sum-of-squares values of the principle components. The proportion of variation explained by components 1 and 2 is also indicated.

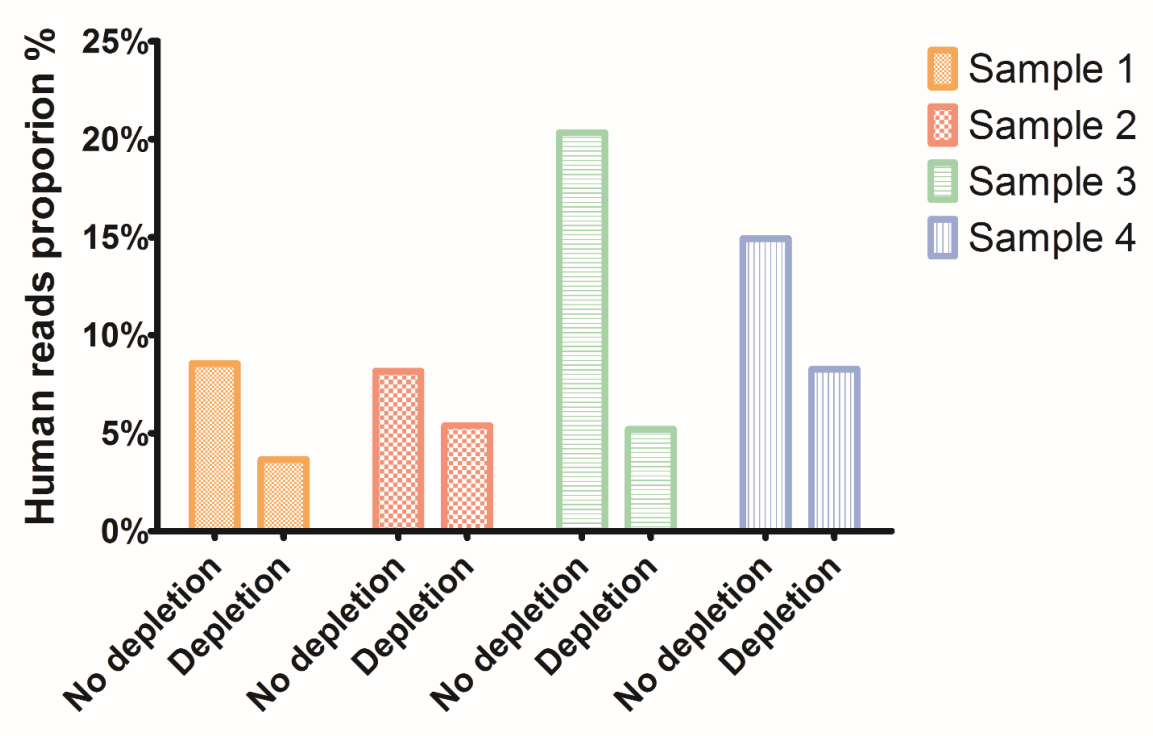

Figure 8. Host depletion effects on human vaginal samples. Bar plot shows the proportion of sequencing reads of human origin. Four human vaginal swab samples went thorough this experiment. Samples labelled as “no depletion” are RNA samples in a transcriptomic experiment, while samples labelled as “depletion” are DNA samples in a metagenomic experiment, treated with NEBnext® Microbiome enrichment Kit. These data showed that microbial enrichment kit was unable to deplete the host component in a sample completely.

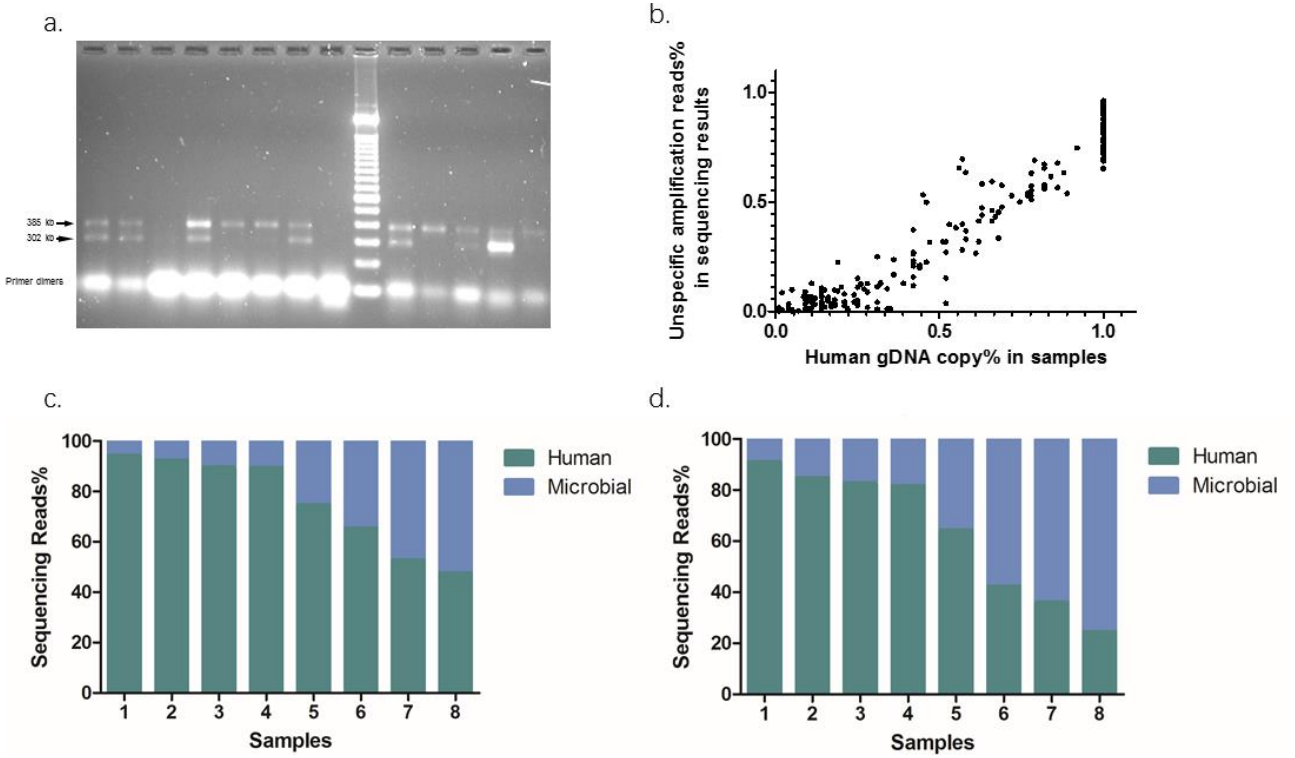

Figure 9. Unspecific amplification of host DNA. (a) Agarose gel electrophoresis of PCR amplification products of a few renal samples. Besides the expected amplification product of V4 primers (385 bp, arrow) there is another band (302 bp, arrowhead), which were confirmed to be unspecific amplifications of human DNA (refer figure s5). (b) Dot plots showing the proportion of unspecific amplification for a batch of renal samples with low microbial abundance. Y-axis = mitochondrial% in sequencing results, x-axis = estimated human DNA% in samples as copy number achieved by qPCR, Nanodrop, and Qubit. For microbial qPCR, if microbial DNA below detection level then considers 100% human DNA. (c) and (d) Stacked bar plot showing proportion of unspecific host DNA amplification from V4 primer before (c) and after (d) treatment from a commercial microbial DNA enrichment kit.

V4 left 5'-3'
Actual primer sequence: GTGCCAGCACGCCGCGGTAA
Mitochondrial sequence: GTGCCAGC CACCGCGGTCA

V4 Right 3'-5'
Actual primer: ATTAGAATACCCTGCTAGGTAGTCC
Mitochondrial: ATTAGA TACCCCACTATGCTT

Figure 10. BLAST alignment of V4 primer and human DNA sequences.

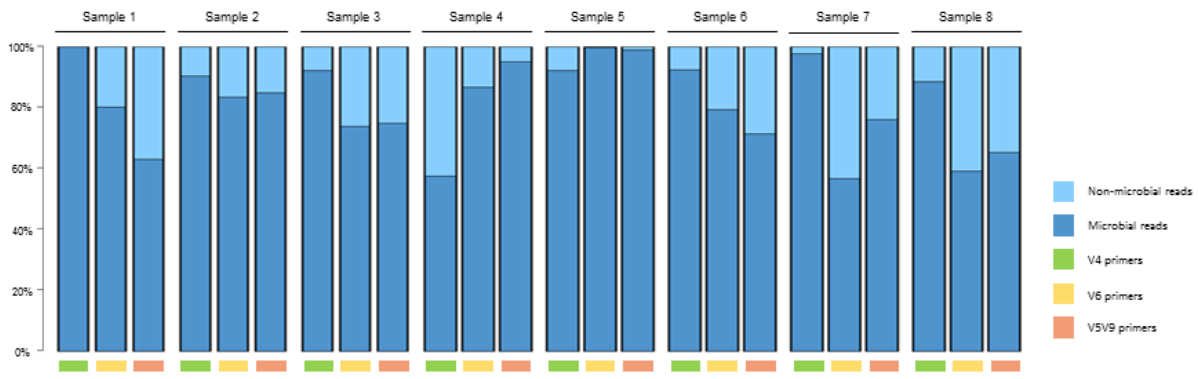

Figure 11. Comparison of the primer combination. A batch of selected renal tissue samples was sent for sequencing with V4, V6 and V5V9 primers on separate sequencing run. Here's a few of them for demonstration. Overall V4 primers had the highest amplification specificity.

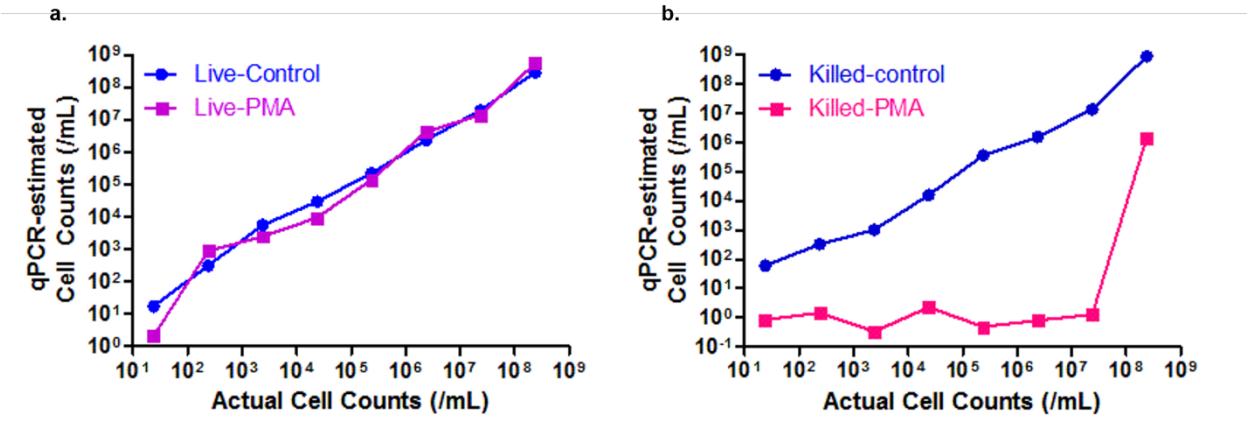

Figure 12. Effects of PMA treatment on live and dead microbial cells and subsequent PCR amplification performance. (a) PMA has little impact on qPCR quantification results of live *E.coli* DH5a cells, while it could deactivate the gDNA of heat-killed *E.coli* DH5a cells. Primer used: V6 (refer to Methods section)

2.3.4 Refined processing for microbiota analysis of High-Low samples.

To better handle the high-low samples, we developed a refined protocol including criteria to exclude the low-quality samples from sequencing (Figure 14b), and processing steps to mitigate the negative effect of host and background DNA, including surface washing, PMA treatment and microbial DNA enrichment (Figure 13). In the quality-checking experiment, the test samples were sent for 16S rRNA gene sequencing, and samples with contaminating bacterial reads lower than 1% is marked as “very good”, 1-10% as “good”, 10-40% as “intermediate”, 40-60% as “poor” and higher than 60% as “very poor” (Figure 14a). The estimated input cell number of *G. vaginalis* of these samples were recorded. After that, all samples sent for microbiota analysis will go through a quality-check algorithm to predict their risk of sequencing results (Figure 14). Samples with microbial abundance in “very poor” level were discarded, and the sequencing results of samples at “intermediate” and “poor” level would go through a self-developed, refined samples processing protocol to reduce the environmental and host DNA (Figure 12). For these samples, the proportion of host reads was taken into consideration, where in our practice

we set microbial reads under refined protocol > 80% as “low host proportion”, 40-80% as “intermediate host proportion” and <40% as “high host proportion”. In mock samples with *G. vaginalis* and bladder 5637 cells, we showed that the refined protocol could mitigate the adverse effect of background noise and non-target host DNA. (Figure 14c).A distinctive microbial consortium, different from the reagent and surface controls could be observed (Figure 15b). A different microbial pattern was identified between the tumour tissue group and normal adjacent tissue group, suggesting a potential different microbial community between these two sites. On samples processed following the kit manufacturer’s protocol, the separation was moderate, and no OTUs were detected as being significant by ALDEx2 test (Figure 15c). On the other hand, samples that went through the refined protocol demonstrated better separation, and 3 OTUs were detected as significantly different between the two sites (Figure 15d). Samples with intermediate to poor quality had better separation in the refined protocol. To confirm the sequencing result we also performed modified expanded quantitative urine culture (EQUC) of selected tissue samples and found that most taxonomies of the sequencing could be recovered from the culture, including *Staphylococcus*, *Escherichia coli*, *Corynebacterium*, *Neisseria*, *Micrococcus*, *Sphingomonas*, *Weeksella*, *Alcaligenes*, *etc.*.

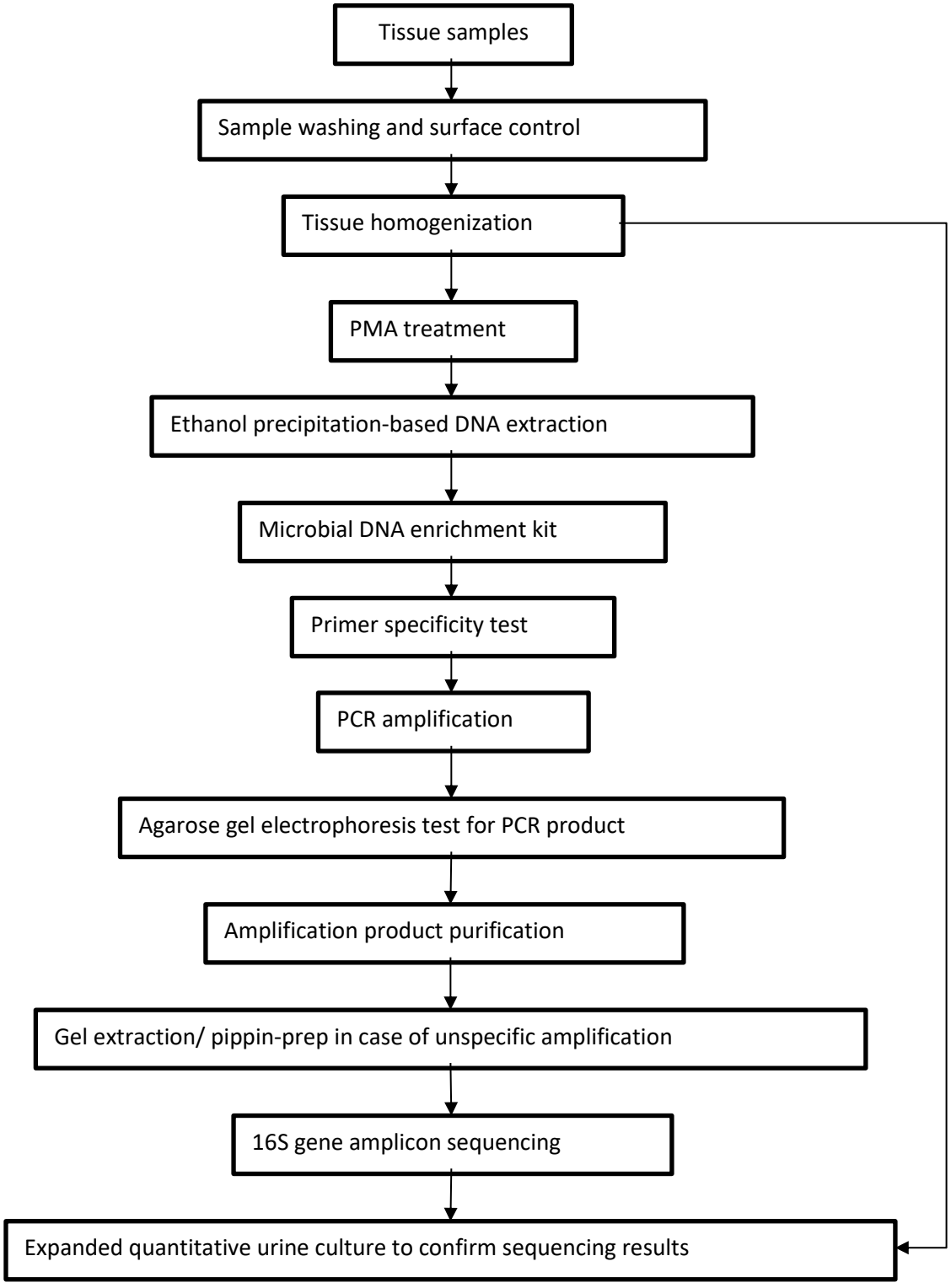

Figure 13. Refined tissue sample processing algorithm for "High-low" samples

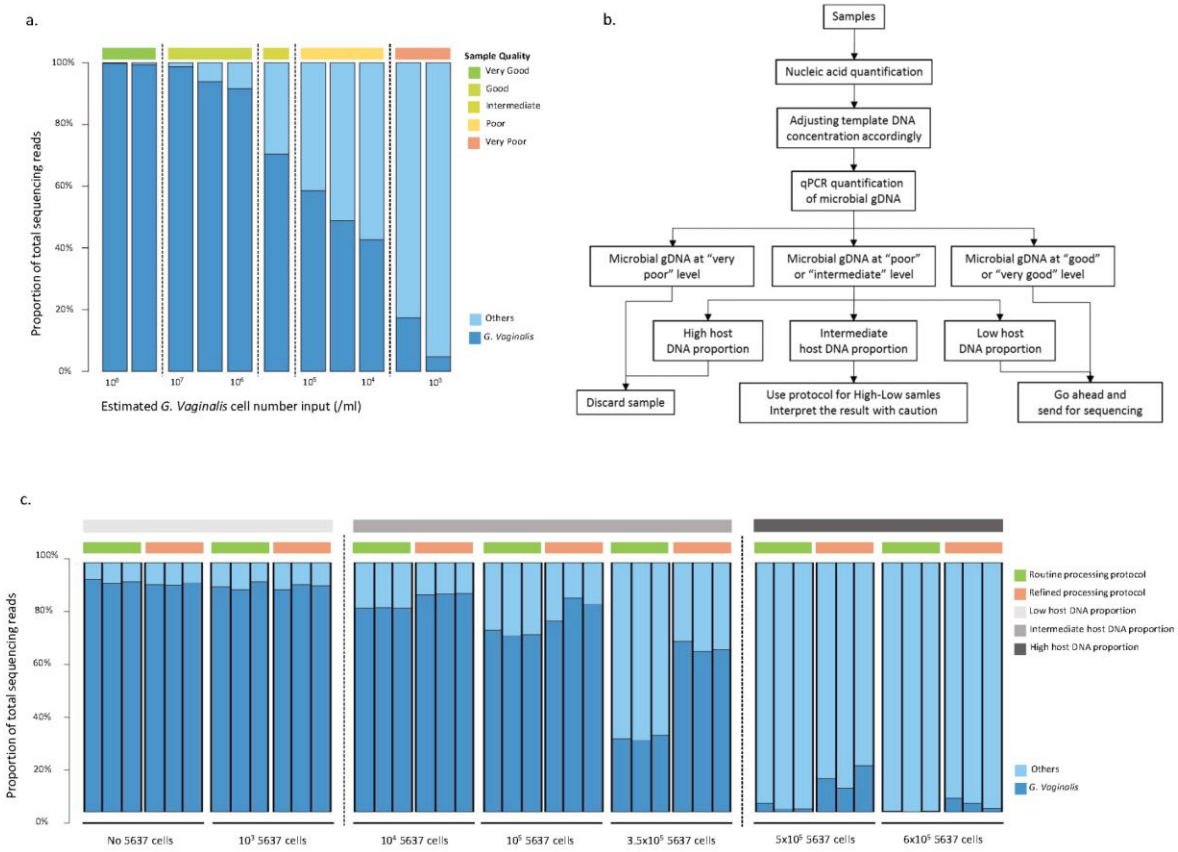

Figure 14. Developing a quality-checking algorithm for microbiota analysis of High-low samples. (a) Stacked bar plot demonstrating 16S rRNA gene sequencing result of serially diluted *G. Vaginalis* culture samples. Undiluted culture has about 10^8 cells. Each column represents a sample. As input cell number decreases, the proportion of non-target, "noise" sequencing reads increases. Sample quality was determined by the proportion of "noise" reads. (b) The processing algorithm for microbiota analysis of High-Low samples. Samples with high microbial abundance usually yield decent sequencing results, and samples with very low microbial abundance should be discarded. Samples with intermediate microbial and host DNA proportion should go through further processing to aaaaaaaoptimise the sequencing results. (c) 16S rRNA gene sequencing result of samples consisted of 10^6 cells of *G. Vaginalis* culture mixed with a various amount of bladder 5637 cells. Triplicates were taken for each sample to go through PCR and 16S rRNA sequencing. For one set of samples both routine sample processing protocols and refined sample processing protocols were utilized and the sequencing results were compared.

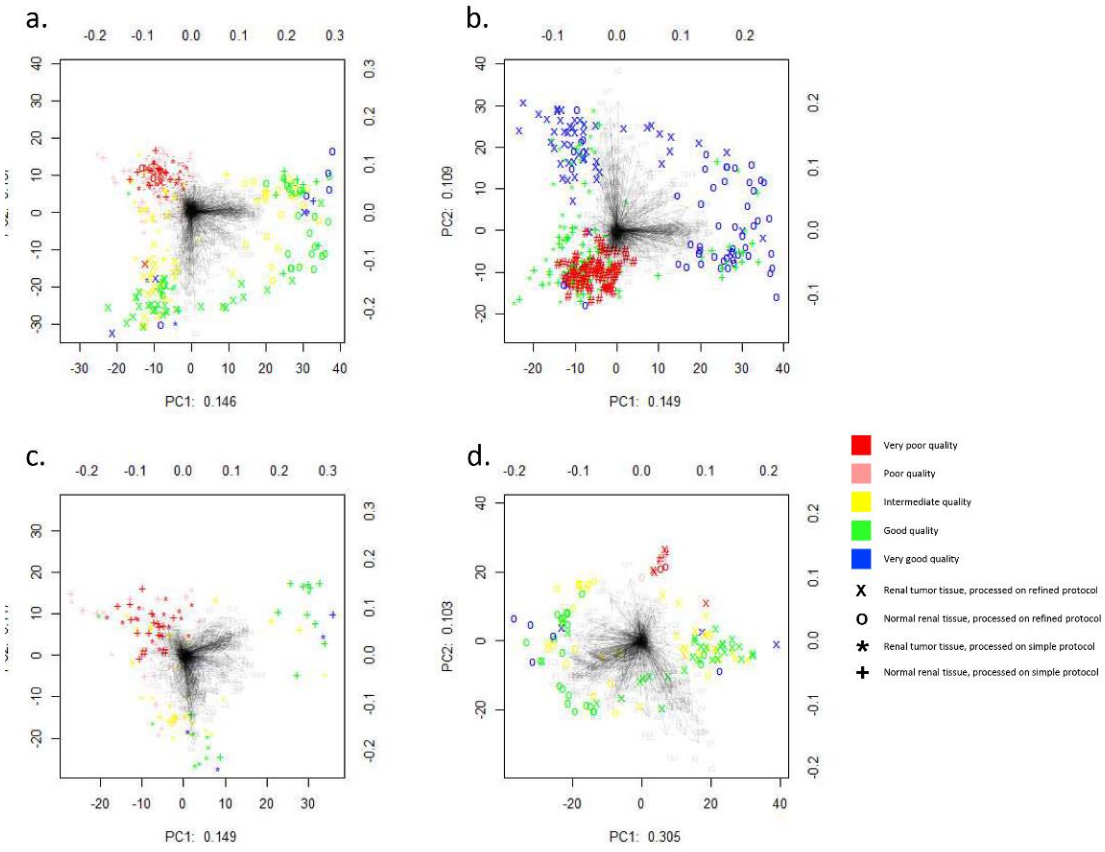

Figure 15. Compositional biplot of the renal tissue sample microbiota.

Samples were coloured to demonstrate their quality. The arrows indicate the amount and direction of variation of the ratio of each OTU to all others in the data set. The top and right axes indicate the values for variable loadings, and the bottom and left axes indicate the unit sum-of-squares values of the principle components. The proportion of variation explained by components 1 and two is also indicated. (a) Biplot of all samples. (b). Biplot of all samples but coloured differently. Here the surface control samples were in red, renal tissue samples that went through the refined protocol were in blue, while samples that went through simple protocol were in green. (c) Biplot the paired renal tumour tissue vs normal adjacent tissue samples that were processed under the simple protocol. (d) Biplot of the paired renal tumour tissue vs normal adjacent tissue samples that were processed under the refined protocol.

2.4 Discussion

The 16S rRNA gene sequencing based microbial identification is one of the most widely used methods to study human microbiota. Thanks to its culture-independent nature, numerous previously undescribed bacterial types were found inhabiting a human body (Lau *et al.*, 2004; Woo *et al.*, 2005). However, 16S rRNA gene sequencing-based methods also have drawbacks, including multiplex-associated biases, limited resolution at the certain species level, inability to differentiate between bacterial DNA and viable bacterial cells (Janda & Abbott, 2007). In this study, we showed that conventional sample processing methods derived from gut microbiota studies were not optimal for handling low microbial, high host abundance DNA samples. We then developed a refined protocol specifically for these sample types. Finally, utilising this protocol, we looked at the microbial consortia from renal carcinoma tissue and their normal adjacent counterparts.

Our results demonstrated that samples with high host eukaryotic DNA abundance and low microbial biomass ("high-low" samples) would compromise the sequencing results in multiple levels: **firstly, the eukaryotic and prokaryotic DNA compete in binding to the silicone filter in the DNA extraction columns** typically used in extraction kits, leading to a compromised prokaryotic DNA yield due to lack of capacity, as is shown in Figure 2b in comparison to 2d. In our experiment, this phenomenon was observed when both cells or free DNA was transferred into the DNA extraction column, suggesting that it is eukaryotic DNA, not cellular components, that compromised microbial DNA yield. Since the estimation of isolated microbial gDNA copy number was performed with qPCR, we also used microbial DNA enrichment kit (New England Biolabs) to deplete human DNA, and received similar results (data not

shown) This suggests the effect is not caused by inhibitive effect of human DNA during qPCR. In fact, commercially available DNA extraction kits can handle common samples without issues. Taking Qiagen QIAamp® and MOBIO Powersoil®, two of the most commonly used DNA extraction kit for gut microbiota studies for example, the QIAamp® column has a binding capability of up to 4-12 ug per filter, while the Powersoil® column has a binding capability of 20 ug per filter, as stated in their user manuals. Given that most gut microbes have a genome size around 4-6 Mb, this binding capability could retrieve presumably as high as 10^9 copies of bacterial gDNA, which will indubitably have a decent representation of gut microbial communities. However, when it comes to high-low samples, given eukaryotic cells' much higher genome size (3000 Mb *homo sapiean* genome), a DNA collection filter could only hold around $6.5x10^6$ cells in theory. This calculation matches our results that the bladder 5637 gDNA yield reaches a plateau at around 10^6 copies (Figure 2a). Since one gram of solid tumour is estimated to have 10^8 to 10^9 tumour cells (Del Monte *et al.*, 2009; Rasnick *et al.*, 2002) a common input of 200 ug tissue sample could contain $2x10^7$ to $2x10^8$ cells, which could easily saturate the DNA extraction columns. Given this and the low prokaryotic/eukaryotic gDNA ratio, the prokaryotic DNA yield could be very low in these columns and bear the risk of poor representation of the sample microbial community. To our knowledge, this column-associated risk factor is not recognised in tissue microbiota studies.

Secondly, high eukaryotic DNA concentration could also affect PCR-amplification of microbial DNA. In our qPCR quantification experiment, the Ct value of 10^9 copies of *E.coli* gDNA increased from 9.4 to 18.97 with the presence of

$3.5x10^5$ copies of bladder 5637 gDNA (Figure 7a). Although appropriate dilution of samples mitigated the inhibitive effect (Figure 7b), when microbial biomass dropped below 10^4 the sequencing result was severely biased by reagent and background DNA, as we shown later in Figure 14a. Therefore, it is crucial to understand which sample can have a better sequencing quality with dilution. Our quality-checking algorithm in Figure 14b helped to solve this issue.

Thirdly, high host DNA abundance leads to unspecific amplification of commonly used primers, consuming sequencing depths. Host DNA consuming sequencing depth is a known issue in whole genome sequencing and RNA sequencing (Feehery *et al.*, 2013). Studies with whole genome shotgun sequencing in the presence of large amounts of host eukaryotic DNA have shown that the background DNA will consume significant amounts of sequencing depth, making sequencing inefficient, time-consuming and inaccurate (Feehery *et al.*, 2013). In our study, it appears that the sequencing depth of 16S rRNA gene sequencing the presence of a high predominance of host DNA is also compromised, due to unspecific amplification of 16s rRNA gene primers. This phenomenon was previously reported in microbiota study of blood samples and heart valve samples (Kühn *et al.*, 2011; Rogina *et al.*, 2014), but to our knowledge, there's no report that this unspecific amplification consumes a remarkable number of reads in 16s rRNA gene sequencing of high-low samples. Based on our results, we recommend testing primer combinations for lowest unspecific amplification, using microbial enrichment kits to deplete human DNA, and perform amplification product purification by agarose gel cutting and extraction for the sample with very high host DNA abundance.

Based on these results, we then proposed that a refined protocol should be utilised when processing high-low samples. To avoid filter competition issue, we used ethanol precipitation as an alternative to column-binding systems to recover sample DNA and showed that it effectively recovers DNA without reaching a "capacity ceiling", and it did increase microbial DNA yield in mock samples (Figure 2c and 2d). We also utilised microbial enrichment kit to deplete eukaryotic DNA. As shown in our study, the

enrichment kit alone is insufficient to deplete host DNA completely (Figure 8), but it did exert positive effect when incorporated in the refined protocol (Figure 15).

In this protocol, besides issues associated with host DNA, we also seek to solve issues of samples with low microbial biomass by identifying the lower limit of sample microbial gDNA concentration level below which the 16S rRNA gene sequencing results will lose its fidelity and value. We used a *Gardnerella vaginalis* strain to perform this quality-checking experiment because it is commonly associated with the urinary tract and has not been reported as a contaminant in DNA extraction kits or other reagents. In our study, samples with a microbial abundance lower than 10^4/ml had >60% contamination reads and were discarded, while samples with 10-60% contamination reads went through the refined protocol for optimised sequencing quality. Since different labs have varying risks to contamination due to their environment and reagent usage, an individualised lower limit determination should be performed for each lab at least once before 16S rRNA gene sequencing, though such quality assurance checks don't appear to be described in the literature.

With our developed protocol, we examined renal tumour and healthy adjacent tissue from the same individuals for the presence of microbiota. Again, the refined protocol produced better sequencing quality, as better separation of tissue samples from reagent controls were observed. Many of the microbial species in the sequencing results were confirmed by enhanced culture results, suggesting exitance of viable microbes in renal tissues. Admittedly, conditions such as pyelonephritis typically occur where obstruction can be present in the urogenital tract. This infection can be the result of urolithiasis and tumour. The former has been reported in early investigations of the bacterial composition of pediatric kidney stones which appear to show the presence of *E. coli* (Barr-Beare *et al.*, 2015).

The limitation of this study includes: the refined protocol embeds several self-developed steps including ethanol precipitation, sample surface washing and PMA treatment *etc*. since no specialised regent or materials are commercially available. These steps bear risks of introducing contamination or decreased microbial DNA yield when not

appropriately handled. On the other hand, culture and sequencing-based techniques are unable to provide direct visual evidence of bacteria in renal tissue, and therefore unable to exclude the possibility of microbial contamination during renal tissue collection, storage and handling.

In summary, host DNA and reagent/environmental DNA have a negative effect on sequencing qualities of high-low samples. Our self-developed refined protocol could mitigate some of these effects and revealed microbiota profiles in renal tissue samples.

2.5 References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J. & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science translational medicine*, 6(237), 237ra65.

Andersen, B. R. & Jackson, G. G. (1961). Pyelitis, an important factor in the pathogenesis of retrograde pyelonephritis. *The Journal of experimental medicine*, 114(3), 375–384.

Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., Semenkovich, C. F., et al. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America*, 101(44), 15718–23.

Barr-Beare, E., Saxena, V., Hilt, E. E., Thomas-White, K., Schober, M., Li, B., Becknell, B., et al. (2015). *The interaction between Enterobacteriaceae and calcium oxalate deposits*. PloS one, 10(10), e0139575.

Berer, K., Mues, M., Koutrolos, M., Rasbi, Z. A., Boziki, M., Johner, C., Wekerle, H., et al. (2011). Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. Nature, 479(7374), 538–41.

Branton, W. G., Ellestad, K. K., Maingat, F., Wheatley, B. M., Rud, E., Warren, R. L., Holt, R. A., et al. (2013). Brain microbial populations in HIV/AIDS: α-proteobacteria predominate independent of host immune status. *PloS one*, 8(1), e54673.

Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I. & Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science*, 326(5960), 1694–1697.

Ding, T. & Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. *Nature*, 509(7500), 357–60.

Dontas, A. S., Kasviki-Charvati, P., Papanayiotou, P. C. & Marketos, S. G. (1981). Bacteriuria and survival in old age. *New England Journal of Medicine*, 304(16), 939–943.

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., et al. (2013). A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PloS one*, 8(10), e76096.

Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., Codelli, J. A., et al. (2013). Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell*, 155(7), 1451–63.

Janda, J. M. & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of clinical microbiology*, 45(9), 2761–4.

Kass, E. H. (1972). The search for asymptomatic pyelonephritis. *New England Journal of Medicine*, 287(11), 563–564.

Koren, O., Spor, A., Felin, J., Fåk, F., Stombaugh, J., Tremaroli, V., Behre, C. J., et al. (2011). Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl 1, 4592–8.

Kühn, C., Disqué, C., Mühl, H., Orszag, P., Stiesch, M. & Haverich, A. (2011). Evaluation of commercial universal rRNA gene PCR plus sequencing tests for identification of bacteria and fungi associated with infectious endocarditis. *Journal of clinical microbiology*, 49(8), 2919–2923.

Lau, S. K., Woo, P. C., Woo, G. K., Fung, A. M., Wong, M. K., Chan, K. M., ... & Yuen, K. Y. (2004). Eggerthella hongkongensis sp. nov. and Eggerthella sinensis sp. nov., two novel Eggerthella species, account for half of the cases of Eggerthella bacteremia. *Diagnostic microbiology and infectious disease*, 49(4), 255-263.

Markle, J. G., Frank, D. N., Mortin-Toth, S., Robertson, C. E., Feazel, L. M., Rolle-Kampczyk, U., ... & Danska, J. S. (2013). Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science*, 339(6123), 1084–1088..

Del Monte, U. (2009). Does the cell number 109 still really fit one gram of tumor tissue? *Cell Cycle*, 8(3), 505–506.

Potgieter, M., Bester, J., Kell, D. B. & Pretorius, E. (2015). The dormant blood microbiome in chronic, inflammatory diseases. *FEMS microbiology reviews*, 39(4), 567–591.

Rasnick, D. (2002). Aneuploidy theory explains tumor formation, the absence of immune surveillance, and the failure of chemotherapy. *Cancer genetics and cytogenetics*, 136(1), 66–72.

Rogina, P., Skvarc, M., Stubljar, D., Kofol, R. & Kaasch, A. (2014). Diagnostic utility of broad range bacterial 16S rRNA gene PCR with degradation of human and free bacterial DNA in bloodstream infection is more sensitive than an in-house developed PCR without degradation of human and free bacterial DNA. *Mediators of inflammation*, 2014.

Segal, L. N. & Blaser, M. J. (2014). A brave new world: the lung microbiota in an era of change. *Annals of the American Thoracic Society*, 11(Supplement 1), S21–S27.

Urbaniak, C., Cummins, J., Brackstone, M., Macklaim, J. M., Gloor, G. B., Baban, C. K., Scott, L., et al. (2014). Microbiota of human breast tissue. *Applied and environmental microbiology*, 80(10), 3007–3014.

Wolfe, A. J., Toh, E., Shibata, N., Rong, R., Kenton, K., FitzGerald, M., Mueller, E. R., et al. (2012). Evidence of uncultivated bacteria in the adult female bladder. *Journal of clinical microbiology*, 50(4), 1376–1383.

Woo, P. C. Y., Lau, S. K. P., Chan, K. M., Fung, A. M. Y., Tang, B. S. F., & Yuen, K. Y. (2005). Clostridium bacteraemia characterised by 16S ribosomal RNA gene sequencing. *Journal of clinical pathology*, 58(3), 301-307.

Chapter 3

3 Exploration and comparison of the urinary microbiota of healthy individuals and urinary stone patients.

3.1 Introduction

Few clinical practitioners believe that the urinary tract of a healthy individual harbours a community of viable microorganisms. The dogma during medical education is that the presence of bacteria in urine correlates with pathological conditions such as asymptomatic bacteriuria or urinary tract infections (UTI). The definition of these conditions are based on one paradigm: that is urine in healthy individuals is "sterile" and when the numbers of bacteria caught from a urine sample exceeds a given threshold (usually 10^5 colony-forming units [CFU] per ml), a determination of a "positive culture result" is made, and the patient is considered as having a pathological condition. However, microbiologists are becoming gradually aware that traditional culture methods favour fast-growing microbial species and ignore the slower-growing, more fastidious commensal organisms (Pace *et al.*, 1997). The later discoveries of intracellular bacteria in urothelial cells and biofilms formation in urinary tract showed evidence of bacteria in urine culture-negative individuals, and further weakened the solidity of the paradigm (Hancock *et al.*, 2007; Salo et al., 2009). Since the appearance of culture-independent, sequencing-based microbial analysing methods over the past two decades from denaturing gradient gel electrophoresis to next-generation sequencing analysis, multiple studies have reported evidence of microorganisms in the urinary tract (Burton *et al.*, 2011).

Compared to other human sites occupied by bacteria, the urinary tract does not harbour the richest habitat for microbes in the human body, which is partly the reason why it may not have been included as a site examined in the original human microbiome project (Consortium *et al.*, 2012). We have shown in the previous chapter, that samples with low microbial mass are more vulnerable to bias. This may be one of the reasons why previous studies of urinary tract microbiota all have discrepancies between their reported urinary

microbial community compositions. Additionally, there are no optimal sample storage and processing methods for urine samples yet described in the literature.

Many links are now emerging between the microbiota and numerous conditions, not just those associated with the gut microbiota and the urinary microbiota should not be overlooked either. For example, urinary stones are among the most common urological conditions, with 16% of the male and 8% of the female population developing at least one symptomatic stone episode in their lifespan (Scales *et al.*, 2012). Previously, urologists only thought that bacteria only played a role in the formation of a certain specified type of urinary stones, namely struvite stones, but in fact, reports have shown that bacteria could be found in most commonly seen calcium oxalate stones (Bratell *et al.* 1990; Dewan *et al.*, 1997; Gault *et al.*, 1995; Hugosson *et al.* 1990; Larsen *et al.*, 1986; Lewi *et al.*, 1984; McCartney *et al.*, 1985). At present, a systemic description of urinary microbiota patterns among urinary stone patients is still lacking. Such a description could provide useful mechanistic and potentially diagnostic information.

In this chapter, I detail how I have optimised the methodology for urine sample collection, storage and processing, and microbiota analysis. Then using these methodologies, I undertook an analysis of the urinary microbiota of urinary stone patients, comparing it to a group of age-matched, healthy individuals. I further explored the longitudinal stability of the urinary microbiota of individuals in these two groups.

3.2 Methods

3.2.1 Evaluation of optimal storage of urine for microbiota analysis

It is well documented that the storage and handling of samples for 16S rRNA gene and other types of NGS analysis should be undertaken in a standard manner (Di Bella *et al.*, 2013). Ideally, the extraction, sequencing and analysis should be done simultaneously and therefore, the optimisation of the storage of clinical samples until this can occur is important (Bao *et al.*, 2017).

To determine the optimal volume and processing for the storage of urine for microbiota analysis, urine samples from four healthy donors were collected. The demographic information of these four donors is shown in table 5.

Table 5. Demographic information of healthy donors for storage methods optimization

Donor	Gender	Age	Active UTI
1	Male	28	No
2	Female	23	No
3	Female	68	No
4	Male	26	No

Urine sample collection methods are the same as the clinical urine sample collection methods, described below in the “Sample collection” section. Each donor provided multiple urine samples that were first pooled together then distributed into multiple forms as described in table 6. Samples in groups A and C were centrifuged at 1000 x g for 10 minutes. The supernatant was discarded, and the pellets were stored at -80 ℃. In groups B and D, whole urine of the desired volume was transferred to a sterile urine sample container (Starplex, VWR) then stored in -80 ℃.

3.2.2 Clinical study design and study population

All specimens used in this study were collected, following a protocol approved by the appropriate human ethics boards at Western University, Canada and West China Hospital, Sichuan University, China. The participants in the stone group were patients admitted to St. Joseph Hospital, London, ON, Canada and West China Hospital, Sichuan University, China. Inclusion criteria were participants at least 18 years of age, non-pregnant, with no other infectious conditions, not receiving antibiotics or chemotherapy.

The diagnosis of urinary stone was confirmed with radiological investigation, and the patient would be scheduled to receive surgical intervention (either extracorporeal shock wave lithotripsy (ESWL), percutaneous nephrolithotomy (PCNL) or ureteroscopy) at St. Joseph's Hospital. Patients were sub-grouped depending on their gender, age and location of the stones, type of surgery they receive, chemical composition of the stone and whether a ureteral stent is inserted. After informed consent, participants were asked to provide two urine samples: one at the patient's clinic visit for preoperative evaluation, and one at the patients visit for post-operative follow up. For healthy volunteers, two urine samples were collected with a gap of three months. All patient information was blinded until all sample processing and DNA sequencing work were completed and only then would the patients identified by their name initials to retrieve their grouping and demographic information for statistical analysis.

Table 6. Urine sample collected for storage methods optimization

Groups	Volume	Forms	Storage period
A	2ml 10ml 20ml 50ml 100ml	Pellets	1 Month
B	2ml 10ml 20ml 50ml 100ml	Whole urine	1 Month
C	20ml	Pellets	1 Month 6 Months 1 Year 2 Years
D	20ml	Whole Urine	1 Month 6 Months 1 Year 2 Years
Control	2ml	Reagents	1 Month 6 Month 1 Year 2 Years

3.2.3 Sample collection

All participants were instructed to wash their hands thoroughly with soap and water, then to clean and wash their external genitalia with an antiseptic towelette. Urine samples were taken with a clean-catch, midstream voided fashion: the first 10 mL was discarded, the majority of the voided volume collected in a sterile container, and the final voided volume was dispensed into the toilet. Once collected, all urine samples were processed within one hour. Microbiological cultures were performed subsequently, and the remaining of urine samples were spun down to pellets and stored at -80 ℃.

3.2.4 Microbiological culture

All collected samples were quantitatively cultured on Colombia Blood Agar with 5% sheep blood, LB agar and MacConkey agar. Urine samples with a CFU > 105 per mL were to be considered as active UTI and excluded. All cultivated strains were isolated and stored. Enhanced quantitative urine culture (EQUC) was also applied following protocols developed by Hilt et al. (Hilt *et al.*, 2014). In brief, urine samples were cultured with higher volume input (100ul and up to 1ml), enriched culture media (CBA, CAN agars and thioglycolate medium), both aerobic and anaerobic culture atmosphere and extended culture time (up to 5 days). For each morphologically unique colony on the plate it was re-streaked into a new plate for taxonomy identification. The culture of *Ureaplasma* species was following the clinical laboratory protocol adapted from ATCC 2616 Medium: basal medium [3.5 g PPLO broth, 10 g pancreatic digest of casein (tryptone), 5 g pancreatic digest of gelatine (peptone) and 500 ml distilled water, followed by aseptic addition of filtered sterilised solutions [50 ml 10X CMRL 1066 medium, 35 ml yeast extract solution, 20 ml yeastolate (10% w/v), 170 ml foetal bovine serum, 20 ml 0.1% phenol red solution, 1 g urea and 205 ml distilled water.

3.2.5 DNA isolation

After storing, the samples were thawed on ice. DNA extraction was performed with MOBIO PowerSoil® 96 well kit. In brief, lysis buffer and beads were added to urine or urine pellets. Lysis of cells was achieved with mechanical and chemical methods. The

suspension was then transferred to a new tube, and the subsequent procedure was following manufacturer's instructions.

3.2.6 16S rRNA gene library generation and sequencing

Extracted DNA from the clinical samples were PCR-amplified with barcoded primers as described by Gloor *et al*. (2010) shown in Appendix A. The PCR was performed in a 42-μl reaction system, with details listed in Appendix B. Following PCR, the amplicons were quantified with the Qubit 2.0 fluorometer (Invitrogen, CA, USA). Equimolar amounts of each amplicon were aliquoted into a single tube purified with the QIAquick purification kit (Qiagen, Mississauga, ON, Canada). Each product was then pooled and sequenced on the Illumina Mi-Seq platform in a paired-end run using a 600 cycle kit with a paired-end 220-bp run at the London Regional Genomics Center (Robarts Research Institute, London, ON, Canada), following standard operating procedures.

3.2.7 Sequencing reads processing and taxonomic assignment.

Description of the processing algorithm has been previously published (Gloor *et al.*, 2010). The SOP and related software are accessible at http://github.com/ggloor/miseq_bin. The sequences were processed following algorithm as stated in Appendix C. Paired-end reads were overlapped with Pandaseq V2.5 (Masella *et al.*, 2012), allowing no ambiguous positions in the overlapped reads. Reads were then demultiplexed and tagged with the name corresponding to the sample identifier and barcode. After demultiplexing, reads from all samples were merged into one single file and were clustered into individual sequence units (ISUs), which were then further grouped into operational taxonomic units (OTUs) with open reference OTU picking method, using Uclust algorithm of USEARCH version 7, which has a *de novo* chimera filtering step built into it (Edgar *et al.*, 2010; Edgar *et al.*, 2015). The identity threshold was 97%. Singleton OTUs and ones with abundance < 1% were discarded. Taxonomical assignment was performed with mothur 1.3.4 to annotate the OTU sequences against the Silva database (v119).

3.2.8 Bioinformatic analysis

For data exploration, OTUs with a zero value was replaced by a small numerical value with the count zero multiplicative methods in the zCompositions package (Palarea-Albaladejo *et al.*, 2015). Custom R codes were generated to perform centred log ratio (clr) transform of the zero-replaced data set. Then a singular value decomposition (SVD) was applied to the data. The output was used for PCA plots exploration. 128 Dirichlet Monte-Carlo (DMC) instances from the data with clr-transform were used for quantitative data analysis. ALDEx 2 V.1.6.0 package were used for differential abundance test. Significance was based on the Benjamini-Hochberg corrected P value of the Wilcoxon rank test (significance threshold $P < 0.1$).

3.2.9 qPCR estimation of 16S rRNA gene copy numbers.

The bacterial load in the urine samples were estimated by qPCR quantification of the DNA extracts. The reaction system includes 10ul of PCR master mix, 5 ul of 800Nm primer stock (V6 hypervariable region of 16S rRNA gene: V6-LT 5' - CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNCWACGCGARGAACCTTACC3'

and V6-RT: 5'CCTCTCTATGGGCAGTCGGTGATACRACACGAGCTGACGAC 3'), 5ul of sample DNA template. A standard curve was made with known copies of *E.coli* DH5a gDNA.

3.3 Results

3.3.1 Methods optimization for sample storage

We first investigated how urine samples storage/preserving techniques could affect their sequencing qualities. Urine samples from healthy donors were either kept in pellets (Group A) or as whole urine (Group B) for one month and then PCR-amplified with V4 primer and sent for sequencing. Among all group A samples, the ones stored at a higher volume had greater weighted UniFrac distances to reagents, suggesting it provided a

representation of the urinary microbiota (Figure 16). Figure 17 presents urine samples from one donor that were either stored as whole urine or in pellets. As shown in the PCOA plot, urine pellets had better separation from the reagents, suggesting a better representation of the urinary microbiota. Urine samples with higher volumes generally had better representation. Pooling the UniFrac distances to reagents of samples from all four donors, in all volume levels the pelletted urine samples had higher weighted UniFrac distances to reagents than whole urine samples. For pelleted urine samples, the weighted UniFrac distances to reagents tended to stabilize after urine samples have a volume higher than 20 ml, while for whole urine samples it seemed to need more than 100ml to reach a decent weighted UniFrac distance to reagents (Figure 18). These results showed that urine stored in pellets might have a better representation of the urinary microbiota of the donor. Samples pelleted from a urine sample with a volume higher than 20 ml seemed to have better representation.

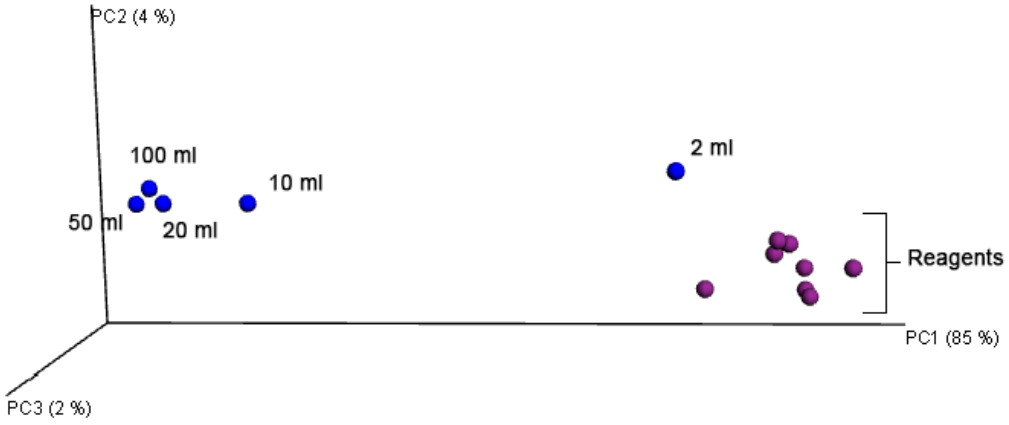

Figure 16. PCOA plot showing weighted UniFrac distances of urine samples stored in pellets (Group A) with various volume and reagent controls. The volume of each sample is marked in the figure. For better representation only one sample of each volume from one donor.

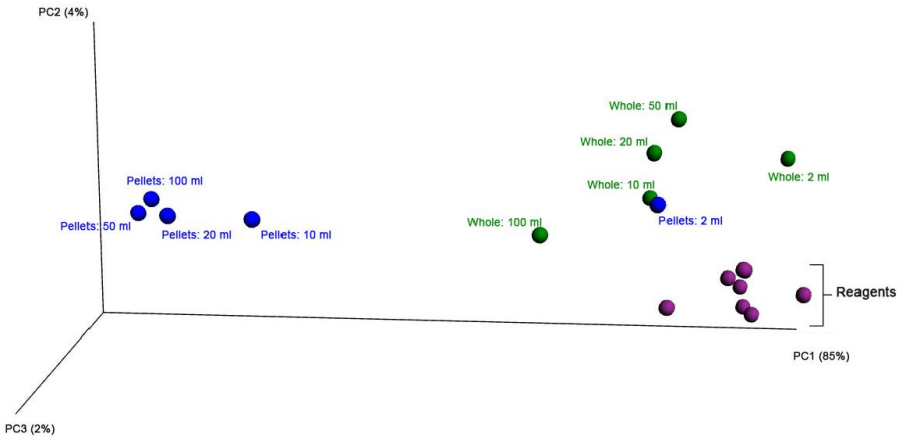

Figure 17. PCOA plot showing weighted UniFrac distances of pelleted urine samples (Group A, blue) and whole urine samples (Group B, green) and reagents. The volume of each sample is marked in the figure. For better representation, only one sample of each volume from one donor is shown here.

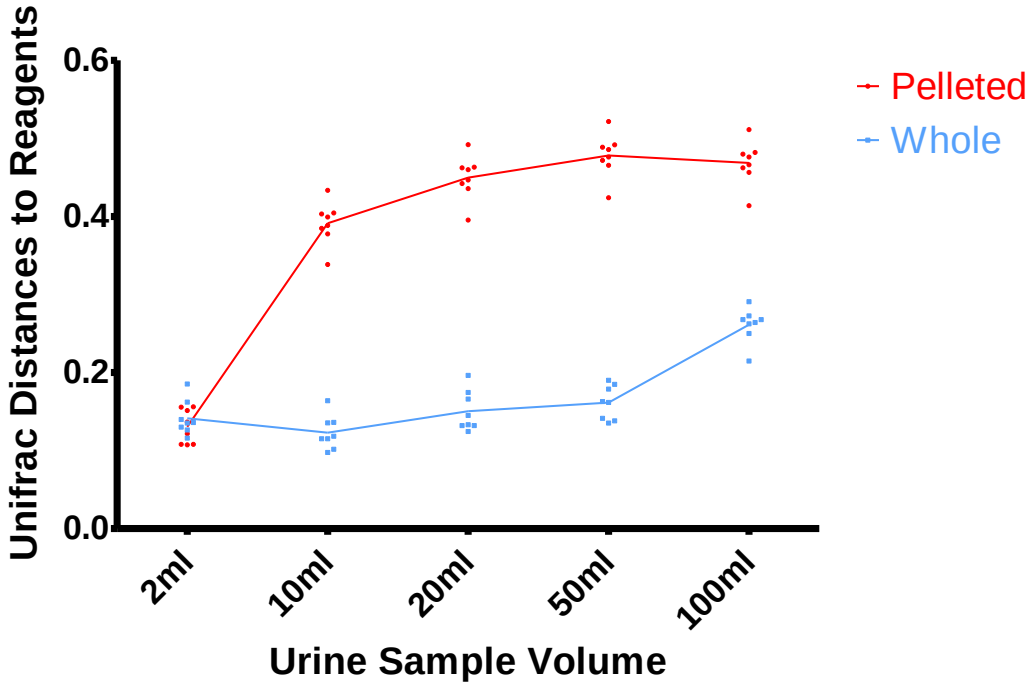

Figure 18. UniFrac distances of samples from all four donors, both in pellets and whole urine (Group A & B), to reagents.

We then proceeded to explore if storage time could affect sequencing results. Both urine pellets (Group C) and whole urine (Group D) were stored in -80℃ for one month, six months, 1 and two years before they were processed and sent for sequencing. Group C samples demonstrated a relatively consistent UniFrac distances to reagent through various time points, while Group D samples showed a more obvious decrease over time (Figure 18 and Figure 19). When compared to samples that only stored in one month, Group D samples over time showed a more remarkable increase in weighted UniFrac distances while Group C sample maintained stable (Figure 19). These results showed that stored sample over time would decrease in quality, but samples kept in pellets would mitigate these effects.

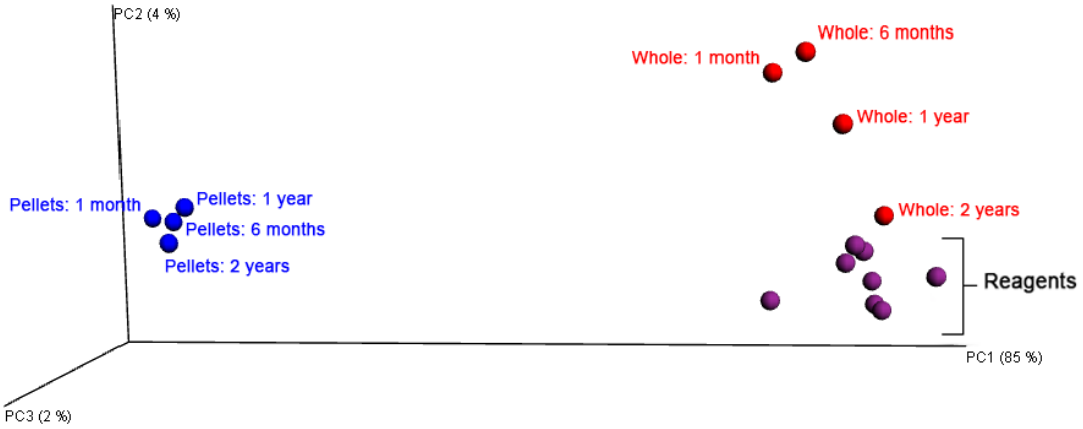

Figure 19. PCOA plot showing weighted UniFrac distances of pelleted urine samples (Group C, blue) and whole urine samples (Group D, red) and reagents. The storage time of each sample is marked in the plot. For better representation, only one sample of each volume from one donor is shown here.

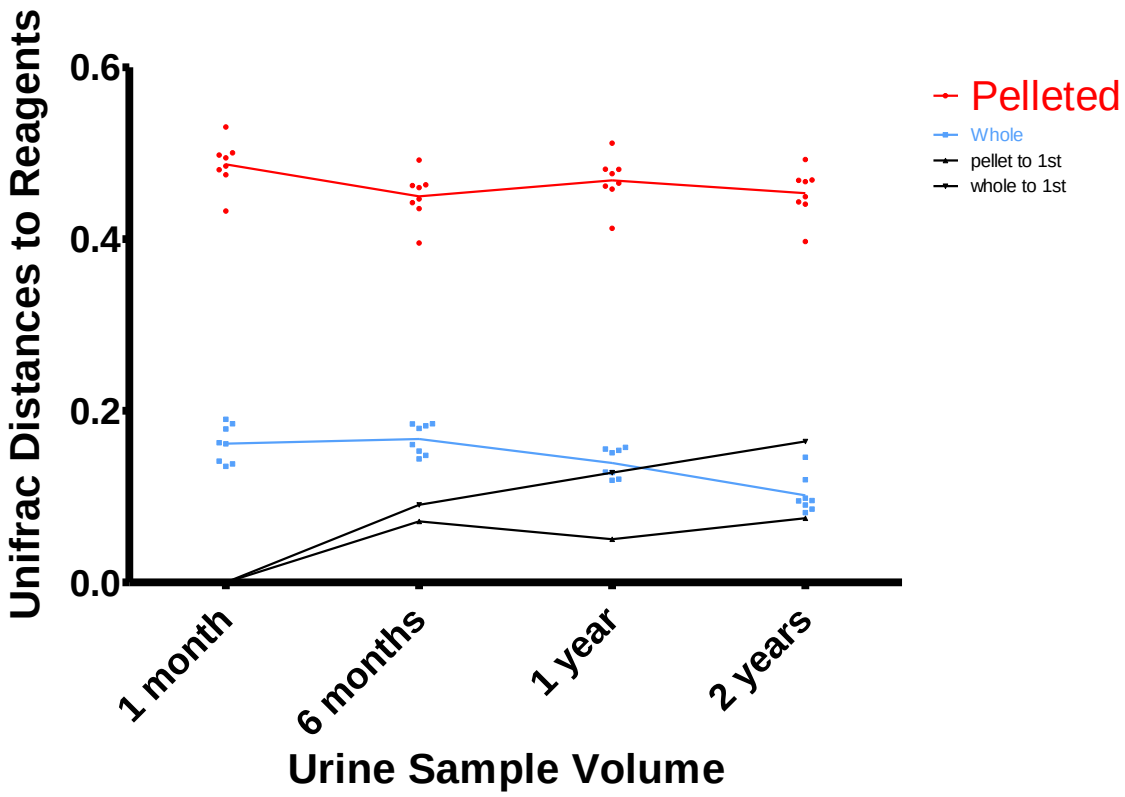

Figure 20. UniFrac distances of samples from all four donors, both in pellets and whole urine (Group C & D), to reagents. The black dots showed UniFrac distances of samples over time to samples stored in one month.

3.3.2 Microbiota pattern of urinary stone patients

3.3.2.1 Characterization of the urinary microbiota of healthy subjects and stone patients.

A total of 116 urine samples were collected. The demographic information of all donors is presented in Table 7. We characterised the microbial composition of urine samples from 41 from healthy individuals (22 female and 19 male) and 61 from stone patients (31 female and 30 male), after excluding samples with bacterial growth exceeding 10^5 CFU/mL and samples with low sequencing reads. Four hundred and fourteen bacterial operational taxonomic units (OTUs) in total were identified in the sequencing result. On average, one sample contained 5338 reads.

The urinary microbiota patterns of women had higher alpha diversity compared to men, although there was no statistical significance (Figure 21). An exploratory compositional biplot of the samples from healthy male and female donors showed that 21.4% and 11.8%, respectively, of the variation could be explained by the first and second components, and was likely due to the relative abundance of *Lactobacillus*, *Corynebacterium* and *Staphylococcus* species. The urinary microbiota of the two genders demonstrated distinct patterns: urine samples from a male donor were most commonly predominated by a single genus, usually *Corynebacterium*, *Staphylococcus* and *Streptococcus,* and occasionally by *Lactobacillus, Enterobacteriaceae* species (presumably *Escherichia coli*), *Anaerococcus, Gardnerella* and *Prevotella*. Urine samples from a female donor, on the other hands, consisted of a more heterogeneous microbial community, usually predominated by *Lactobacillus* species (Figure 22). When comparing the urinary microbiota of healthy male and female, it was noted that female urinary microbiota contained higher proportions of *Lactobacillus* species, while male urinary microbiota contained higher proportions of *Corynebacterium* or *Staphylococcus* species (Figure 23-28). qPCR quantification of 16S rRNA gene operon copy numbers showed that a voided urine sample from a female donor has slightly higher copy number of rRNA operons than her male counterpart, both healthy subjects and stone patients, but without statistical significance (Figure 29)

Table 7. Demographic information of the urinary microbiota study participants (3).

	Urinary stone patients		Healthy volunteers	
Female (n, %)	31	50.8	22	53.7
Age (n, %)				
20-29	3	4.9	2	4.9
30-39	6	9.8	8	19.5
40-49	16	26.2	8	19.5
50-59	20	32.8	13	31.7
60-69	13	21.3	5	14.6
70-79	3	4.9	5	9.7
Median	51	-		
Operation (n, %)				
SWL	17	27.9	-	-
URS	37	60.7	-	-
PCNL	7	11.2	-	-

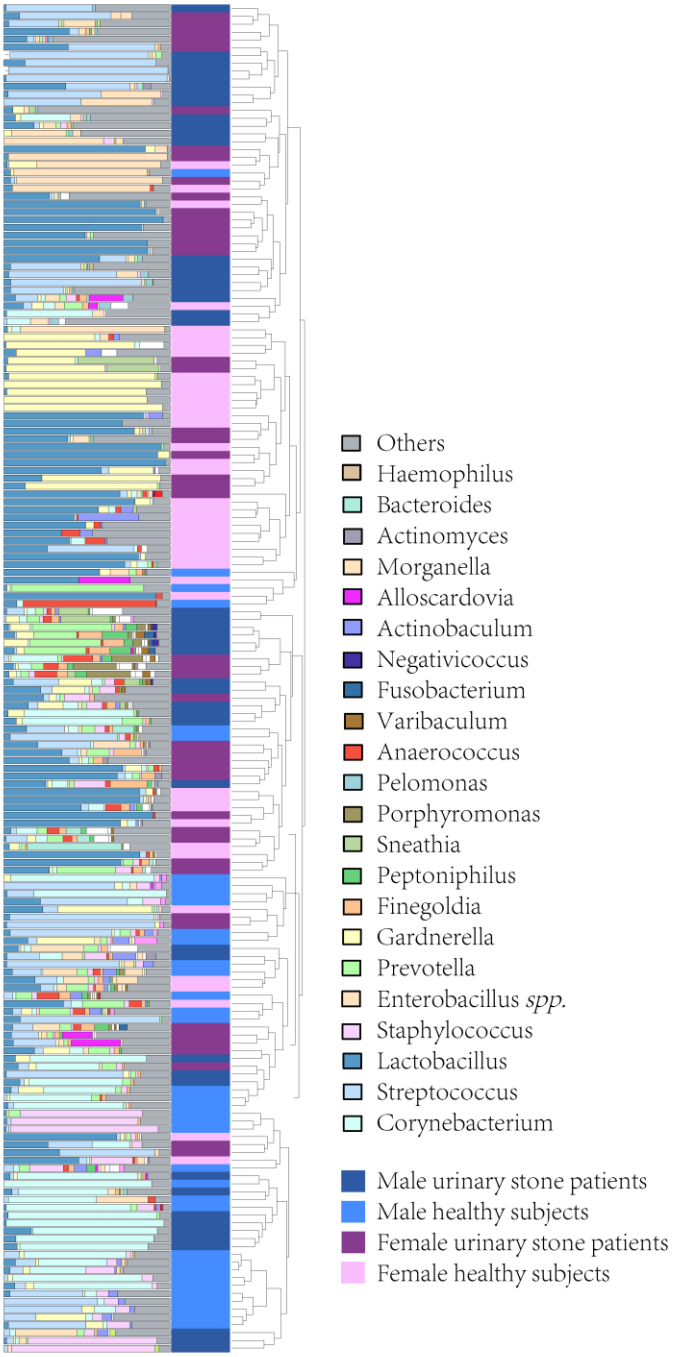

Figure 21. Relative abundance taxa bar blot of urine samples from healthy individuals and stone patients. Taxa (groups of 16s rRNA gene sequences) are represented at the genus level. The dendrogram represents the hierarchical average-linkage clustering of the microbiota profiles based on the computed weighted UniFrac distances.

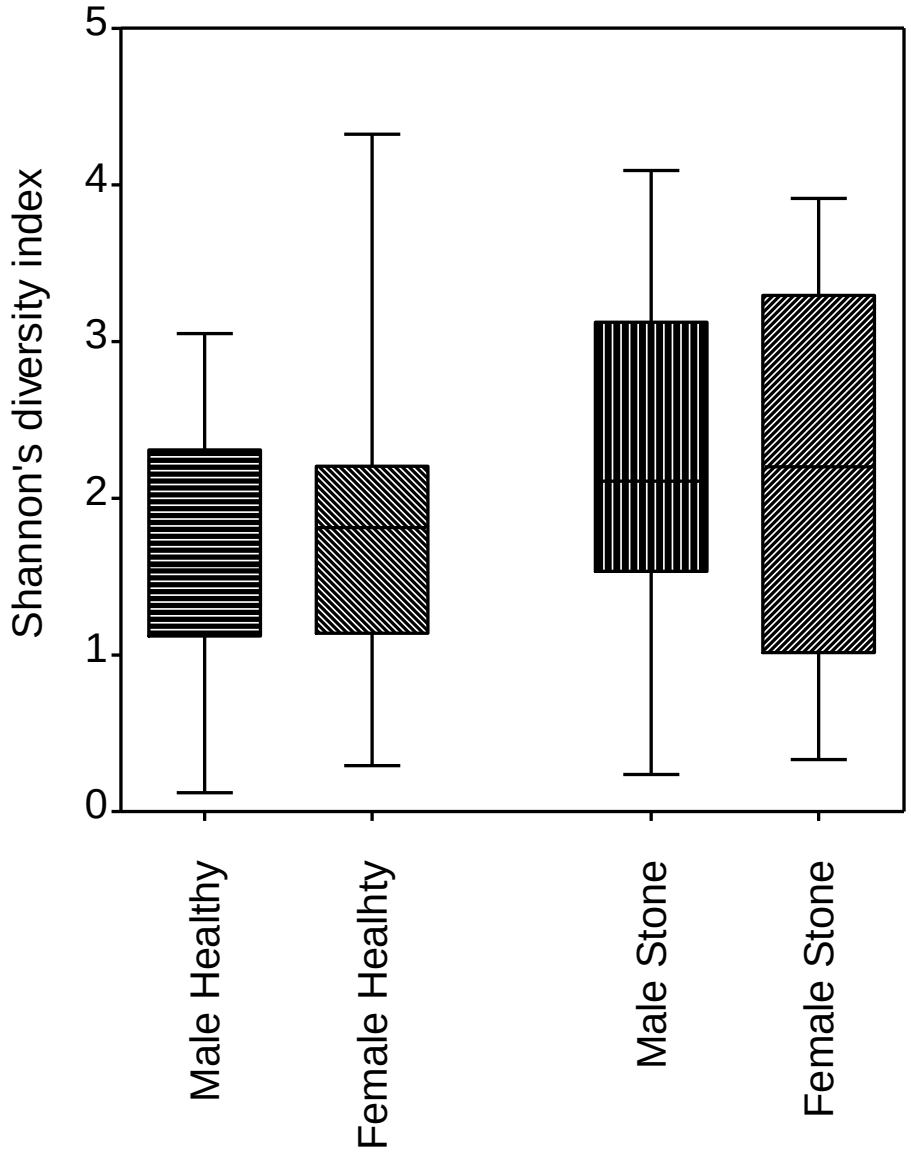

Figure 22. Distribution of alpha diversity within groups as measured by Shannon index. The left two boxes represent the calculated Shannon index for urine microbiota samples from healthy male and female donors while the right two boxes were samples from male and female patients of urinary stones.

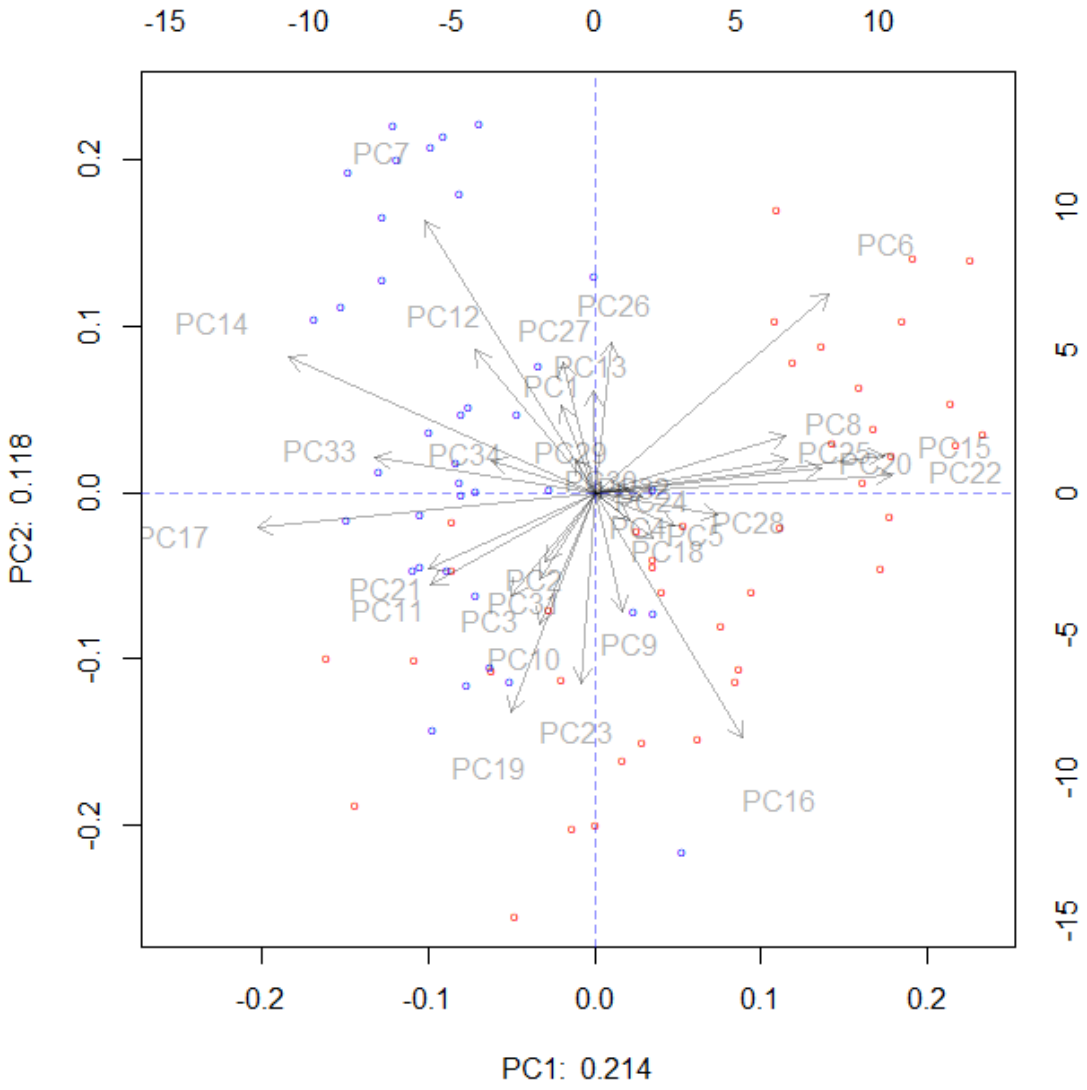

Figure 23. Compositional biplot of urinary microbiota of healthy male donors (blue) and healthy female donors (red)

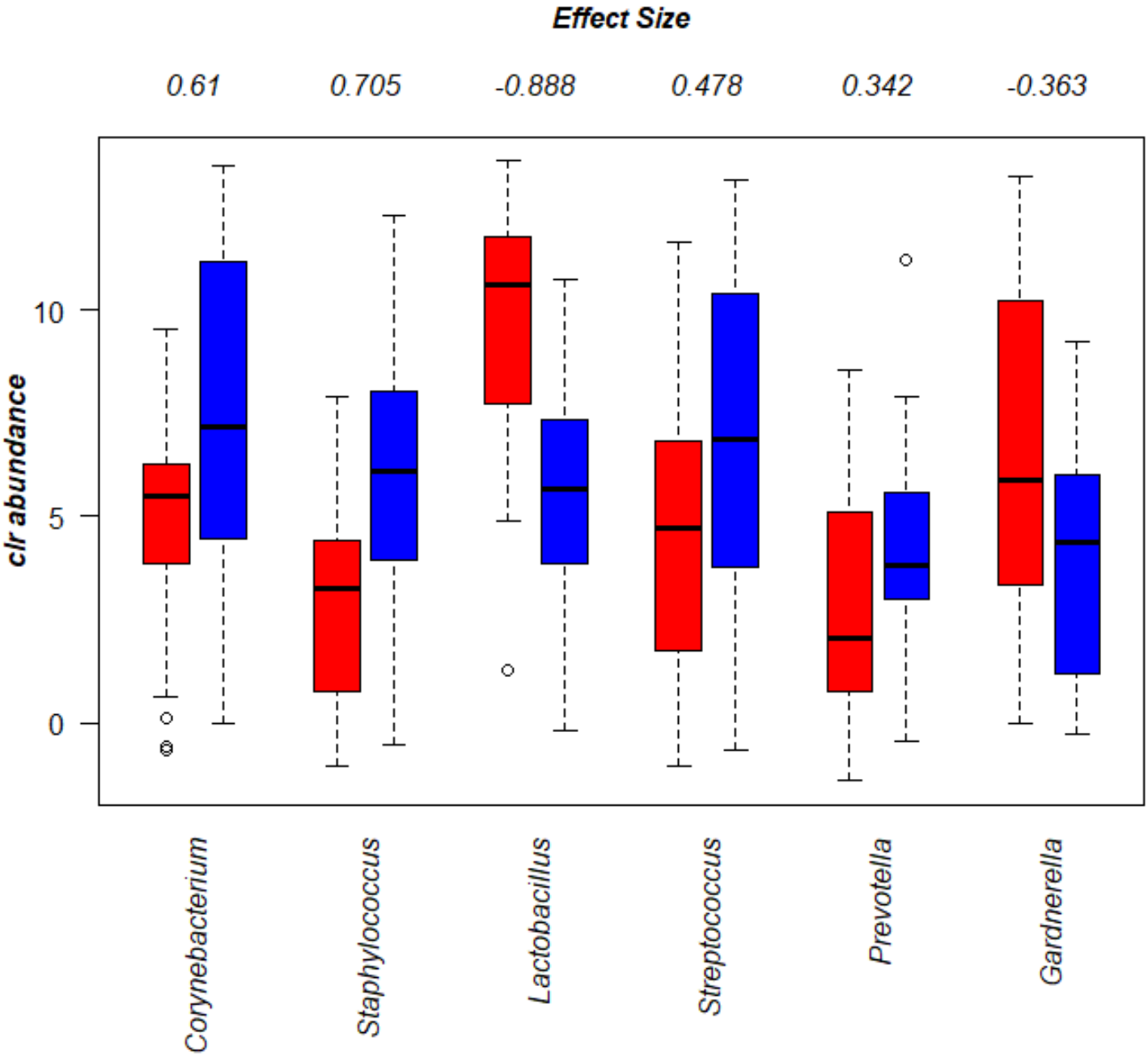

Figure 24. Box plot of the difference in bacterial abundance between healthy male and female. Taxa with high effect sizes are shown. The bacterial abundances were presented in relative abundance as computed by ALDEx2.

When comparing the urinary microbiota between the healthy individuals and the stone patients, a trend of separation could be observed in both male and female groups on biplot, but there was no clear separation (Figure 25 and 27). Male stone patients had slightly higher relative abundances in *Escherichia* and *Fingoldia* species comparing to healthy donors, while female stone patients had slightly higher relative abundances in *Prevotella, Streptococcus, Finegoldia* and *Peptoniphilus* species, although none of these differences reached statistical significance (Figure 26 and 28). The analysis based on weighted UniFrac did not observe a clear clustering of healthy or stone samples at the genus level (Figure 31). Similarly, no differences in alpha and beta diversity were observed among the level of stone size, location and composition (data not shown). In summary, no taxa were found significantly different between stone patient and healthy individual in both genders in this study.

3.3.2.2 Longitudinal stability of urinary microbiota

The aim was to investigate how the urinary microbiota changed throughout time. Two samples were collected from each person with approximately three months between collections (refers to the demographic table). Of all included participants, 70 had their follow-up sample collected (18/19 healthy male, 19/22 healthy female, 17/30 male stone, 16/31 female stone).

The UniFrac phylogenetic distance metric was calculated to measure the distance of the samples from one donor of two time points to samples from one donor to another donor. The UniFrac score ranges from 0 to 1. A lower UniFrac score means higher similarity of the microbiota between the compared samples to samples with higher UniFrac scores. If the UniFrac score was 0, it means the microbiota profiles of the two samples were exactly the same. If the UniFrac score was 1, it suggests two completely microbiota profiles with no overlapping species. Figure 30 demonstrated the UniFrac distances of samples from one donor and samples from different donors. Across all conditions (male

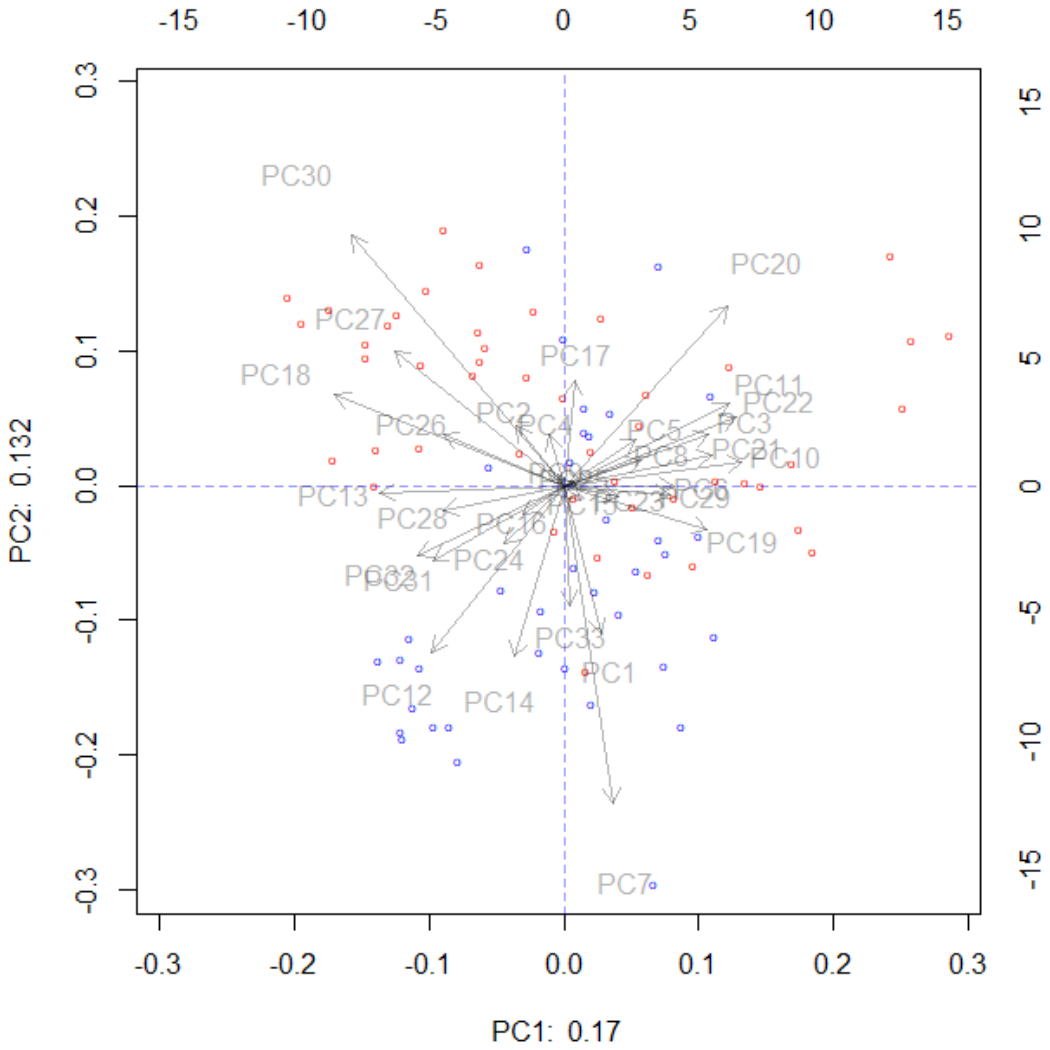

Figure 25. Compositional biplot of urinary microbiota of healthy male donors (blue) and male stone patients (red)

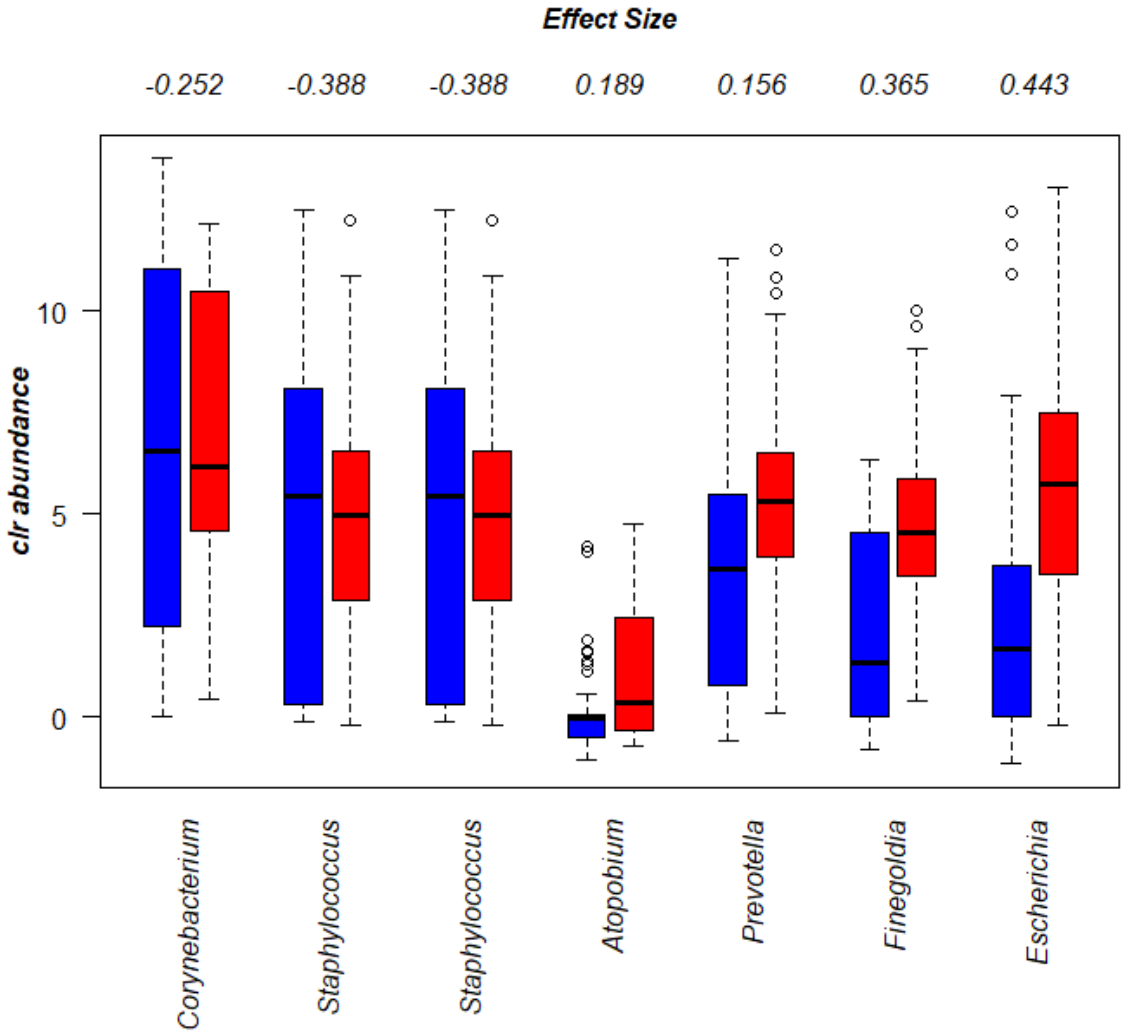

Figure 26. Box plot of the difference in bacterial abundance between healthy male individuals and male urine stone patients. Taxa with high effect sizes are shown. The bacterial abundances were presented in relative abundance as computed by ALDEx2.

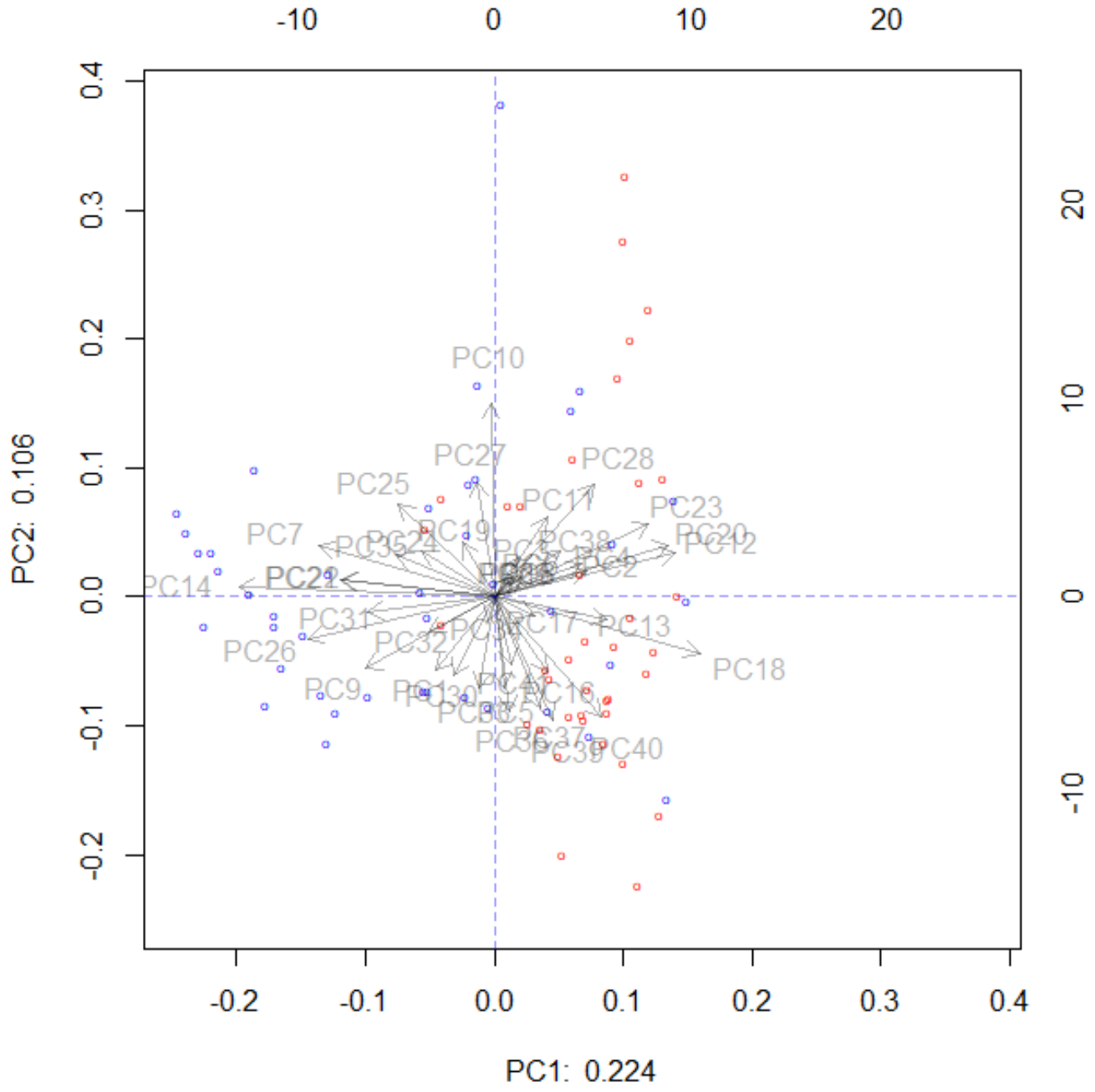

Figure 27. Compositional biplot of urinary microbiota of healthy male donors (blue) and male stone patients (red)

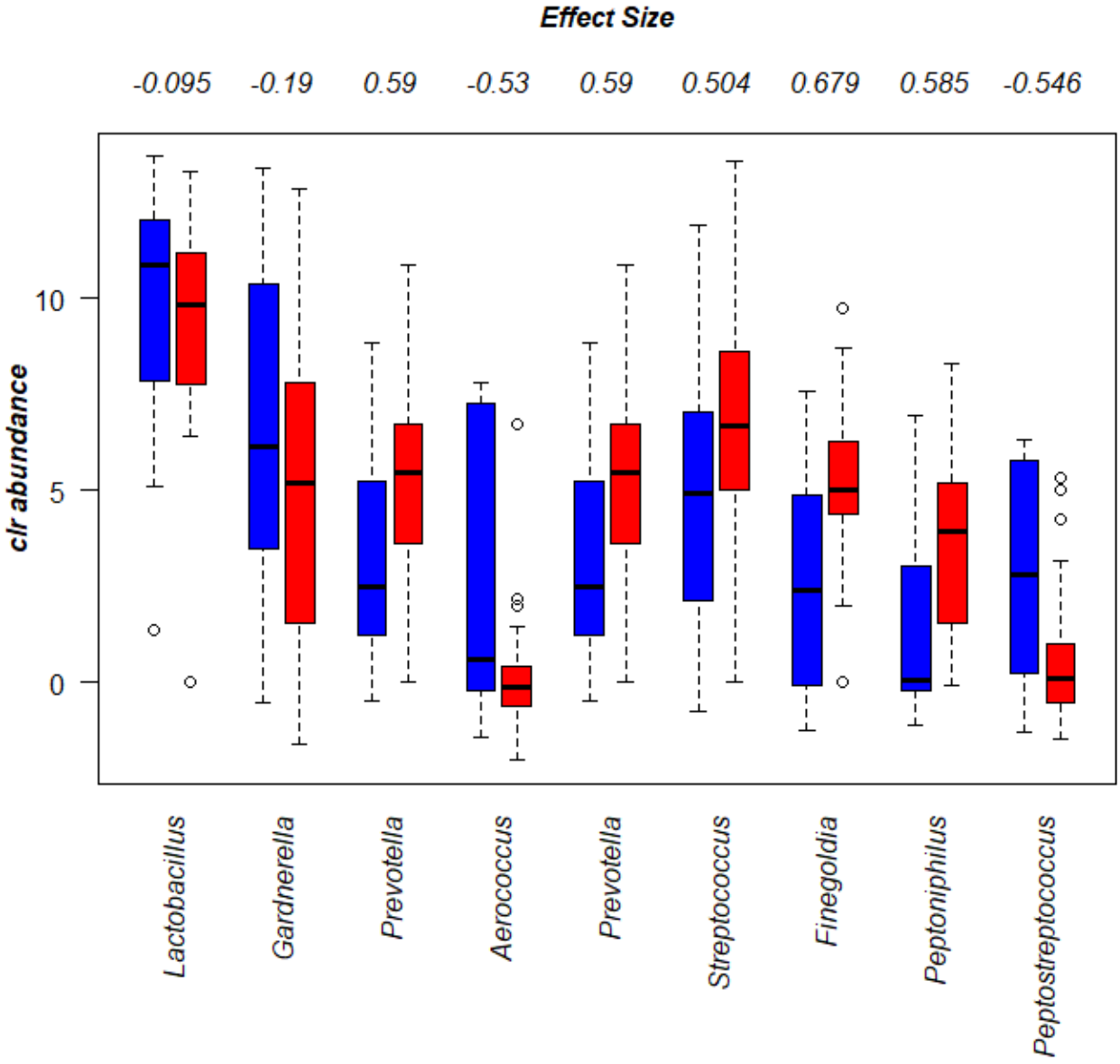

Figure 28. Box plot of the difference in bacterial abundance between healthy female individuals and female urine stone patients. Taxa with high effect sizes are shown. The bacterial abundances were presented in relative abundance as computed by ALDEx2.

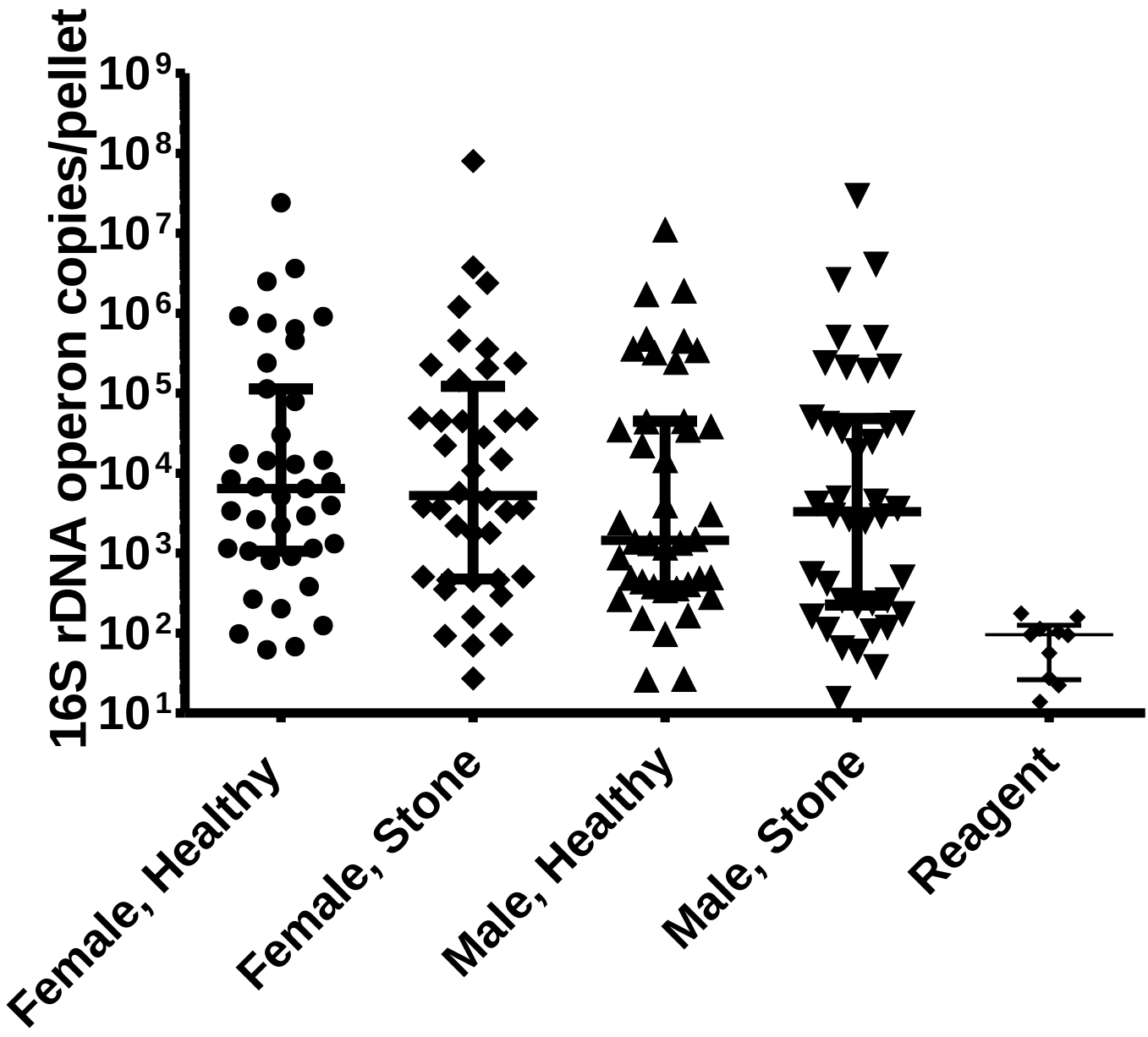

Figure 29. qPCR estimation of bacterial load in urine pellet samples. The result was reported as the estimated number of operon copies of 16S rRNA present per urine pellet.

Figure 30. Weighted UniFrac distance between samples from one donor (black dots) and between samples of different donors (white squares). For clearer presentation, each white square represents the mean weighted UniFrac distance of the first sample from one donor to the samples of all other donors.

or female, stone or healthy), samples from one donor collected from two time points had significantly lower UniFrac scores than samples from different donors, which suggested the variation of the urinary microbiota of one individual is always smaller than the difference of urinary microbiota profiles of two individuals. These results suggested that the urinary microbiota of one individual stays relatively stable throughout time.

3.3.2.3 Mutual support culture-based methods and 16S rRNA sequencing in microbial detection

Among all 116 urine samples collected, 14 had higher than 10^5 CFU/mL, and the participants were excluded. Of the remaining 102 samples, 76 had positive cultures. Among them, 43 samples were negative for traditional culture but had positive EQUC culture results. Almost all cultured species had corresponding OTUs in the sequencing results, and usually, they were the most predominant or second predominant genus in the sample. Also, there were two cases of interest: in one sample, a colony was retrieved, and Sanger sequencing confirmed it to be *Weeksella virosa*. Interestingly, the 16S rDNA sequencing result of this sample the reads of the OTU attributed to *Weeksella* only accounted for only 1.32% of the total reads, while *Weeksella virosa* is the only microbial species that was retrieved by culture-based methods. On the other hand, in another sample, 32.32% of reads were unusually attributed to *Ureaplasma* species, while routine culture methods failed to grow any *Ureaplasma.* However, when *Ureaplasma*-specific media and atmosphere (refer to methods section) were utilized, it was possible to culture *Ureaplasma ureealyticum.* These results showed that the combination of culture-based techniques and sequencing-based techniques could yield the highest sensitivity and accuracy in microbial detection in clinical samples.

3.4 Discussion

In this study, we first sought to optimise the methodology for urine sample storage to obtain the highest fidelity in 16S rRNA gene sequencing; then we explored the urinary

microbiota profiles of healthy individuals with culture-based and sequencing-based techniques. The urinary microbiota of an individual maintained its profile features throughout the multiple sampling points, making it distinct from the urinary microbiota from other individuals.

The value of proper sample collection and handling methods can never be under-emphasised in microbiota studies. Each sample type has its unique features that must be taken into account for sample acquisition and processing. Common sample type for human microbiota studies such as stool, saliva and vaginal fluids have well-established handling methods (Abrahamson *et al.*, 2015; Goodrich *et al.*, 2014; Kumar *et al.*, 2014; Ott *et al.*, 2004). However, no consensus has yet been reached on the optimal protocol for urine sample collection and storage. Unlike stool or saliva, urine samples have their own features relating to how they perform when processed for microbiota analysis: (a) Low microbial biomass in urine sample means it tends to be affected by reagent and environmental contamination (see chapter two) and (b) The liquid form of a urine sample makes it more susceptible to the freeze-thaw process, as explained below.

Since the microbial biomass in a urine sample from a healthy subject is low, a relatively large amount of urine is required to have a decent representation of the urinary microbiota. In this study, as the volume of the urine sample increased, a higher UniFrac distance score was achieved to reagent controls, suggesting a better representation of the original urinary microbiota. The UniFrac score was also noted to reach its “plateau phase” after volume reaches 20 mL, suggesting that this volume may be the minimal appropriate for urinary microbiota sampling. Several previously published studies investigating urinary microbiota would use a low urine volume such as 1 ml, 2 ml or 10 ml (Lewis *et al.*, 2013; Nelson *et al.*, 2010; Pearce *et al.*, 2014). Their sequencing quality may, therefore, be compromised.

Since the DNA extraction of urine samples includes a step of centrifuging the urine sample and discarding the supernatant, all DNA released into that urine would be lost. The freeze-thaw process may lead to microbial cellular rupture, releasing their DNA into the urine and loss following centrifugation. Common cryoprotectants in traditional

microbiology experiments such as glycerol do not have a DNA-free form that is compatible with microbiota investigation. Therefore, when the whole urine is frozen in -70℃ without any cryoprotectants, its quality will have a remarkable drop. Keeping urine in pellets avoids discarding the supernatant, thus transferring all microbial components into the DNA extraction experiment.

To my knowledge, this is the first report of the optimal storage form of urine samples. As the technical progress of high-throughput sequencing, many study groups and commercial companies are planning on sequencing-based studies of urine samples. Therefore, developing a standard protocol for sample handling and storage is crucial. The data generated here may provide some reference.

The present study observed different urinary microbiota features between male and females but did not identify a "stone-related" microbiota, as no clear distinction from healthy individual and stone patients of the same gender could be found. One possibility is that there's no such "stone-related" microbiota, especially in patients with no acute infective episodes. Another possibility is the inclusion criteria did not specify frequently recurrent stone patients, or patients with a specific stone composition, which may be more likely to have a specific urinary microbiota (Maalouf *et al.*, 2012). Alternatively, the impedance of urine through one ureter may not be detected by used here, or if only small numbers of bacteria are released this may also be below the levels of detection. Barr-Beare *et al.* explored the stone microbiota with both sequencing-based methods and EQUC and found several known uropathogens including *E.coli*, *Pseudomonas*, *Gardnerella et al.* in the stone (Barr-Beare *et al.*, 2015), which suggests microbes do inhabit non-infectious, calcium-based stones. However, other studies also report that urine samples have lower sensitivity to detect stone-harboured microbes compared to stone samples, which may explain the inability to identify a stone-specific microbiota in the urine samples tested here.

The urinary microbiota was found to have its longitudinal stability throughout three months. Urine samples from one individual in different time points were always more similar than samples from different individuals. This phenomenon persisted in both

genders, and in both healthy, and stone donors. It is also worth noting that the urinary microbiota maintained its stability in stone patients, even after their stone had been surgically removed. This could also suggest that calculi per se are not major factors of urinary microbiota profiles, or the time gap of three months was too long to identify the perioperative disruption of the urinary microbiota. This will be further investigated this in the next chapter.

To my knowledge, this is the first study reporting the longitudinal changes of the urinary microbiota. The understanding of the relative resilience of urinary microbiota is crucial when further studying the effect of antibiotics, surgical intervention and stone fragmentation to the urinary microbiota, as explored in the next chapter.

In the last series of studies in the present chapter, culture-based methods were compared to sequencing-based techniques. In most cases, the results overlapped, in the sense that the culture-retrieved species were invariably identified among the five most predominant OTUs obtained by sequencing. However, exceptions did exist as the case of *Weeksella* species which only consisted of 1.32% of the total reads. This demonstrated the sensitivity differences of culture-based methods and sequencing-based techniques. The former, although unable to culture all species, was sensitive enough to recover even one single microbial cell, at least in theory. While sequencing-based methods tended to lose their detection accuracy for microbes at very low abundance since the reagent noise was at 103 copy/ml. Therefore, if the microbe of interest can be cultivated, the culture-based methods have higher sensitivity than sequencing-based techniques in microbial detection. On the other hand, there's no single culture method that could cover all spectrum of known microbes, making it time-consuming work. Therefore, sequencing techniques can act as a "guide" for specific culture methods to be employed to recover microbes of interest.

In conclusion, urine samples for microbiota studies should be stored in centrifuged pellet form. The urinary microbiota has gender-specific features, and have relative resilience in its microbial structure over time. No specific microbiota profiles were identified in this study.

3.5 References

Abrahamson, M., Hooker, E., Ajami, N. J., Petrosino, J. F., & Orwoll, E. S. (2017). Successful collection of stool samples for microbiome analyses from a large community-based population of elderly men. *Contemporary clinical trials communications*, 7, 158-162.

Bao, Y., Al, K. F., Chanyi, R. M., Whiteside, S., Dewar, M., Razvi, H., Reid, G., et al. (2017). Questions and challenges associated with studying the microbiome of the urinary tract. *Annals of translational medicine*, 5(2).

Barr-Beare, E., Saxena, V., Hilt, E. E., Thomas-White, K., Schober, M., Li, B., Becknell, B., et al. (2015). The interaction between Enterobacteriaceae and calcium oxalate deposits. *PloS one*, 10(10), e0139575.

Di Bella, J. M., Bao, Y., Gloor, G. B., Burton, J. P. & Reid, G. (2013). High throughput sequencing methods and analysis for microbiome research. *Journal of microbiological methods*, 95(3), 401–414.

Bratell, S., Brorson, J., Grenabo, L., Hedelin, H. & Pettersson, S. (1990). The bacteriology of operated renal stones. *European urology*, 17, 58–61.

Burton, J., Dixon, J. & Reid, G. (2003). Detection of Bifidobacterium species and Gardnerella vaginalis in the vagina using PCR and denaturing gradient gel electrophoresis (DGGE). *International Journal of Gynecology & Obstetrics*, 81(1), 61–63.

Choo, J. M., Leong, L. E. & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific reports*, 5.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A., et al. (2008). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic acids research*, 37(suppl_1), D141–D145.

Consortium, H. M. P. & others. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207.

Dewan, B., Sharma, M., Nayak, N. & Sharma, S. (1997). Upper urinary tract stones \& Ureaplasma urealyticum. *The Indian journal of medical research*, 105, 15–21.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461.

Edgar RC, Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* 31: 3476–3482.

Gault, M., Longerich, L., Crane, G., Cooper, R., Dow, D., Best, L., Stockall, E., et al. (1995). Bacteriology of urinary tract stones. *The Journal of urology*, 153(4), 1164–1170.

Gloor, G. B., Hummelen, R., Macklaim, J. M., Dickson, R. J., Fernandes, A. D., MacPhee, R., Reid, G., et al. (2010). Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PloS one*, 5(10), e15406–e15406.

Goodrich, J. K., Di Rienzi, S. C., Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., Knight, R., et al. (2014). Conducting a microbiome study. *Cell*, 158(2), 250–262.

Hancock, V., Ferrières, L. & Klemm, P. (2007). Biofilm formation by asymptomatic and virulent urinary tract infectious Escherichia coli strains. *FEMS microbiology letters*, 267(1), 30–7.

Hilt, E. E., McKinley, K., Pearce, M. M., Rosenfeld, A. B., Zilliox, M. J., Mueller, E. R., Brubaker, L., et al. (2014). Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *Journal of clinical microbiology*, 52(3), 871–876.

Hugosson, J., Grenabo, L., Hedelin, H., Pettersson, S. & Seeberg, S. (1990). Bacteriology of upper urinary tract stones. *The Journal of urology*, 143(5), 965–968.

Kumar, R., Eipers, P., Little, R. B., Crowley, M., Crossman, D. K., Lefkowitz, E. J. & Morrow, C. D. (2014). Getting started with microbiome analysis: sample acquisition to bioinformatics. *Current Protocols in Human Genetics*, 18–8.

Larsen, E. H., Gasser, T. & Madsen, P. (1986). Antimicrobial prophylaxis in urologic surgery. *The Urologic clinics of North America*, 13(4), 591–604.

Lewi, H., White, A., Hutchinson, A. & Scott, R. (1984). The bacteriology of the urine and renal calculi. *Urological research*, 12(2), 107–109.

Lewis, D. A., Brown, R., Williams, J., White, P., Jacobson, S. K., Marchesi, J. R. & Drake, M. J. (2013). The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. *Frontiers in cellular and infection microbiology*, 3.

Maalouf, N. M. (2012). Approach to the adult kidney stone former. *Clinical reviews in bone and mineral metabolism*, 10(1), 38–49.

Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. 2012. PANDAseq: paired-end assembler for Illumina sequences. *BMC Bioinfor- matics* 13:31.

McCartney, A. C., Clark, J. & Lewi, H. J. (1985). Bacteriological study of renal calculi. *European journal of clinical microbiology*, 4(6), 553–5.

Nelson, D. E., Van Der Pol, B., Dong, Q., Revanna, K. V., Fan, B., Easwaran, S., Sodergren, E., et al. (2010). Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. *PloS one*, 5(11), e14116.

Ott, S. J., Musfeldt, M., Timmis, K. N., Hampe, J., Wenderoth, D. F. & Schreiber, S. (2004). In vitro alterations of intestinal bacterial microbiome in fecal samples during storage. *Diagnostic microbiology and infectious disease*, 50(4), 237–245.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science*, 276(5313), 734–740.

Pearce, M. M., Hilt, E. E., Rosenfeld, A. B., Zilliox, M. J., Thomas-White, K., Fok, C., ... & Wolfe, A. J. (2014). The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *MBio*, 5(4), e01283–14.

Rosen, D. A., Hooton, T. M., Stamm, W. E., Humphrey, P. A. & Hultgren, S. J. (2007). Detection of intracellular bacterial communities in human urinary tract infection. *PLoS* Med, 4(12), e329.

Salo, J., Sevander, J., Tapiainen, T., Ikäheimo, I., Pokka, T., Koskela, M. & Uhari, M. (2009). Biofilm formation by Escherichia coli isolated from patients with urinary tract infections. *Clinical nephrology*, 71(5), 501–507.

Scales, C. D., Smith, A. C., Hanley, J. M., Saigal, C. S., & Urologic Diseases in America Project. (2012). Prevalence of kidney stones in the United States. *European urology*, 62(1), 160–165.

Siddiqui, H., Nederbragt, A. J., Lagesen, K., Jeansson, S. L. & Jakobsen, K. S. (2011). Assessing diversity of the female urine microbiome by high throughput sequencing of 16S rDNA amplicons. *BMC microbiology*, 11(1), 244.

Chapter 4

4 Perioperative urinary microbiota and its association with infectious complications following extracorporeal shock wave lithotripsy of metabolic stones.

4.1 Introduction

Urolithiasis, also known as urinary stones, is among the most common urological conditions (Scales et al., 2012). In the United States, urinary stones take up about two million outpatient visits annually. The incidence seems to be on the rise, in the United States, from 3.8% in 1976-1980 to 8.4% in 2007-2010 (Scales *et al.*, 2012; Stamatelou *et al.*, 2003). The condition has a detrimental effect on the kidney, with obstruction compromising renal functions, which can lead to renal failures. Studies have shown that 28% of staghorn calculi (a large stone that takes up more than one calyx) patients will develop renal function deterioration in eight years (Teichman *et al.*, 1995). Urinary stones can also induce infections due to urinary stasis. This can develop into lethal urosepsis (O'Keeffe *et al.*, 1993).

Bacterial infections have themselves been associated with certain types of urinary stones. These are usually called "infections stones or struvite stones". These are thought to be caused by urease-producing bacteria (Naber *et al.*, 2011), although this has recently been challenged because these microorganisms are not always recovered from the stone (Yoshida *et al.*, 1984).

The chemical composition of infectious stones can comprise mono-ammonium urate, struvite (magnesium ammonium phosphate) and carbonate apatite. Common urease-producing bacteria causing infection stones include *E.coli, Proteus* species, *Klebsiella* species, *Staphylococcus* species, and *Pseudomonas* species, *etc*. (Bichler *et al.*, 2002). Other stone types, for example, those composed of calcium oxalate are thus considered "non-infective stones". They are believed to form without microbial involvement. However, despite meticulous preoperative screening and prophylaxis, infective

complications still occur frequently, following lithotripsy surgery (Borghi *et al.*, 2012; Golechha *et al.*, 2001).

Common surgical options for urinary stone management include extracorporeal shock wave lithotripsy (ESWL), percutaneous nephrolithotomy (PCNL) and endoscopic procedures.

Firstly introduced in the early 1980s (Chaussy *et al.*, 1984), ESWL has revolutionised the treatment of urinary stones. Its ease of use, non-invasive nature and high efficiency have rapidly won global acceptance. Despite these advantages, complications can occur in three categories: stone fragment-associated complications, infection-associated complications and tissue damage (Table 8). Infectious complications include bacteriuria, UTI and sepsis. For the more serious complications such as sepsis, bacteria in urine enter the bloodstream facilitated by renal trauma and vascular disruption induced by the ESWL. This is especially the case for infectious stones, where bacteriuria can be detected through postoperative urine culture (Michaels *et al.*, 1988). However, even in non-infection stone patients, postoperative bacteriuria still happens in 5.8% of cases (Mira Moreno *et al.*, 2014). In 1997, Pearle et al. performed a meta-analysis performed on eight randomised controlled trials (RCTs) investigating the effect of antibiotic prophylaxis on ESWL patients, concluded that antibiotic prophylaxis is beneficial and reduces the post-ESWL complication rate (Pearle & Roehrborn, 1997). This led to the American Urological Association (AUA) recommending antibiotic prophylaxis for ESWL until 2012 when new evidence questioned the efficacy (Honey *et al.*, 2013; Lu et al., 2012; Wiesenthal *et al.*, 2011). Today, both the AUA and European Association of Urology (EAU) guidelines do not recommend antibiotic prophylaxis, unless the patient has an increased risk of infection.

Table 8. Common complications following ESWL and PCNL (Skolarikos et al., 2006)

Complications after ESWL for urinary stones
Related to stone fragments
• Incomplete fragmentation
• Residual stone fragments
• Steinstrasse
• Obstruction
Infections
• Asymptomatic bacteriuria • Urinary tract infection • Sepsis
Tissue effects
• Renal (hematoma, hemorrhage)
• Cardiovascular effects
• Gastrointestinal effects
• Genital system
• Foetus
Complications after PCNL for urinary stones
Acute
• Damage to vascular endothelium
• Damage to nephron, renal tubules and interstitium
• Loss of corticomedullary demarcation
• Increased extraction in urine of metabolites indicating renal damage
• Haematuria
• Hematoma
• Decrease in GFR
• Decrease in effective RPF
Chronic
• New onset of hypertension
• Perirenal fibrosis
• Loss of renal function

Compared to ESWL, PCNL is a surgical procedure required to remove larger stones or those refractory to non-invasive procedures. Infective complications of PCNL are more common, in that one-third of cases have an infection-related episode (Cadeddu *et al.*, 1998). Despite the fact that only struvite stones are currently considered as having a bacterial nidus, bacteria have been detected in other common stone types, such as calcium oxalate and calcium phosphate stones (Bratell *et al.*, 1990; Dewan *et al.*, 1997; Gault et al., 1995; Hugosson *et al.*, 1990; Larsen *et al.*, 1986; Lewi *et al.*, 1984; McCartney *et al.*, 1985). During lithotripsy, stone fragmentation may potentially release bacteria into the bloodstream, increasing the risk of urosepsis. Therefore, evaluating the risk of postoperative UTI especially urosepsis is of importance. In routine practice, a urine culture is usually performed one week prior to surgery to exclude active infection, and antibiotic prophylaxis (commonly fluoroquinolones) is applied if positive. Patients do not typically undergo their surgery until the urine culture is clear of any evidence of infection, as the EAU guidelines recommend (Türk *et al.*, 2017). Despite these efforts, life-threatening systemic infections can occur after lithotripsy surgery. This raises the question whether there are bacteria present in non-infectious stones that contribute to postoperative bacteriuria and other infective complications.

In this study, we hypothesise that bacteria exist within non-infectious stones, such as those composed of calcium oxalate and phosphate are released into urine during lithotripsy surgeries. We sought out to determine whether their 16S rRNA gene sequencing and enhanced culture from urine and stone samples would be able to detect these organisms.

4.2 Materials and methods

4.2.1 Clinical samples and study design

The protocol of this study was approved by the human research ethics board, Western University, Canada and West China Hospital, Sichuan University, China. Participants in this study were patients admitted to St. Joseph Hospital or West China Hospital. Inclusion criteria were that the participants: at least 18 years of age, non-pregnant, no other infectious diseases, not receiving antibiotic or chemotherapy. The diagnosis of

urinary stone was confirmed with the radiological investigation, and the patient then received either ESWL or PCNL. Patient information including gender, age, comorbidities, size of the stone along with number, location and composition of the stones were recorded. After inclusion, participants in the ESWL group were asked to provide three urine samples: one in the operating room before ESWL, plus one collected as the first voided urine after ESWL, and one at the follow-up outpatient clinic. Participants of the PCNL group provided preoperative voided urine 1-3 days before the surgery, and a postoperative urine collected via Foley catheter within 24 hours of surgery. A renal pelvic urine sample and a stone sample were also collected during the PCNL surgery. Complications were noted from medical records. All patient information was maintained by a research coordinator, and were blind to other members of the research group until all sample processing, and DNA sequencing work was completed. Patients were de-identified after retrieval of their demographic information for statistical analysis.

4.2.2 Sample collection

4.2.2.1 Voided urine

All participants were instructed to wash their hands thoroughly with soap and water, then to clean and wash their external genitalia with an antiseptic towelette. Urine samples were taken with a clean-catch, midstream voided fashion: the first 10 mL was discarded, the majority of the voided volume collected in a sterile container, and the final voided volume dispensed into the toilet. Once collected, all urine samples were processed within one hour. Microbiological cultures were performed subsequently, and the remaining of urine samples were spun down to pellets and stored at -80 ℃.

4.2.2.2 Renal pelvic urine samples

During PCNL surgeries, after the ureteral catheter were inserted the patient and set in prone position. At this point, after sterile preparation, the percutaneous access was established with a 14-gauge Kellet needle. When the needle tip reached the renal collection system, urine was emitted from the rear end of the needle, and collected as the renal pelvic urine sample.

4.2.2.3 Stone samples

Urinary stone fragments were collected during PCNL, and transferred by the surgeon or nurse to a sterile container using aseptic techniques. Stone samples were then transferred to the lab, where the procedures for cultural analysis was performed within 4 hours. In brief, the surface of the stone was washed five times with sequential falcon tubes containing 75% ethanol in the first one and DNA-free water in the remaining four tubes. The stone fragments were then transferred to a sterile mortar, crushed, then aliquoted for traditional or EQUC culture analysis and 16S rRNA gene sequencing.

4.2.3 Microbiological culture

All collected samples were quantitatively cultured on self-made Colombia Blood Agar with 5% sheep blood, LB agar and MacConkey agar. Urine samples with a CFU > 10^5 per mL were considered to represent an active UTI and excluded, as per standard protocol (Pickard *et al.*, 2016). All cultivated strains were isolated and stored. EQUC was also applied following protocols developed by Hilt et al. (Hilt *et al.*, 2014). In brief, urine samples were cultured with higher volume input (100 ul and up to 1 ml), enriched culture media (CBA, CAN agars and thioglycolate medium), both aerobic and anaerobic culture atmosphere and extended culture time (up to 5 days). Morphologically unique colony-types were re-isolated for taxonomic identification. The details of microbiological culture media used in this study are in the appendix.

4.2.4 DNA extraction

The samples were thawed on ice and DNA-extracted with MOBIO Powersoil® kit. The DNA extraction protocol was similar to ones used in Chapter three. In brief, lysis buffer and beads were added to urine or urine pellets. Lysis of cells was achieved with mechanical and chemical methods. The suspension was then transferred to a new tube, and the subsequent procedure was following manufacturer's instructions.

4.2.5 16S rRNA gene library generation and sequencing

Extracted DNA from the clinical samples were PCR-amplified with barcoded primers as described by Gloor *et al*. (2010) shown in Appendix A. The PCR was performed in a 42-

µl reaction system, with details listed in Appendix B. Following PCR, the amplicons were quantified with the Qubit 2.0 fluorometer (Invitrogen, CA, USA). Equimolar amounts of each amplicon were aliquoted into a single tube purified with the QIAquick purification kit (Qiagen, Mississauga, ON, Canada). Each product was then pooled and sequenced on the Illumina Mi-Seq platform in a paired-end run using a 600 cycle kit with a paired-end 220-bp run at the London Regional Genomics Center (Robarts Research Institute, London, ON, Canada), following standard operating procedures.

4.2.6 Sequence processing and taxonomic assignment

Description of the processing algorithm has been previously published (Gloor *et al.*, 2010). The SOP and related software are accessible at http://github.com/ggloor/miseq_bin. The sequences were processed following algorithm as stated in Appendix C. Paired-end reads were overlapped with Pandaseq V2.5 (Masella *et al.*, 2012), allowing no ambiguous positions in the overlapped reads. Reads were then demultiplexed and tagged with the name corresponding to the sample identifier and barcode. After demultiplexing, reads from all samples were merged into one single file and were clustered into individual sequence units (ISUs), which were then further grouped into operational taxonomic units (OTUs) with open reference OTU picking method, using Uclust algorithm of USEARCH version 7, which has a *de novo* chimera filtering step built into it (Edgar *et al.*, 2010; Edgar *et al.*, 2015). The identity threshold was 97%. Singleton OTUs and ones with abundance < 1% were discarded. Taxonomical assignment was performed with mothur 1.3.4 to annotate the OTU sequences against the Silva database (v119).

4.2.7 Bioinformatic analysis

Compositional data transforming was done using centred log ratios (clr) by custom R codes, linear sample independence, and normalization of read counts. K-means clustering of the data was performed using Euclidean distances on CLR-transformed data with a uniform prior of 0.5 added to each value before transformation. The ALDEx R package version 2 was used to compare the relative abundances of genera (fernandes et al., 2014). Values reported in the manuscript represent the expected values of 128 Dirichlet Monte-

Carlo instances of clr-transformed data. A value of zero indicated that organism abundance was equal to the geometric mean abundance. Thus, organisms more abundant than the mean would have positive values, and those less abundant than the mean would have negative values. Base 2 was used for the logarithm so that differences between values would represent fold changes. Significance was based on the Benjamini-Hochberg corrected P value of the Wilcoxon rank test (significance threshold $P < 0.1$) (Gloor *et al.*, 2017).

4.2.8 qPCR estimation of 16S rRNA gene copy numbers.

The bacterial load in the urine samples were estimated by qPCR quantification of the DNA extracts. The reaction system includes 10ul of PCR master mix, 5 ul of 800Nm primer stock (V6 hypervariable region of 16S rRNA gene: V6-LT 5' - CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNCWACGCGARGAACCTTACC3'

and V6-RT: 5'CCTCTCTATGGGCAGTCGGTGATACRACACGAGCTGACGAC 3') Apprill *et al.,* 2015), 5ul of sample DNA template. A standard curve was constructed using the values returned using the above methodology with known copy numbers of E.coli DH5a gDNA.

4.3 Results

4.3.1 Characterization of urinary microbiota of urinary stone patients receiving ESWL

A total of 57 urine samples from 21 urinary stone patients receiving ESWL were collected. The median age of participants was 53. After exclusion of urine samples that had positive preoperative routine urine culture results, sequencing results of 51 urine samples from 12 male and five female stone patients went through downstream bioinformatic analysis. A total of 414 OTUs were identified from the returned sequences. On average, each sample contained 1938 DNA sequencing reads.

The composition of the urinary microbiota appeared to be similar to those detailed in other chapters. The female urinary microbiota had a comparatively higher alpha diversity

than males and was commonly predominated by *Lactobacillus* species. The male urinary microbiotas were usually predominated by a single genus, commonly *Corynebacterium*, *Staphylococcus* or *Streptococcus* (Figure 31). qPCR quantification of 16S RNA operon copy numbers showed that voided urine samples from female donors had a slightly higher copy number of rRNA operons than males, irrespective of healthy or stone status, but this was not statistically significant (Figure 32).

The sequencing results had considerable overlap with the EQUC results of the urine samples. A total of 24 isolated strains from nine genera were cultured, with the majority of which also confirmed in 16S rRNA gene sequencing (Figure 33). The overlap of sequencing and culture results further confirmed the evidence of viable bacteria inhabiting in the urinary tract of urinary stone patients.

4.3.2 Immediate postoperative urine showed signs of bacteria released into urine

The urinary microbiota profile immediately before and then after ESWL was undertaken with urine samples from 17 urinary stone patients receiving ESWL. The Shannon index was calculated as alpha diversity (species richness and evenness of a given sample) and weighted UniFrac phylogenetic distance metric as beta diversity (the difference of species composition between two samples) of immediate preoperative and immediate postoperative samples. The alpha diversity of immediate post-operative urine samples demonstrated remarkable change comparing to preoperative samples. The follow-up samples, on the other hand, had their Shannon diversity index score more similar to preoperative samples, making the difference of Shannon diversity index score between follow-up samples and preoperative samples significantly lower than the difference between immediate post-operative samples and preoperative samples (Figure 34).The beta diversity of urine samples before and after ESWL demonstrated similar fluctuations. The weighted Unifrac distances of samples from one donor at various time points were measured. The weighted UniFrac distances between immediate postoperative samples and preoperative samples were significantly higher than distances between follow-up samples and preoperative samples (Figure 35).

Of the immediate postoperative samples, there were OTUs that were not present in preoperative samples and reagent controls, or only constituted a small proportion in the preoperative samples. In the 16s rRNA gene sequencing results, 12 of 17 immediate postoperative samples had increased relative abundance of certain taxa, and in 7 of these 12, the microbial taxa with relative abundance change were retrieved using culture-based methods (routine urine culture and EQUC) (Table 9). This group of immediate postoperative urine samples that have "microbial fluctuation", also had higher weighted UniFrac distances to preoperative urine samples compared to the rest of the immediate postoperative urine samples (Figure 36). However, no significant difference in qPCR-estimated 16s rDNA operon copy number was observed. (Figure 37). In summary, these results suggest the release of microbes into urine during ESWL.

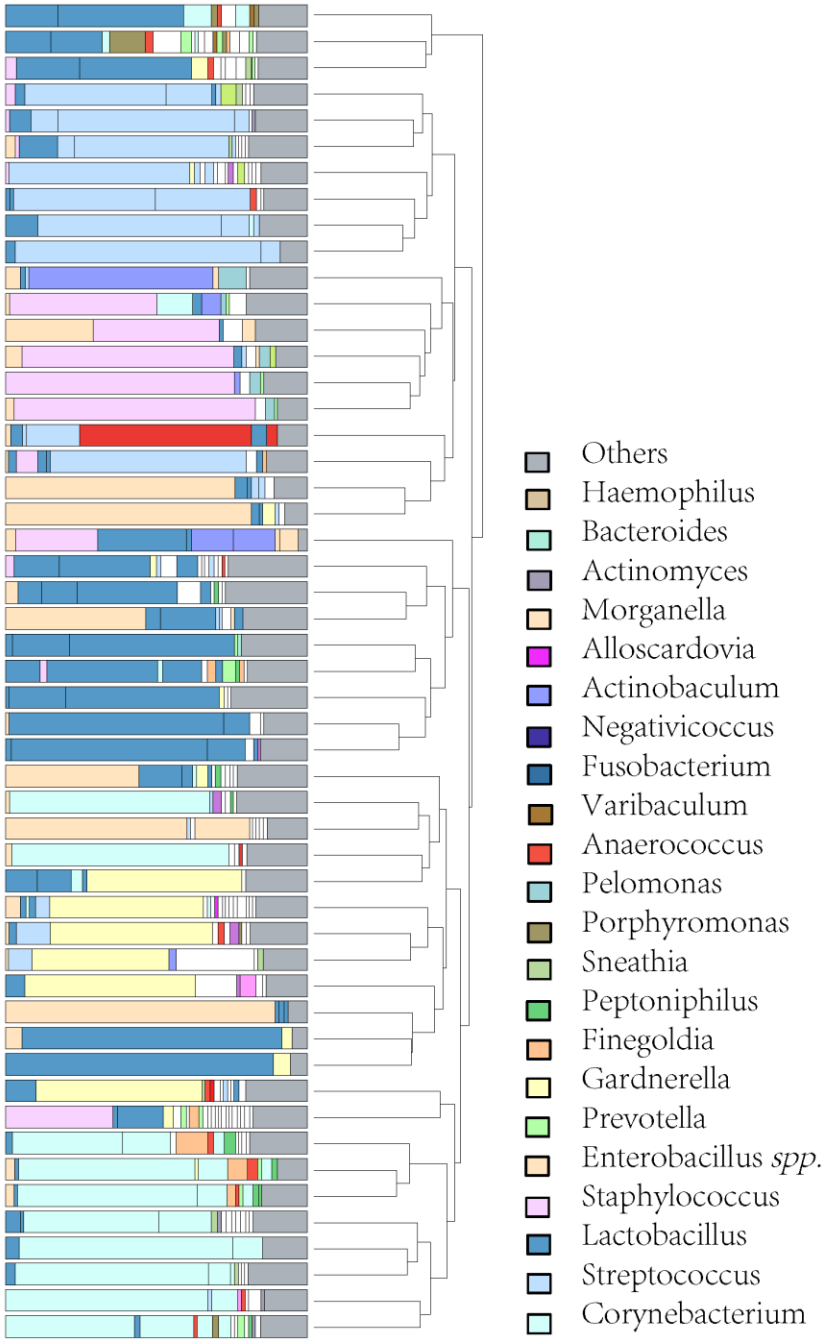

Figure 31. Relative abundance taxa bar plot of urine samples from urinary stone patients receiving extracorporeal shock wave lithotripsy (ESWL). Taxa (groups of 16s rRNA gene sequences) are represented at the genus level. The dendrogram represents the hierarchical average-linkage clustering of the microbiota profiles based on the computed weighted UniFrac distances.

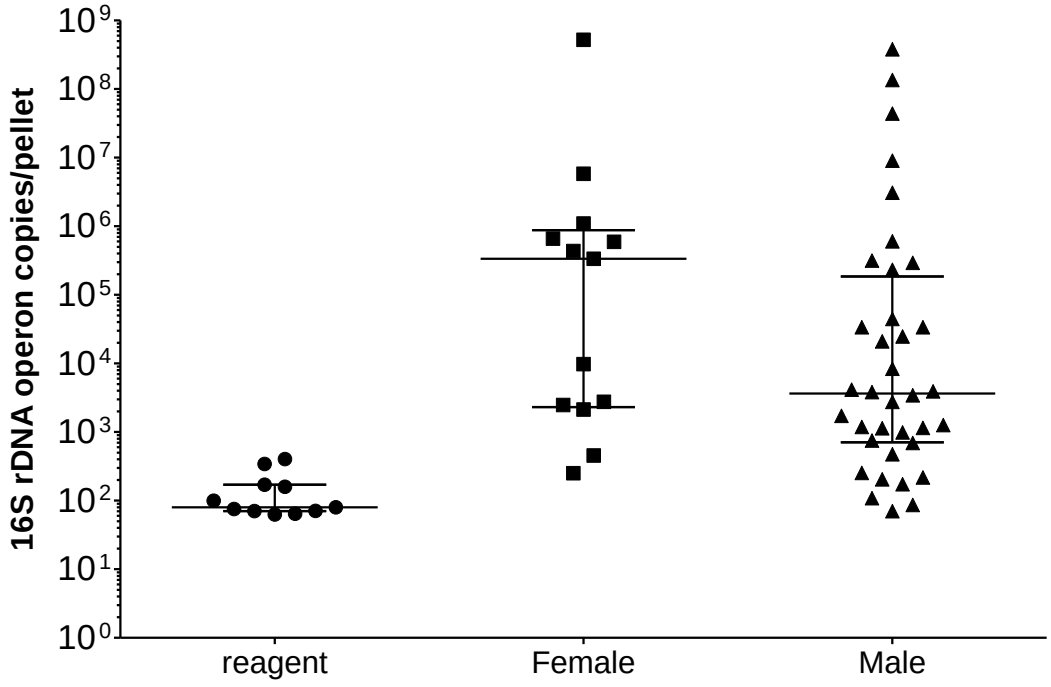

Figure 32. qPCR estimation of bacterial load in urine pellet samples. The result was reported as the estimated number of operon copies of 16S rRNA present per urine pellet. Bars in plot indicate median and interquartile range

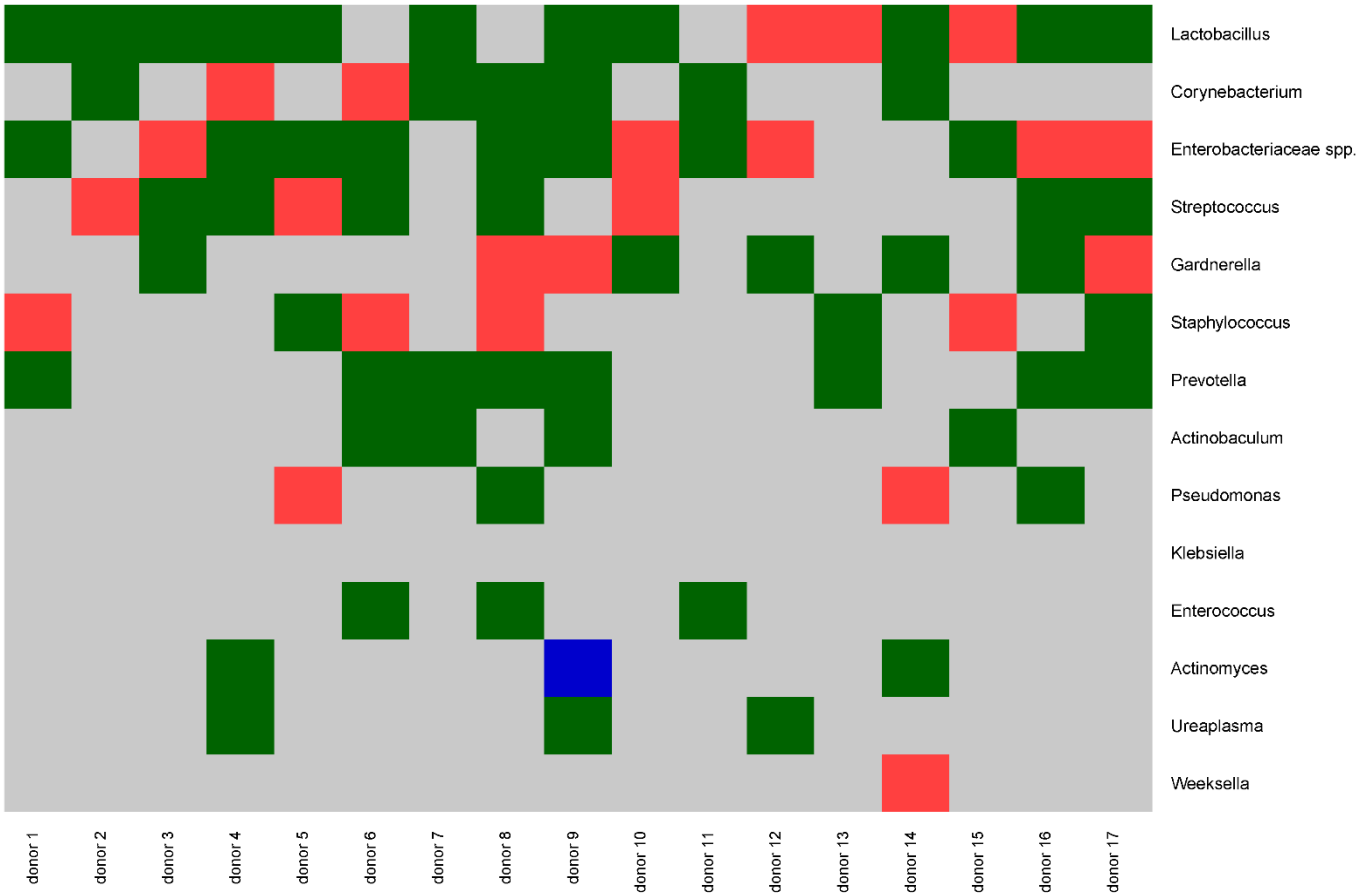

Figure 33. Heatmap comparing taxa detected by 16S rRNA gene sequencing and culture-based techniques. Each square was colour-coded based on the detection status of a given taxa. Green means detected by sequencing only, red means detected by both methods, blue means detected by culture only and grey means not detected by either method.

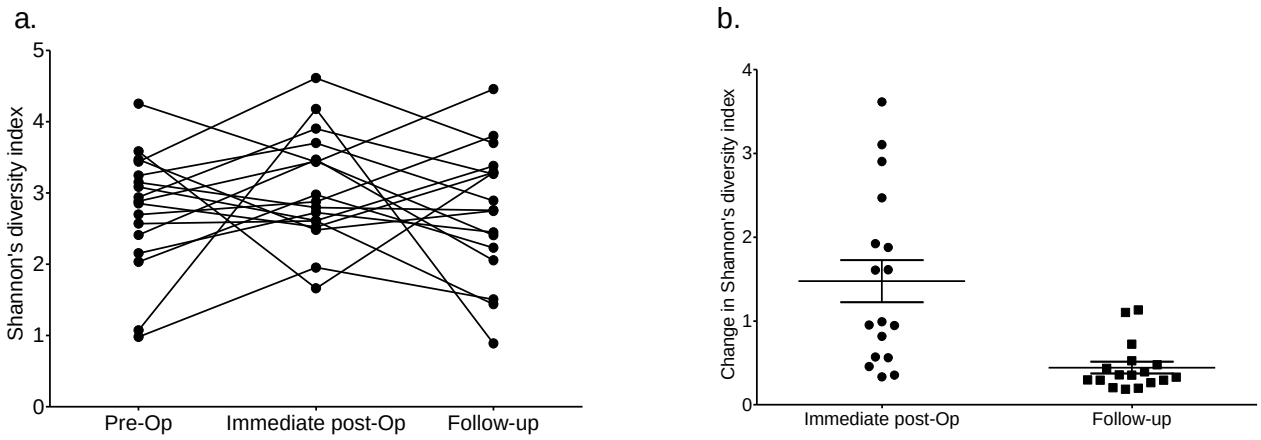

Figure 34. Shannon's diversity of urine samples from stone patient receiving ESWL. (a) Change in Shannon's diversity of urinary microbiota before and after ESWL. Samples from one donor were connected by lines. (b) Dot plot shows differences of alpha diversity from immediate post-operative samples and follow-up samples to pre-operative samples.

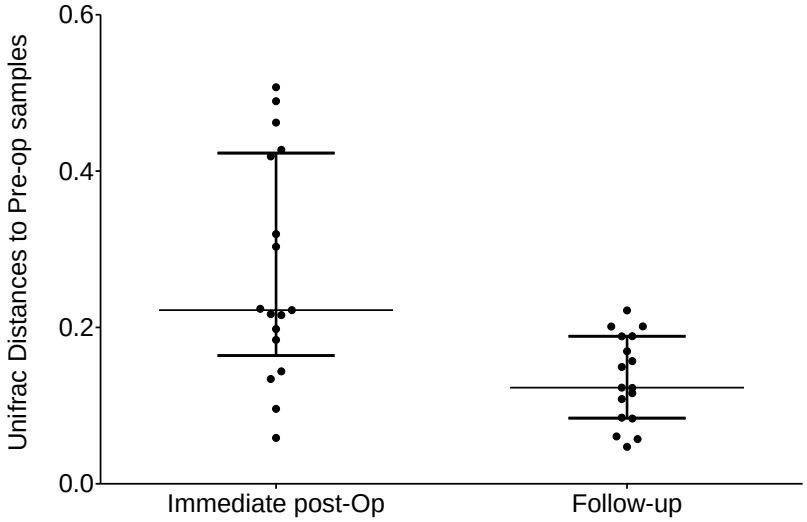

Figure 35. Dot plot showing weighted UniFrac distances index between preoperative samples to immediate postoperative samples and follow-up samples. Each dot stands for the weighted UniFrac distances index from the current condition (immediate post-operative or follow-up) to the preoperative urine sample from the same donor.

Table 9. Changes in relative abundance of taxa in immediate postoperative urine samples.

Donor	Urotype	Postoperative sequencing fluctuation	Culture results
1	*Staphylococcus*	*Enterobacteriaceae spp.*	*Staphylococcus*
2	*Streptococcus*		*Streptococcus*
3	*Gardnerella*	*Enterobacteriaceae spp.，Lactobacillus*	*Enterobacteriaceae spp.*
4	*Corynebacterium*		*Corynebacterium*
5	*Streptococcus*	*Pseudomonas*	*Streptococcus, Pseudomonas*
6	*Corynebacterium*	*Staphylococcus*	*Corynebacterium, Staphylococcus*
7	*Corynebacterium*	*Enterobacteriaceae spp.*	
8	*Gardnerella*	*Staphylococcus*	*Gardnerella, Staphylococcus*
9	*Gardnerella*	*Actinobaculum*	*Gardnerella, Actinomyces*
10	*Streptococcus*	*Staphylococcus*	*Streptococcus, Enterobacteriaceae*
11	*Corynebacterium*		
12	*Lactobacillus*	*Enterobacteriaceae spp.*	*Enterobacteriaceae spp.，Lactobacillus*
13	*Lactobacillus*		*Lactobacillus*
14	*Lactobacillus*	*Porphyromonas*	*Pseudomonas, Weeksella*
15	*Lactobacillus*	*Staphylococcus, Actinobaculum*	*Staphylococcus, Lactobacillus*
16	*Corynebacterium*		*Enterobacteriaceae spp.*
17	*Lactobacillus*	*Staphylococcus*	*Enterobacteriaceae spp.,Gardnerella*

* In this table, “Urotype” shows the most predominant taxa in the 16s rRNA gene sequencing result, and “Postoperative sequencing fluctuation” shows taxa that have the highest degree of change in their relative abundance in the in the 16s rRNA gene sequencing result. “Culture results” showed strains isolated from routine urine culture and EQUC.

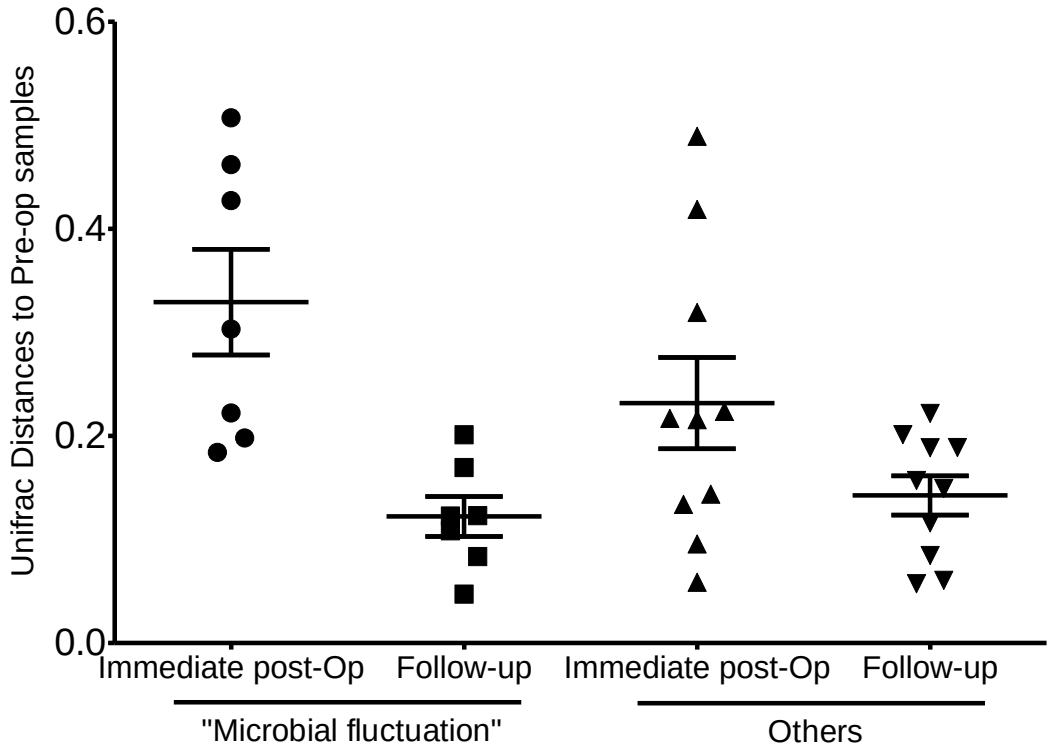

Figure 36. Comparison of weighted UniFrac distances of urine samples between groups that have "microbial fluctuation" in immediate postoperative samples. That is, taxa that have a remarkable increase in relative abundance and also confirmed in culture-based methods, and other samples. Each dot in the plot represents the weighted UniFrac distance of this sample to the preoperative sample of the same donor.

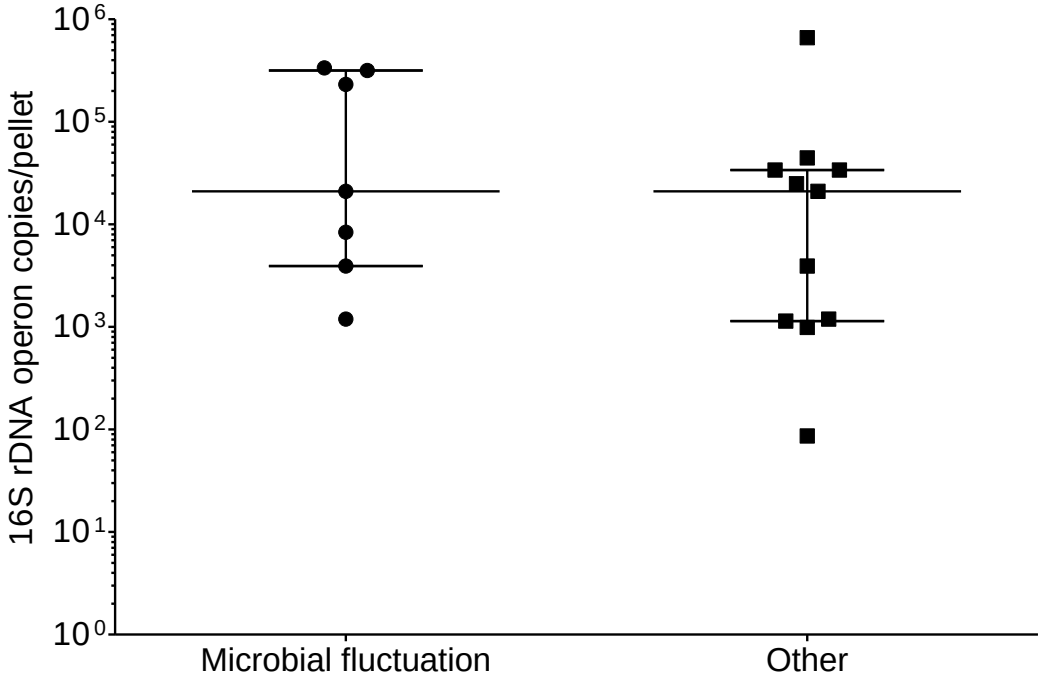

Figure 37. qPCR estimation of bacterial load in immediate postoperative urine samples. The "microbial fluctuation" groups stand for urine samples that have taxa with a remarkable increase in relative abundance and confirmed by culture-based methods. The result was reported as the estimated number of operon copies of 16S rRNA present per urine pellet.

4.3.3 Correlation of stone and urine microbiota profile further confirmed bacteria in stone

To further confirm that the post-ESWL detected bacteria cames from the stones, pelvic urine samples and stone samples were collected together with preoperative and postoperative voided urine sample from six patients receiving percutaneous nephrolithotomy (PCNL). Of these, the stone samples of two patients of six patients had positive 16S rRNA gene sequencing results (*i.e.*, a microbiota profile distinct from reagent control profiles).

The first patient was a 46-year-old male whose preoperative urine sample presented a *Corynebacterium*-predominated profile, while his stone sample had DNA sequences attributable to *Staphylococcus*. The pelvic urine samples and voided postoperative urine sample also had a high proportion of *Staphylococcus* in sequencing results which coincided with the stone sequencing results.

The second patient was a 52-year-old female whose preoperative voided urine sample was predominated by *Gardnerella* and *Lactobacillus*. The stone microbiota consisted mainly of *Actinobaculum* and *Enterobacteriaceae* spp, which are also major components of the urinary microbiota in the patient's pelvic urine and a postoperative voided urine sample (Figure 38).

These results suggest that it is highly likely that the newly-emerged bacteria in the postoperative urine may come from the stone during lithotripsy.

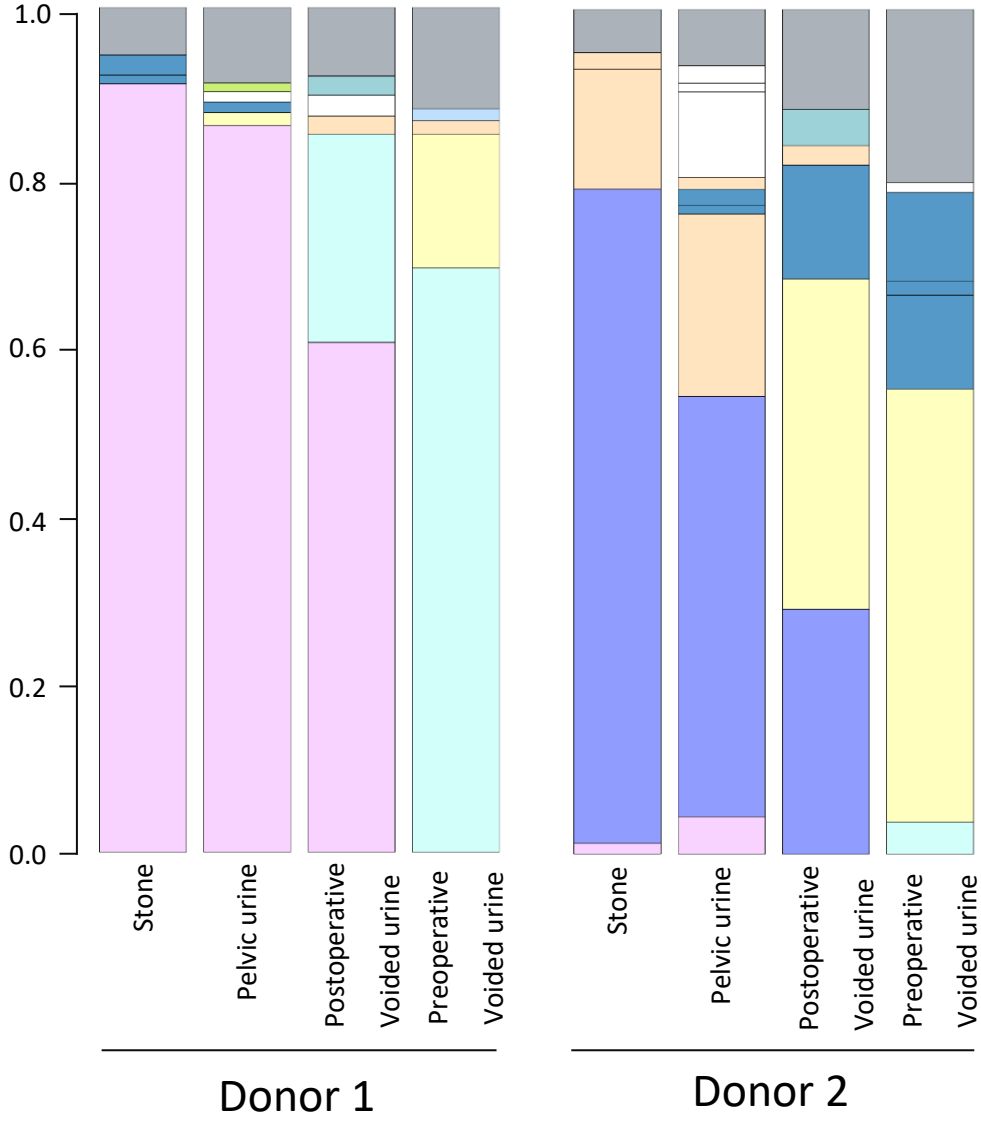

Figure 38. Relative abundance taxa bar plot show microbiota profiles from two PCNL patients kidney stones plus various urine collections.

4.4 Discussion

In this study, the urinary microbiota was noted to fluctuate following lithotripsy surgery, suggesting the origin of the bacteria was the stone itself.

Traditionally, urinary bacteria have not been implicated in calcium oxalate or calcium phosphate stones. However, postoperative infective complications do happen after lithotripsy surgeries, despite negative preoperative urine culture and antibiotic prophylaxis. Several studies now have reported that bacteria can be isolated from non-struvite urinary stone samples, even with routine urine culture methods (Mariappan *et al.*, 2005; Mufarrij *et al.*, 2012; Songra *et al.*, 2016). These reports suggest the existence of bacteria in "non-infective" stones. In our study, we observed a "microbial fluctuation" phenomenon in the immediate postoperative urine samples, that is, an increase in relative abundance of several microbial taxa that either did were not present before ESWL or were in low abundance.

When designing the study, attempts were made to eliminate confounding factors that could have an impact on the urinary microbiota. None of the patients in this study was receiving antibiotic prophylaxis during ESWL. Since ESWL uses externally applied focused shock wave to fragment the stone, no invasive instrumentation or procedures were taken during stone fragmentation in ESWL. This setting excluded most factors that affect the urinary microbiota. Although, theoretically bacteria could be liberated from epithelial cells during lithotripsy, the correlation of stone and urine sequencing results suggested that the "microbial fluctuation" likely resulted from the fragmentation of the stone.

It was interesting to note the correlation of microbiota profiles between stone samples, renal pelvic urine samples and immediate postoperative samples. The microbiota associated with stones and upper tract urine samples had different features than preoperative voided urine samples: the latter was invariably similar to the genitalia: *Lactobacillus*-predominant for female, and *Corynebacterium* or *Staphylococcus* for a

male. The stone samples and "microbial fluctuation" taxa in postoperative urine samples in our study, on the other hand, had *Staphylococcus*, *Actinobaculum*, *Pseudomonas* and *Enterobacteriaceae* spp. (presumtively *E.coli*) as the major taxa in sequencing results. Barr-Beare *et al*. (2015) sequenced stones from five paediatric patients and found *Enterobacteriaceae, Pseudomonas, Gardnerella*, *Bradyrhizobium*, *Phyllobacterium* and *Brucella* 16S rRNA gene homologies. Of note, many of these stone-associated microbes are known uropathogens (Kass *et al.*, 2002; Salem *et al.*, 2015; Tavassoli *et al.*, 2012; Woolfrey *et al.*, 1986), which again links UTI and urinary stone formation. Urinary stone patients are more likely to develop UTIs (Holmgren *et al.*, 1989; Hugosson *et al.*, 1989), and stone culture studies also showed 19-32% of calcium oxalate stones are bacteria positive (Tavichakorntrakool *et al.*, 2012; Wang *et al.*, 2014). Barr-Beare *et al.* (2015)also showed that bacteria and urinary stones have mutual facilitating effects on their pathogenesis progress in the formation of calcium oxalate crystals. The role of bacteria in the formation of calcium oxalate and other non-infective stones warrants further study.

Post-ESWL bacteriuria and bacteremia are not uncommon for patients with infection stones (Michaels *et al.*, 1988; Müller-Mattheis *et al.*, 1991). However, the incidence of postoperative bacteriuria for non-infection stone patients is as low as 5.8%, when using routine urine culture methods (Mira Moreno *et al.*, 2014). In our study, the postoperative "microbial fluctuation" phenomenon was identified from 70.6% of postoperative urine samples(12 out of 17), and both 16S rRNA gene sequencing and EQUC could detect this "microbial fluctuation" phenomenon in 41.2% of all immediate postoperative urine samples (7 out of 17 samples), suggesting a higher sensitivity than traditional culture methods. Admittedly, the "microbial fluctuation" phenomenon observed in this study is much milder than the traditional definition of bacteriuria, as it neither reached the required CFU threshold on routine urine culture nor caused any symptoms.

In summary, some patients with non-infection urinary stones appear to have bacteria released from their kidney stones when disrupted by ESWL. This has clinical ramifications if release of these organisms induces infection. The results should not be

taken as recommending antibiotic prophylaxis for all stone patients, as it is not clear that these drugs will prevent infection, or that their use is justified if the infection rate is low.

4.5 References

16s Illumina Amplicon Protocol. Retrieved from *http://press.igsb.anl.gov/earthmicrobiome/protocols-and-standards/16s/*

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, 75(2), 129-137.

Barr-Beare, E., Saxena, V., Hilt, E. E., Thomas-White, K., Schober, M., Li, B., Becknell, B., et al. (2015). The interaction between Enterobacteriaceae and calcium oxalate deposits. *PloS one*, 10(10), e0139575.

Bichler, K.-H., Eipper, E., Naber, K., Braun, V., Zimmermann, R. & Lahme, S. (2002). Urinary infection stones. *International journal of antimicrobial agents*, 19(6), 488–498.

Borghi, L., Nouvenne, A. & Meschi, T. (2012). Nephrolithiasis and urinary tract infections:“the chicken or the egg”dilemma? *Nephrology Dialysis Transplantation*, 27(11), 3982–3984.

Bratell, S., Brorson, J., Grenabo, L., Hedelin, H. & Pettersson, S. (1990). The bacteriology of operated renal stones. *European urology*, 17, 58–61.

Cadeddu, J. A., Chen, R., Bishoff, J., Micali, S., Kumar, A., Moore, R. G. & Kavoussi, L. R. (1998). Clinical significance of fever after percutaneous nephrolithotomy. *Urology*, 52(1), 48–50.

Chaussy, C., Schmiedt, E., Jocham, D., Schüller, J., Brandl, H. & Liedl, B. (1984). Extracorporeal shock-wave lithotripsy (ESWL) for treatment of urolithiasis. *Urology*, 23(5), 59–66.

Dewan, B., Sharma, M., Nayak, N. & Sharma, S. (1997). Upper urinary tract stones \& Ureaplasma urealyticum. *The Indian journal of medical research*, 105, 15–21.

Gault, M., Longerich, L., Crane, G., Cooper, R., Dow, D., Best, L., Stockall, E., et al. (1995). Bacteriology of urinary tract stones. *The Journal of urology*, 153(4), 1164–1170.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Frontiers in Microbiology*, 8, 2224.

Golechha, S., Solanki, A. & others. (2001). Bacteriology and chemical composition of renal calculi accompanying urinary tract infection. *Indian Journal of Urology*, 17(2), 111.

Hilt, E. E., McKinley, K., Pearce, M. M., Rosenfeld, A. B., Zilliox, M. J., Mueller, E. R., Brubaker, L., et al. (2014). Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *Journal of clinical microbiology*, 52(3), 871–876.

Holmgren, K., Danielson, B. G., Fellström, B., Ljunghall, S., Niklasson, F. & Wikström, B. (1989). The relation between urinary tract infections and stone composition in renal stone formers. *Scandinavian journal of urology and nephrology*, 23(2), 131–136.

Honey, R. J. D., Ordon, M., Ghiculete, D., Wiesenthal, J. D., Kodama, R. & Pace, K. T. (2013). A prospective study examining the incidence of bacteriuria and urinary tract infection after shock wave lithotripsy with targeted antibiotic prophylaxis. The Journal of urology, 189(6), 2112–2117.

Hugosson, J., Grenabo, L., Hedelin, H., Lincoln, K. & Pettersson, S. (1989). Chronic urinary tract infection and renal stones. *Scandinavian journal of urology and nephrology*, 23(1), 61–66.

Hugosson, J., Grenabo, L., Hedelin, H., Pettersson, S. & Seeberg, S. (1990). Bacteriology of upper urinary tract stones. *The Journal of urology*, 143(5), 965–968.

Kass, E. H. (2002). Asymptomatic infections of the urinary tract. *The Journal of urology*, 167(2), 1016–1020.

Larsen, E. H., Gasser, T. & Madsen, P. (1986). Antimicrobial prophylaxis in urologic surgery. *The Urologic clinics of North America*, 13(4), 591–604.

Lewi, H., White, A., Hutchinson, A. & Scott, R. (1984). The bacteriology of the urine and renal calculi. *Urological research*, 12(2), 107–109.

Lu, Y., Tianyong, F., Ping, H., Liangren, L., Haichao, Y. & Qiang, W. (2012). Antibiotic prophylaxis for shock wave lithotripsy in patients with sterile urine before treatment may be unnecessary: a systematic review and meta-analysis. *The Journal of urology*, 188(2), 441–448.

Mariappan, P., Smith, G., Bariol, S. V., Moussa, S. A. & Tolley, D. A. (2005). Stone and pelvic urine culture and sensitivity are better than bladder urine as predictors of urosepsis following percutaneous nephrolithotomy: a prospective clinical study. *The Journal of urology*, 173(5), 1610–1614.

McCartney, A. C., Clark, J. & Lewi, H. J. (1985). Bacteriological study of renal calculi. *European journal of clinical microbiology*, 4(6), 553–5.

Michaels, E. K., Fowler Jr, J. E. & Mariano, M. (1988). The Risk of Persistent Bacteriuria after Extracorporeal Shock Wave Lithotripsy of Infection Stones: A Prospective Study. *Shock Wave Lithotripsy* (pp. 201–206).

Mira Moreno, A., Montoya Lirola, M. D., García Tabar, P. J., Galiano Baena, J. F., Tenza Tenza, J. A. & Lobato Encinas, J. J. (2014). Incidence of infectious complications after extracorporeal shock wave lithotripsy in patients without associated risk factors. *The Journal of urology*, 192(5), 1446–9.

Mufarrij, P. W., Lange, J. N., Assimos, D. G., Mirzazadeh, M. & Holmes, R. P. (2012). Multibacterial growth from a surgical renal stone culture: a case report and literature review. *Reviews in urology*, 14(3-4), 108.

Müller-Mattheis, V., Schmale, D., Seewald, M., Rosin, H. & Ackermann, R. (1991). Bacteremia during extracorporeal shock wave lithotripsy of renal calculi. *The Journal of urology*, 146(3), 733–736.

Naber, K. G., Vahlensieck, W., Wagenlehner, F. M. E., Hummers-Pradier, E., Schmiemann, G., Hoyme, U. B., ... & Sester, U. (2011). Epidemiologie, Diagnostik, Therapie und Management unkomplizierter bakterieller ambulant erworbener Harnwegsinfektionen bei der erwachsenen Patientin. *Geburtshilfe und Frauenheilkunde*, 71(11), R94-R114.

O'keeffe, N., Mortimer, A., Sambrook, P. & Rao, P. (1993). Severe sepsis following percutaneous or endoscopic procedures for urinary tract stones. *BJU International*, 72(3), 277–283.

Pearle, M. S. & Roehrborn, C. G. (1997). Antimicrobial prophylaxis prior to shock wave lithotripsy in patients with sterile urine before treatment: a meta-analysis and cost-effectiveness analysis. *Urology*, 49(5), 679–686.

Pickard, R and Bartoletti, R and Bjerklund-Johansen, TE and Bonkat, G and Bruyère, F and Çek, M and Grabe, M and Tenke, P and Wagenlehner, F and Wullt, B and others. (2016). EAU guidelines on urological infections. *European Association of Urology, Arnhem, The Netherlands.*

Salem, N., Salem, L., Saber, S., Ismail, G. & Bluth, M. H. (2015). Corynebacterium urealyticum: a comprehensive review of an understated organism. *Infection and drug resistance*, 8, 129.

Scales, C. D., Smith, A. C., Hanley, J. M., Saigal, C. S., Project, U. D. in A. & others. (2012). Prevalence of kidney stones in the United States. *European urology*, 62(1), 160–165.

Skolarikos, A., Alivizatos, G. & de la Rosette, J. (2006). Extracorporeal shock wave lithotripsy 25 years later: complications and their prevention. *European urology*, 50(5), 981–90; discussion 990.

Songra, M., Damor, M., Namdev, R. K., Patbamniya, N. K., Nawalakhe, P. & Jain, R. (2016). A study on positive stone culture and its association with rate of sepsis after urological procedures. *International Surgery Journal*, 2(2), 239–246.

Stamatelou, K. K., Francis, M. E., Jones, C. A., Nyberg, L. M. & Curhan, G. C. (2003). Time trends in reported prevalence of kidney stones in the United States: 1976-1994. *Kidney international*, 63(5), 1817–1823.

Tavassoli, P., Paterson, R. & Grant, J. (2012). Actinobaculum schaalii: an emerging uropathogen? *Case reports in urology*, 2012.

Tavichakorntrakool, R., Prasongwattana, V., Sungkeeree, S., Saisud, P., Sribenjalux, P., Pimratana, C., Bovornpadungkitti, S., et al. (2012). Extensive characterizations of bacteria isolated from catheterized urine and stone matrices in patients with nephrolithiasis. *Nephrology Dialysis Transplantation*, 27(11), 4125–4130.

Teichman, J. M., Long, R. D. & Hulbert, J. C. (1995). Long-term renal fate and prognosis after staghorn calculus management. *The Journal of urology*, 153(5), 1403–1407.

Türk, C., Neisius, A., Petrik, A., Seitz, C., Skolarikos, A., Tepeler, A., Thomas, K., et al. (2017). EAU guidelines on Urolithiasis. Retrieved September 2017, from *http://uroweb.org/guideline/urolithiasis/*

Wang, X., Krambeck, A. E., Williams, J. C., Tang, X., Rule, A. D., Zhao, F., Bergstralh, E., et al. (2014). Distinguishing characteristics of idiopathic calcium oxalate kidney stone formers with low amounts of Randall's plaque. *Clinical Journal of the American Society of Nephrology*, 9(10), 1757–1763.

Wiesenthal, J. D., Ghiculete, D., Ordon, M., Pace, K. T. & Honey, R. J. D. (2011). 1176 a prospective study examining the incidence of bacteriuria and urinary tract infection post-shockwave lithotripsy: the case against universal antibiotic prophylaxis. *The Journal of Urology*, 185(4), e472.

Woolfrey, B. F., Ireland, G. K. & Lally, R. T. (1986). Significance of Gardnerella vaginalis in urine cultures. *American journal of clinical pathology*, 86(3), 324–329.

Chapter 5

5 General Discussion

5.1 Study of the microbiota in the urinary tract

5.1.1 Problems of applying 16S rRNA gene sequencing analysis to low abundance microbiota

The introduction of culture-independent, sequencing-based microbial analytic methods has greatly reshaped the understanding of microbes inhabiting a human body that many of these microbes can't be retrieved by routine traditional microbial culture methods. The awareness that not all bacteria present in clinical samples could be cultivated dates back to "The Great Plate Count Anomaly" phenomenon that microscopically identified microbial cell numbers might not coincide with the number of colonies in the Petri plates (Staley & Konopka, 1985). Now with the power of sequencing-based techniques, the number of bacterial phyla (*i.e.,* the highest level of taxonomical divisions under bacterial kingdom) has increased from 11 in 1987 to at least 85 at present, with many that have still to be cultivated (Achtman & Wagner, 2008; Keller & Zengler, 2004; Rappé & Giovannoni, 2003; Woese, 1987). These findings further promoted the investigation of the presence of bacteria at sites considered likely to be microbes-rich, including brain, breast tissue, blood, placenta and subepidermal tissue *etc.* in human body (Aagaard *et al.*, 2014; Branton *et al.*, 2013; Emery *et al.*, 2017; Justesen *et al.*, 2010; Nakatsuji *et al.*, 2013). Most of these studies claim the existence of "tissue-associated microbiota", and their potential role in the pathogenesis of diseases. However, a study by Salter et al. revealed that a remarkable proportion of sequencing reads were attributed to background noise from environment and reagents (Salter *et al.*, 2014). This study revealed the awareness of increased risk in analysing samples with low microbial biomass (Weiss *et al.*, 2014).

In this study, we further demonstrated that besides contamination from environment and reagent, the host DNA could also compromise the quality of microbiota study in multiple levels.

First, we showed that high host DNA abundance would saturate commonly-used DNA extraction kit columns. These extraction kits rely upon columns that have silicone filters to bind the DNA from the sample. Two of the most commonly used DNA extraction kits for gut microbiota studies include the Qiagen QIAamp® and MOBIO Powersoil® (now Qiagen) products. The QIAamp® column has a binding capability of up to 4-12 ug DNA per filter, whereas, the Powersoil® has a capacity up to 20 ug per filter (Manufacturer's manuals). Given that most gut microbes have a genome size around 4-6 Mb, this binding capability could retrieve presumably as high as 10^9 copies of bacterial gDNA, which when amplified, sequenced and analysed would provide a reliable representation of gut microbial communities. However, when it comes to tissue samples with low prokaryotic and high eukaryotic cell components, given that eukaryotic cells have a much higher genome size (*e.g.* 3000 Mb *homo sapiens* genome), a DNA collection filter as used in a commercial extraction kit, could only hold the DNA contents of around 6.5×10^6 cells in theory. If we assume that the rate of binding is the same for bacterial to host DNA, it would mean that only a small amount of bacterial DNA would be captured because the binding capacity of the column would rapidly be exhausted by the eukaryotic DNA. Since one gram of solid tumor is estimated to have 10^8 to 10^9 cells (Del Monte *et al.*, 2009; Rasnick *et al.*, 2002), therefore, a common input of 200 ug of this tissue as used in some of our studies, could contain 2×10^7 to 2×10^8 cells, rapidly saturating the DNA extraction columns. Given this and the low prokaryotic/eukaryotic cell ratio, the prokaryotic DNA yield could be very low in these columns and bear the risk of poor representation of the sample microbial communities that they are obtained from and the importance of confirmatory analysis by culture.

Secondly, we showed that the presence of disproportionately high host DNA abundance inhibits PCR amplification when using the bacterial specific 16S rRNA gene primers. The inhibiting effect of non-target DNA on PCR amplification performance is well known in molecular biology (Ludwig & Schleifer, 2000). However, compromised microbial rRNA gene amplification seems to be less recognised. Grice *et al*. (2008) reported that a eukaryotic/prokaryotic gDNA ratio of 1 to 100 would not significantly alter 16S rRNA gene amplification, however, our data shows that as the ratio continues to increase, amplification efficiency drops remarkably when eukaryotic/prokaryotic gDNA

ratio above 10^3 (Figure 6a and 6b in Chapter Two). Samples with a higher ratio would have a poor microbial community representation after DNA extraction and PCR amplification (Figure 6c and 6d).

Thirdly, we showed that the unspecific amplification of host DNA could take a remarkable proportion of the sequencing depth. Host DNA consuming sequencing depth is a known issue in whole genome sequencing and RNA sequencing (Feehery *et al.*, 2013). Studies with whole genome shotgun sequencing in the presence of large amounts of host eukaryotic DNA have shown that the background DNA will consume significant amounts of sequencing depth, making sequencing inefficient, time-consuming and inaccurate (Feehery et al., 2013). From these studies, it appears that the sequencing depth of 16S rRNA gene sequencing the presence of a high predominance of host DNA is also compromised.

Besides host DNA, samples with low microbial biomass are associated with risks of poor sequencing results as well. Recently Salter *et al.* showed that a low level of background DNA contamination is ubiquitous and almost inevitable for a sequencing-based microbiota studies (Salter *et al.* 2014; Weiss *et al*., 2014). Fecal samples have tremendously rich microbial content, whose 16S rRNA gene sequencing reads will indubitably overshadow the background noise. However, samples with low microbial biomass have their microbial DNA at a comparable level to contaminating DNA and is thus more vulnerable to sequencing bias and within-sample heterogeneity (Kennedy *et al*., 2014). In fact, as microbial gDNA concentration in a sample goes lower, the reads from background contamination will take increasing proportions of the result, which will lead to a complete background contamination sequencing result for an extremely low abundance sample (Salter *et al.*, 2014). It is, therefore, important to identify a lower limit of the microbial biomass of samples that could yield a decent sequencing result. We thus recommended sequencing serially diluted mock samples to identify this lower limit, then estimate the sequencing quality before the sequencing the samples.

5.1.2 Value of an optimised sample processing algorithm

As the fore-mentioned risk factors exist, processing samples with the high host to low bacterial gDNA ratios (“high-low” samples) with a protocol similar to ones used in the gut microbiota studies may greatly compromise analysis quality. Therefore, an adapted protocol is required. In the second part of Chapter Two, we developed a quality-checking tool and a refined algorithm which is considerate of these sample types. The aim of the refined algorithm is to mitigate the negative effect of high host DNA abundance and low microbial biomass.

The photo-activated DNA binding compound, propidium monoazide (PMA) has previously been used in microbiota studies when analyzing environmental samples with very low estimated microbial biomass, including ocean (Galand *et al.*, 2009), cleanroom environments (Weinmaier *et al.*, 2015), and even international space stations(Checinska *et al.*, 2015; Venkateswaran *et al.*, 2014) . It has been used to differentiate DNA from living microbial cells to that from free extracellular DNA or DNA present within dead cells (Emerson *et al.*, 2017). Living microbial cells have intact cell membrane, which prevents PMA cellular uptake and subsequent DNA binding. This technique is also being used in fluorescence staining with flow cytometry, where the proportion of living microbes in drinking water are approximately 70-80% (Berney *et al*., 2008; Luna *et al*., 2002; Yokomaku, Yamaguchi, & Nasu, 2000), and 50-60% in surface fresh water, estuarine and coastal marine stations (Freese, *et al.*, 2006; Schumann *et al.*, 2003). The limitation of this membrane integrity-based technique includes (1). It lacks differentiability in dying cells because the membrane does not instantly disintegrate. (2). PMA cant evaluate the viability of microbial forms with strong wall structure such as spores (Berney *et al*., 2008; Bosshard *et al*., 2009; Lisle *et al*. 1999).

Microbial DNA enrichment kits are commonly used in whole genome sequencing, where eukaryotic host DNA and microbial DNA are both present. Various mechanisms are utilised in commercial microbial enrichment kits to deplete eukaryotic DNA. The NEBNext® Microbiome DNA Enrichment kit from New England Biolab uses a methylated CpG-specific binding protein to bind methyl-CpG that only present in eukaryotic cells, while the Molzym's MolYsis kit selectively lyse human cells using

chaotropic reagents, then degrades all extracellular DNA with DNase (Thoendel *et al.*, 2016). Sequencing of 16S rRNA gene amplicon specific primers typically won't amplify human DNA, however, we have shown in chapter two that in "high-low" samples, host DNA could inhibit PCR amplification and that unspecific amplification consumes sequencing reads. Therefore, microbial enrichment kits remain as a useful tool to deplete host DNA when handling "high-low" samples for 16S rRNA gene sequencing.

Although the techniques mentioned above have been previously reported, here we are first to report that silicon-based filters in DNA extraction kits should be avoided. By doing this and substituting for an alternative DNA collection technology, such as the traditional ethanol precipitation, much higher microbial DNA yield will be realized for these "high-low" samples.

5.1.3 Verification of a microbiota in renal tissue

In Chapter two, we examined renal tumour and healthy adjacent tissue from the same individuals for the presence of bacterial microbiota. The conventional medical doctrine is that the upper urinary tract is considered to be "sterile". Conditions such as pyelonephritis are thought to typically occur where urinary obstruction can be present in the urogenital tract and the result of reflux and opportunistic ascending microorganisms. Obstruction can be the result of urolithiasis, stricture and the presence of tumours. The former has been reported in early investigations of the bacterial composition of kidney stones which appear to show the presence of *E. coli* (Barr-Beare *et al*., 2015) and is potential evidence which suggests that trace amounts of bacteria exist in the upper urinary tract.

Of note, uropathogens such as *E.coli* were able to invade into kidney urothelial cells(Chassin *et al.*, 2008; Pichon *et al.*, 2009) and form intracellular bacterial communities (Eto *et al.*, 2006) , and then move even deeper and form quiescent intracellular reservoirs (Hunstad & Justice, 2010) . In this way, rather than adhering to the luminal surface of urothelial cells, bacteria could also move deeper into the parenchymal tissue, where they were able to avoid antibiotics and host immune response. Bacteria in these reservoirs maintained minimal metabolic activities and usually could not be cultivated, but they could restore their activity and cause subsequent recurrence or

asymptomatic infection (Lewis *et al.*, 2007). Through the development of DNA extraction methods, our study supports the existence of bacteria in renal parenchyma tissue.

5.2 Urinary microbiota profiles of healthy individuals and urinary stone patients

5.2.1 Urine sample storage form and sequencing quality

Sample processing and storage methods are arguably the most critical steps in microbiota studies since it has a high risk of influencing microbial profiles. It has been reported in gut microbiota studies that sample storage methods can significantly affect faecal microbiota profiles (Choo *et al.*, 2015; Gorzelak *et al.*, 2015; Nsubuga *et al.*, 2004). Optimal preparation, handling and storage protocols for stool samples have also been proposed as a standard methodology (Gorzelak *et al*., 2015). Unfortunately, for urine samples, there have no optimised storage methods yet reported. In the first part of Chapter Three, we performed experiments to investigate how sample volume and storage form could affect the microbiota sequencing analysis results. Urine samples with greater than 20 ml in volume, which were stored in centrifuged pellets gave the same consistent results. Any smaller volumes collected tended to reflect the background contaminating sequences to some degree.

With the exception of stool samples, studies focusing on the optimal preparation, handling and storage methods for next generation sequencing-based microbiological studies are rather limited. Part of the reason for that is that simple act of culturing urine can be performed almost immediately after the samples collection, so there's no need for extended sample storage solutions. However, for sequencing-based microbiota analysis, clinical samples are usually processed in batches, making it a requirement to store urine samples after collections. As the interest in urinary microbiota grows, it will become increasingly crucial to develop a standardised operating protocol (SOP) for the handling of urine samples.

In most biobank sample handling SOPs, urine samples are collected and stored for biomarker detection. Therefore, relatively low volumes are typically captured, and no precautions are taken to maintain the viability of microorganisms (Elliott & Peakman, 2008). In our study, we showed that these conditions are not optimal for microbiota studies. Given the relatively low human and microbial cell numbers in a urine sample, the volume requirement to capture amounts of microbial gDNA required for a reasonable analysis is higher than metabolites analysis. Roewer *et al*. reported that 200 ml of urine are required to perform a human DNA type with multilocus probes (Roewer *et al.,* 1990) . Brinkman et al. showed that 10 ml of the urine sample are generally insufficient to perform DNA polymorphism using Southern blotting (Brinkmann *et al.,* 1992) . In our study, urine samples with a volume higher than 20 ml seem to have a decent representation of the urinary microbiota profiles. As a growing number of microbiome biobank being established ("Harvard Chan School awarded $4.9 million to create microbiome biobank," the "PennCHOP Microbiome Program," webpage) , our results could be helpful to set their urine sample collection and storage SOPs.

5.2.2 Longitudinal stability of urinary microbiota

In our study, we showed that the urinary microbiota from one donor when analysed at different time points over three months have high similarity, but much less so than urine samples from other donors of similar gender and age. We thus can infer that the urinary microbiota has its stability and resilience as an ecosystem. To our knowledge, this is no previous study on the ecological stability of the urinary microbiota, but multiple studies demonstrated the ability of gut microbiota to recover after a pathological condition or iatrogenic procedures (David *et al.,* 2015; Dethlefsen & Relman, 2011; Fukuyama *et al.*, 2017; Raymond *et al*., 2016) . Similarly, we also demonstrated the recovery of urinary microbiota after lithotripsy surgery in Chapter four. Reported factors contributing to the stability of gut microbiota includes topical microenvironment, genetic background, diet and host age, *etc.* (Relman *et al.,* 2012) . in reverse, gut microbiota profiles also reflect the health, nutritional and drug administration information of the host (Lozupone *et al.*, 2012). The urinary microbiota is believed to be shaped and affected by two adjacent ecosystem: the gastrointestinal tract and the external genitalia (including vagina in

females), as many common urinary inhabitant microbial species can find their counterparts in these sites (Dong *et al.*, 2011; Khasriya *et al.*, 2013; D. A. Lewis *et al.*, 2013; Siddiqui *et al*., 2011; Willner *et al.*, 2014). Under pathological conditions, these two sites also serve as a source of microbial invasion and pathogen reservoir which cause UTI recurrence and relapse (Bao *et al.*, 2014). Other key factors shaping the urinary microbiota, on the other hand, awaits further studies.

In fact, the stability of the ecosystem is a well-discussed topic in ecology. Elton in 1958 claimed population variability, population recovery, ease of invasion and the consequences of invasion as aspects and components of ecosystem stability (Elton *et al.*, 2000). Pimm *et al.* set a few criteria of ecosystem stability, including parameters showing the system returning to previous equilibrium after displacement, the time for the system to return to equilibrium is relatively short, the main parameters of a given system remains stable during a displacement, the system has the drive against the displacement, and the variation of main parameters over time is limited (Pimm *et al.*, 1991). In our study, the urinary microbiota demonstrated its resilience as its alpha and beta diversity index (Shannon's index and Weight UniFrac distance scores) maintained low variation level throughout lithotripsy surgery and could return to the pre-disturbance equilibrium in a relatively short period (three months or shorter). Recognizing this resilience of urinary microbiota is crucial when further investigating how displacing events (such as antibiotics and surgical procedures as discussed in Chapter four) affect the urinary microbiota and the degree of resilience of urinary microbiota beyond which it can't restore the pre-disturbance conditions.

5.3 Bacteria in metabolic urinary stones

In chapter four, we showed that some of the non-infectious, calcium-based metabolic urinary stones could also harbour bacteria, as can be observed in immediate postoperative urine samples, despite that all these patients have a negative preoperative urine culture. In fact, several other studies also reported bacteria in metabolic urinary stones. McCartney *et al.* reported that among 16 patients with calcium oxalate stones and negative preoperative urine culture, four of them has positive culture results on their stone samples (McCartney *et al.*, 1985). Gault *et al.* investigated all culture-positive urine stone samples

of 258 patients, and found only 43% of kidney stones, none of the ureteric stones and 50% of the bladder stones contained magnesium ammonium phosphate composition (Gault *et al.*, 1995). Hugosson *et al.* performed stone and urine cultures from 215 patients and found among all calcium oxalate phosphate stones 32% has positive stone cultures (Hugosson *et al.*, 1990). These results suggest that urinary stone of any composition could harbour bacteria and release them during lithotripsy. In our study, we observed a "bacterial fluctuation" in immediate postoperative urine samples which contained microbial species that could not be detected in preoperative samples. Interestingly, the OTUs that showed fluctuation in sequencing correlated with the culture results of stone and pelvic urine samples, which further confirmed that the source of these is bacteria likely liberated from the stone. We also noticed that stones with routine type culture analysis that were positive seemed to have a broader spectrum of microbial species identified by 16S rRNA gene sequencing than those that didn't, potentially demonstrating the ability of sequencing-based techniques to characterise microbial populations that are not cultivated during routine analysis.

The findings of bacteria in metabolic stone arouse questions of the role bacteria in the metabolic stone formation and whether they a nidus for inflammation leading to crystal nucleation. Clinical studies found that urinary stone patients are more likely to have UTIs, even though such stones may not cause an obstruction (Holmgren *et al*., 1989; Hugosson *et al.*, 1989). Interestingly, the most common metabolic stone-associated bacteria are *E.coli*, while *Proteus* species are more common in struvite stones (Gault *et al.*, 1995). Barr-Beare *et al.* showed in a rodent model that *E.coli* infection caused increased kidney calcium oxalate deposits (Barr-Beare *et al.*, 2015). The exact mechanism of *E.coli* inducing metabolic stone formation has yet to be elucidated and beyond the scope of this study.

5.4 Value of voided urine samples in studying the kidney stone microbiota and its use in predicting postoperative infective complications

Although most patients had a negative preoperative urine culture before lithotripsy, postoperative infective complications are not rare. Many studies have also reported

infective complications such as bacteriuria in patients with clean voided preoperative urine cultures. Many urinary stone patients also develop infective complications with a negative preoperative urine culture (Larsen *et al.*, 1986). Charton *et al*. documented 35% of patients receiving PCNL could develop postoperative bacteriuria, while all these patients had negative preoperative urine culture (Charton *et al.*, 1986). Mariappan *et al*. carried out a study among patient receiving ureteroscopic lithotripsy, and found 12 out of 75 included patients developed postoperative bacteriuria (Mariappan & Loong, 2004). Bratell *et al.* showed that many of the bacterial species cultured from stone samples are unidentifiable on voided preoperative urine samples, emphasizing the importance of stone cultures (Bratell *et al.*, 1990). Of note, in our studies, most patients had negative postoperative urine cultures by routine methods, but 16S rRNA gene sequencing could still identify evidence of bacteria liberated from the stone. These results suggested metabolic stones release bacteria during lithotripsy, and in many cases, it stayed below clinical detection range. In fact, the threshold of a routine urine culture is always under some controversy. The commonly used 10^5 colony forming units (CFU)/mL potentially miss many clinically significant infections, and there are also recommendations favouring a threshold of 103 CFU/mL (Jungwirth *et al.*, 2012; Schmiemann, Gebhardt, Matejczyk, & Hummers-Pradier, 2009). Both of these are subjective determinants. These controversies to some extent reflect the difficulty of differentiating the pathogens from the bladder or kidney from commensal microbes in urethral and genitalia when using voided urine samples. This issue also presents when screening bacteriuria in stone patients. Several studies have reported that the microbes isolated from the stone culture cannot be identified from voided urine samples (Mariappan *et al.*, 2005; Mufarrij *et al.*, 2012; Songra *et al.*, 2016). In our study of the paired sequencing of stone and urine samples, we can see that OTUs that predominated in the stone sequencing results may only occupy a small proportion in the voided urine sample, with the remaining reads attributed to the urethra and genitalia-associated species. These may be part of the reasons why people claim that voided urine has inferior performance in detection of bacteria in stones. Lewi et al. showed that the positive rate of preoperative voided urine, stone sample and pelvic urine cultures are 29%, 38% and 30% respectively, with the voided urine sample to be the lowest (Lewi *et al.*, 1984). The advantage of the voided

urine sample, however, is that (a) it can be obtained non-invasively, and (b). it can be obtained preoperatively for risk evaluation of infective complications. In our study, none of the six patients who had their stone samples sequenced had positive post-operative urine culture when analyzed by standard culture techniques, however, for the two of the patients that had positive stone culture and sequencing results, microbial species that presented in stone samples also had their OTUs increased in relative abundance in the sequencing results. These results suggested that sequencing-based techniques could increase the sensitivity to detect bacteria release from stone during lithotripsy with voided urine samples.

5.5 Future directions

Detection of microorganisms in very low microbial biomass is always challenging, especially when the microbes are fastidious. In our study, we utilised sequencing-based techniques to identify the DNA of these microbes in kidney tissues, and undertook methodologies to limit contamination by using techniques such as PMA to deactivate any extracellular DNA. We then confirmed many of these microbes with enhanced culture techniques. The disadvantage of this strategy is the risk of contamination during sample handling. A more direct method would be using staining methods that have live/dead differentiation abilities, for sample the viability stains that based on membrane integrity, then visualise the bacteria within a tissue section using confocal microscopy. In this way, bacteria could be visualised in situ, making it possible to identify their inhabiting locus within a given tissue.

For the stability and resilience of urinary microbiota, one aspect that was not covered in our present study is the “elasticity” of the urinary microbiota, i, the degree of disturbance or displacement to the urinary microbiota that it could be still able to restore, and the rate at which the microbiota returns to its previous conditions. In our study, most of the participants maintained their urinary microbiota stability partly because they are either healthy individuals or uncomplicated urinary stone patient with relatively low risk of severe disturbance. How urinary microbiota responds to a more severe condition, including chronic infection with anatomical abnormalities, long-term foreign body

implantation, direct connection with gastrointestinal tract such as fistula or urinary diversion awaits further study.

To carry on studies on roles of bacteria on the metabolic stone formation, a cross-sectional study should be carried out to explore the incidence of metabolic stones harbouring bacteria. Multiple studies have reported positive bacterial culture from urinary samples, but most of these studies have a relatively small sample size (<= 150 donors) (Sohshang *et al.*, 2000; Takeuchi *et al.*, 1989; Tavichakorntrakool et al., 2012; Yoshida *et al.*, 1984). Microbiological screening of a larger cohort needs to be applied to confirm the observation that *E.coli* is the most commonly found microorganism in metabolic stones.

There have been reports that inflammation could promote urinary stone formation (Ghazali *et al.*, 1997; Rule *et al.*, 2009). The most extreme case would be the matrix stones, which mainly consists of inflammatory proteins (Canales *et al.*, 2009). Metabolic stones in the kidney may also induce inflammation by mechanic effects or urinary obstruction through Il-1β released by dendritic cells, and urine Il-6 would also be elevated (Canales *et al.*, 2010; Mulay *et al.*, 2012). The innate immune reaction may later be activated, and the immune proteins may form the inner core of the stone which in turn facilitate crystalluria precipitation (Mushtaq *et al.*, 2007). How bacteria may modulate the inflammation cascade, and the interaction of immune proteins would be the interest of further studies.

5.6 Final statement

As our results and other studies have shown, the study of samples with high host DNA and low microbial biomass is technically challenging. Described by Joanne *et al.* as “Schrödinger’s microbes”, analysis of microbial composition and function in these samples is fraught with complications. Extra caution, meticulously optimised methods and combination with enhanced culture methods are required to obtain a relatively accurate result. Despite these limitations, sequencing-based techniques remain as a novel and sensitive methods for microbiota research, and have the potential for further clinical application.

5.7 References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J. & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science translational medicine*, 6(237), 237ra65.

Achtman, M. & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, 6(6), 431–440.

Bao, Y., Welk, B., Reid, G. & Burton, J. P. (2014). Role of the microbiome in recurrent urinary tract infection. *Novel Insights into Urinary Tract Infections and their Management*, 49.

Barr-Beare, E., Saxena, V., Hilt, E. E., Thomas-White, K., Schober, M., Li, B., Becknell, B., et al. (2015). The interaction between Enterobacteriaceae and calcium oxalate deposits. *PloS one*, 10(10), e0139575.

Berney, M., Vital, M., Hülshoff, I., Weilenmann, H.-U., Egli, T. & Hammes, F. (2008). Rapid, cultivation-independent assessment of microbial viability in drinking water. *Water research*, 42(14), 4010–4018.

Bosshard, F., Berney, M., Scheifele, M., Weilenmann, H.-U. & Egli, T. (2009). Solar disinfection (SODIS) and subsequent dark storage of Salmonella typhimurium and Shigella flexneri monitored by flow cytometry. *Microbiology*, 155(4), 1310–1317.

Branton, W. G., Ellestad, K. K., Maingat, F., Wheatley, B. M., Rud, E., Warren, R. L., ... & Power, C. (2013). Brain microbial populations in HIV/AIDS: α-proteobacteria predominate independent of host immune status. *PloS one*, 8(1), e54673.

Bratell, S., Brorson, J., Grenabo, L., Hedelin, H. & Pettersson, S. (1990). The bacteriology of operated renal stones. *European urology*, 17, 58–61.

Brinkmann, B., Rand, S. & Bajanowski, T. (1992). Forensic identification of urine samples. *International journal of legal medicine*, 105(1), 59–61.

Canales, B. K., Anderson, L., Higgins, L., Ensrud-Bowlin, K., Roberts, K. P., Wu, B., Kim, I. W., et al. (2010). Proteome of human calcium kidney stones. *Urology*, 76(4), 1017–e13.

Canales, B. K., Anderson, L., Higgins, L., Frethem, C., Ressler, A., Kim, I. W. & Monga, M. (2009). Proteomic analysis of a matrix stone: a case report. *Urological research*, 37(6), 323.

Charton, M., Vallancien, G., Veillon, B. & Brisset, J. (1986). Urinary tract infection in percutaneous surgery for renal calculi. *The Journal of urology*, 135(1), 15–17.

Chassin, C., Vimont, S., Cluzeaud, F., Bens, M., Goujon, J.-M., Fernandez, B., Hertig, A., et al. (2008). TLR4 facilitates translocation of bacteria across renal collecting duct cells. *Journal of the American Society of Nephrology*, 19(12), 2364–74.

Checinska, A., Probst, A. J., Kumar, D., Pierson, D. L., Fox, G. E., Nilsson, H. R., White, J. R., et al. (2015). Microbiomes of the dust particles collected from the International Space Station and Spacecraft Assembly Facilities. *Microbiome*, 3(1), 50.

Choo, J. M., Leong, L. E. & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific reports*, 5.

David, L. A., Weil, A., Ryan, E. T., Calderwood, S. B., Harris, J. B., Chowdhury, F., Begum, Y., et al. (2015). Gut microbial succession follows acute secretory diarrhea in humans. MBio, 6(3), e00381–15.

Dethlefsen, L. & Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiome to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences*, 108(Supplement 1), 4554–4561.

Dong, Q., Nelson, D. E., Toh, E., Diao, L., Gao, X., Fortenberry, J. D. & Van Der Pol, B. (2011). The microbial communities in male first catch urine are highly similar to those in paired urethral swab specimens. *PloS one*, 6(5), e19709.

Elliott, P. & Peakman, T. C. (2008). The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *International journal of epidemiology*, 37(2), 234–44.

Elton, C. S. (2000). *The ecology of invasions by animals and plants.* University of Chicago Press.

Emerson, J. B., Adams, R. I., Román, C. M. B., Brooks, B., Coil, D. A., Dahlhausen, K., Ganz, H. H., et al. (2017). Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome*, 5(1), 86.

Emery, D. C., Shoemark, D. K., Batstone, T. E., Waterfall, C. M., Coghill, J. A., Cerajewska, T. L., Davies, M., et al. (2017). 16S rRNA next generation sequencing analysis shows bacteria in Alzheimer's post-mortem brain. *Frontiers in aging neuroscience*, 9, 195.

Eto, D. S., Sundsbak, J. L. & Mulvey, M. A. (2006). Actin-gated intracellular growth and resurgence of uropathogenic Escherichia coli. *Cellular microbiology*, 8(4), 704–717.

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., et al. (2013). A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PloS one*, 8(10), e76096.

Freese, H., Karsten, U. & Schumann, R. (2006). Bacterial abundance, activity, and viability in the eutrophic river Warnow, Northeast Germany. *Microbial Ecology*, 51(1), 117–127.

Fukuyama, J., Rumker, L., Sankaran, K., Jeganathan, P., Dethlefsen, L., Relman, D. A. & Holmes, S. P. (2017). Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment. *PLoS Computational Biology*, 13(8), e1005706.

Galand, P. E., Casamayor, E. O., Kirchman, D. L. & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proceedings of the National Academy of Sciences*, 106(52), 22427–22432.

Gault, M., Longerich, L., Crane, G., Cooper, R., Dow, D., Best, L., Stockall, E., et al. (1995). Bacteriology of urinary tract stones. *The Journal of urology*, 153(4), 1164–1170.

Ghazali, A., Fuentes, V., Desaint, C., Bataille, P., Westeel, A., Brazier, M., Prin, L., et al. (1997). Low bone mineral density and peripheral blood monocyte activation profile in calcium stone formers with idiopathic hypercalciuria. *The Journal of Clinical Endocrinology & Metabolism*, 82(1), 32–38.

Gorzelak, M. A., Gill, S. K., Tasnim, N., Ahmadi-Vand, Z., Jay, M. & Gibson, D. L. (2015). Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. *PloS One,* 10(8), e0134802

Grice, E. A., Kong, H. H., Renaud, G., Young, A. C., Bouffard, G. G., Blakesley, R. W., Wolfsberg, T. G., et al. (2008). A diversity profile of the human skin microbiome. *Genome research,* 18(7), 1043–50.

Harvard Chan School awarded $4.9 million to create microbiome biobank. Retrieved from *https://www.hsph.harvard.edu/news/features/harvard-chan-school-awarded-4-9-million-to-create-microbiome-biobank/*

Holmgren, K., Danielson, B. G., Fellström, B., Ljunghall, S., Niklasson, F. & Wikström, B. (1989). The relation between urinary tract infections and stone composition in renal stone formers. *Scandinavian journal of urology and nephrology*, 23(2), 131–136.

Hugosson, J., Grenabo, L., Hedelin, H., Lincoln, K. & Pettersson, S. (1989). Chronic urinary tract infection and renal stones. *Scandinavian journal of urology and nephrology*, 23(1), 61–66.

Hugosson, J., Grenabo, L., Hedelin, H., Pettersson, S. & Seeberg, S. (1990). Bacteriology of upper urinary tract stones. *The Journal of urology*, 143(5), 965–968.

Hunstad, D. A. & Justice, S. S. (2010). Intracellular lifestyles and immune evasion strategies of uropathogenic Escherichia coli. *Annual review of microbiology*, 64, 203–21.

Jungwirth, A., Giwercman, A., Tournaye, H., Diemer, T., Kopa, Z., Dohle, G., Krausz, C., et al. (2012). European Association of Urology guidelines on Male Infertility: the 2012 update. *European urology*, 62(2), 324–332.

Justesen, U. S., Skov, M. N., Knudsen, E., Holt, H. M., Søgaard, P. & Justesen, T. (2010). 16S rRNA gene sequencing in routine identification of anaerobic bacteria isolated from blood cultures. *Journal of Clinical Microbiology*, 48(3), 946–948.

Keller, M. & Zengler, K. (2004). Tapping into microbial diversity. *Nature Reviews Microbiology*, 2(2), 141–150.

Kennedy, K., Hall, M. W., Lynch, M. D., Moreno-Hagelsieb, G. & Neufeld, J. D. (2014). Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Applied and environmental microbiology*, 80(18), 5717–5722.

Khasriya, R., Sathiananthamoorthy, S., Ismail, S., Kelsey, M., Wilson, M., Rohn, J. L. & Malone-Lee, J. (2013). Spectrum of bacterial colonization associated with urothelial cells from patients with chronic lower urinary tract symptoms. *Journal of clinical microbiology*, 51(7), 2054–62.

Larsen, E. H., Gasser, T. & Madsen, P. (1986). Antimicrobial prophylaxis in urologic surgery. *The Urologic clinics of North America*, 13(4), 591–604.

Lewi, H., White, A., Hutchinson, A. & Scott, R. (1984). The bacteriology of the urine and renal calculi. *Urological research*, 12(2), 107–109.

Lewis, D. A., Brown, R., Williams, J., White, P., Jacobson, S. K., Marchesi, J. R. & Drake, M. J. (2013). The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. *Frontiers in cellular and infection microbiology*, 3.

Lewis, K. (2007). Persister cells, dormancy and infectious disease. *Nature Reviews Microbiology*, 5(1), 48–56.

Lisle, J., Pyle, B. & McFeters, G. (1999). The use of multiple indices of physiological activity to access viability in chlorine disinfected Escherichia coli O157: H7. *Letters in Applied Microbiology*, 29(1), 42–47.

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiome. *Nature*, 489(7415), 220–30.

Ludwig, W. & Schleifer, K.-H. (2000). How quantitative is quantitative PCR with respect to cell counts? *Systematic and applied microbiology*, 23(4), 556–562.

Luna, G., Manini, E. & Danovaro, R. (2002). Large fraction of dead and inactive bacteria in coastal marine sediments: comparison of protocols for determination and ecological significance. *Applied and Environmental Microbiology*, 68(7), 3509–3513.

Mariappan, P. & Loong, C. W. (2004). Midstream urine culture and sensitivity test is a poor predictor of infected urine proximal to the obstructing ureteral stone or infected stones: a prospective clinical study. *The Journal of urology*, 171(6), 2142–2145.

Mariappan, P., Smith, G., Bariol, S. V., Moussa, S. A. & Tolley, D. A. (2005). Stone and pelvic urine culture and sensitivity are better than bladder urine as predictors of urosepsis following percutaneous nephrolithotomy: a prospective clinical study. *The Journal of urology*, 173(5), 1610–1614.

McCartney, A., Clark, J. & Lewi, H. (1985). Bacteriological study of renal calculi. *European Journal of Clinical Microbiology & Infectious Diseases*, 4(6), 553–555.

Del Monte, U. (2009). Does the cell number 109 still really fit one gram of tumor tissue? *Cell Cycle*, 8(3), 505–506.

Mufarrij, P. W., Lange, J. N., Assimos, D. G., Mirzazadeh, M. & Holmes, R. P. (2012). Multibacterial growth from a surgical renal stone culture: a case report and literature review. *Reviews in urology*, 14(3-4), 108.

Mulay, S. R., Kulkarni, O. P., Rupanagudi, K. V., Migliorini, A., Darisipudi, M. N., Vilaysane, A., ... & Anders, H. J. (2012). Calcium oxalate crystals induce renal inflammation by NLRP3-mediated IL-1β secretion. *The Journal of clinical investigation*, 123(1).

Mushtaq, S., Siddiqui, A. A., Naqvi, Z. A., Rattani, A., Talati, J., Palmberg, C., & Shafqat, J. (2007). Identification of myeloperoxidase, α-defensin and calgranulin in calcium oxalate renal stones. *Clinica chimica acta*, 384(1), 41-47.

Nakatsuji, T., Chiang, H.-I., Jiang, S. B., Nagarajan, H., Zengler, K. & Gallo, R. L. (2013). The microbiome extends to subepidermal compartments of normal skin. *Nature communications*, 4, 1431.

Nsubuga, A. M., Robbins, M. M., Roeder, A. D., Morin, P. A., Boesch, C. & Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular ecology*, 13(7), 2089–2094.

PennCHOP Microbiome Program. (n.d.). Retrieved 3, 2017, from *http://www.med.upenn.edu/penn-chop-microbiome/acutecarebiobanking.html*

Pichon, C., Héchard, C., Du Merle, L., Chaudray, C., Bonne, I., Guadagnini, S., Vandewalle, A., et al. (2009). Uropathogenic Escherichia coli AL511 requires flagellum to enter renal collecting duct cells. *Cellular microbiology*, 11(4), 616–628.

Pimm, S. L. (1991). *The balance of nature: ecological issues in the conservation of species and communities*. University of Chicago Press.

Rappé, M. S. & Giovannoni, S. J. (2003). The uncultured microbial majority. *Annual Reviews in Microbiology*, 57(1), 369–394.

Rasnick, D. (2002). Aneuploidy theory explains tumor formation, the absence of immune surveillance, and the failure of chemotherapy. *Cancer genetics and cytogenetics*, 136(1), 66–72.

Raymond, F., Déraspe, M., Boissinot, M., Bergeron, M. G. & Corbeil, J. (2016). Partial recovery of microbiomes after antibiotic treatment. *Gut microbes*, 7(5), 428–434.

Relman, D. A. (2012). The human microbiome: ecosystem resilience and health. *Nutrition reviews*, 70(s1).

Roewer, L., Nürnberg, P., Fuhrmann, E., Rose, M., Prokop, O. & Epplen, J. (1990). Stain analysis using oligonucleotide probes specific for simple repetitive DNA sequences. *Forensic science international*, 47(1), 59–70.

Rule, A. D., Bergstralh, E. J., Melton, L. J., Li, X., Weaver, A. L. & Lieske, J. C. (2009). Kidney stones and the risk for chronic kidney disease. *Clinical Journal of the American Society of Nephrology*, 4(4), 804–811.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology*, 12(1), 87.

Schmiemann, G., Gebhardt, K., Matejczyk, M., & Hummers-Pradier, E. (2009). DEGAM-Leitlinie Nr. 1: Brennen beim Wasserlassen–Update 2009. *Düsseldorf: Omikron publishing*.

Schumann, R., Schiewer, U., Karsten, U. & Rieling, T. (2003). Viability of bacteria from different aquatic habitats. II. Cellular fluorescent markers for membrane integrity and metabolic activity. *Aquatic Microbial Ecology*, 32(2), 137–150.

Siddiqui, H., Nederbragt, A. J., Lagesen, K., Jeansson, S. L. & Jakobsen, K. S. (2011). Assessing diversity of the female urine microbiome by high throughput sequencing of 16S rDNA amplicons. *BMC microbiology*, 11(1), 244.

Sohshang, H. L., Singh, M., Singh, N. & Singh, S. (2000). Biochemical and bacteriological study of urinary calculi. *The Journal of communicable diseases*, 32(3), 216–221.

Songra, M., Damor, M., Namdev, R. K., Patbamniya, N. K., Nawalakhe, P. & Jain, R. (2016). A study on positive stone culture and its association with rate of sepsis after urological procedures. *International Surgery Journal*, 2(2), 239–246.

Staley, J. T. & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology*, 39(1), 321–346.

Takeuchi, H., Okada, Y., Yoshida, O., Arai, Y. & Tomoyoshi, T. (1989). Urinary tract infection associated with urinary calculi. 1. The significance of urinary tract infection in urinary calculi. Hinyokika kiyo. *Acta urologica Japonica*, 35(5), 749–754.

Tavichakorntrakool, R., Prasongwattana, V., Sungkeeree, S., Saisud, P., Sribenjalux, P., Pimratana, C., Bovornpadungkitti, S., et al. (2012). Extensive characterizations of bacteria isolated from catheterized urine and stone matrices in patients with nephrolithiasis. *Nephrology Dialysis Transplantation*, 27(11), 4125–4130.

Thoendel, M., Jeraldo, P. R., Greenwood-Quaintance, K. E., Yao, J. Z., Chia, N., Hanssen, A. D., Abdel, M. P., et al. (2016). Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *Journal of microbiological methods*, 127, 141–145.

Venkateswaran, K., Vaishampayan, P., Cisneros, J., Pierson, D. L., Rogers, S. O. & Perry, J. (2014). International Space Station environmental microbiome—microbial inventories of ISS filter debris. *Applied microbiology and biotechnology*, 98(14), 6453–6466.

Weinmaier, T., Probst, A. J., Duc, M. T., Ciobanu, D., Cheng, J.-F., Ivanova, N., Rattei, T., et al. (2015). A viability-linked metagenomic analysis of cleanroom environments: eukarya, prokaryotes, and viruses. *Microbiome*, 3(1), 62.

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J. & Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology*, 15(12), 564.

Willner, D., Low, S., Steen, J. A., George, N., Nimmo, G. R., Schembri, M. A. & Hugenholtz, P. (2014). Single clinical isolates from acute uncomplicated urinary tract infections are representative of dominant in situ populations. *MBio*, 5(2), e01064–13.

Woese, C. R. (1987). Bacterial evolution. *Microbiological reviews*, 51(2), 221.

Yokomaku, D., Yamaguchi, N. & Nasu, M. (2000). Improved direct viable count procedure for quantitative estimation of bacterial viability in freshwater environments. *Applied and environmental microbiology*, 66(12), 5544–5548.

Yoshida, O., Kiriyama, T., Okada, K., Okada, Y., Watanabe, H., Mishina, T., Uchida, M., et al. (1984). A bacteriological study on urinary calculi associated with infections. Hinyokika kiyo. *Acta urologica Japonica*, 30(2), 191–198.

Appendices

Appendix A: Primer details for 16S rRNA gene amplification

The target of these primer sets was the V4/V6 hypervariable region of the 16S rRNA gene: V4-forward, 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCTnnnn(8)GTGCCAGCMGCCGCGGTAA 3'; and V4-reverse, 5' CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTnnnn(8)GGACTACHVGGGTWTCTAAT 3'. V6-forward, 5′ACACTCTTTCCCTACACGACGCTCTTCCGATCTnnnn(8)CWACGCGARGAACCTTACC3′; and V6-reverse, 5′CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTnnnn(8)ACRACACGAGCTGACGAC3′.

In the primers, the first 35-37 capitalized letters were the Illumina paired-end sequencing adapters, "nnnn" were four random nucleotides, and "(8)" was a specific sample barcode region which allows for sample multiplexing.

Appendix B: PCR reaction system of V4 amplicon library construction

The PCR was performed in a 42-μl reaction system, including 2 μl of DNA template (for negative controls nuclease-free water was used instead), 0.15 μg/μl of bovine serum albumin, 20 μl of 2× GoTaq hot-start colorless master mix (Promega), 10 μl of left and right primers (3.2 pmol/μl). Samples were amplified individually with an Eppendorf Mastercycler under the following conditions: initial denaturation at 95°C for 2 min followed by 25 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min.

Appendix C: The exact workflow for 16S sequencing

This part is adapted from the supplementary material with permission of

Bian G, Gloor GB, Gong A, Jia C, Zhang W, Hu J, Zhang H, Zhang Y, Zhou Z, Zhang J, Burton JP, Reid G, Xiao Y, Zeng Q, Yang K, Li J. 2017. The gut microbiota of healthy aged Chinese is similar to that of the healthy young. mSphere 2:e00327-17. https://doi.org/10.1128/mSphere.00327-17.

Step 1: Download and de-compress the MiSeq reads. Place the reads into the reads/directory. Reads are compressed with 7Zip. The following commands were used to decompress the file:

7z e filename

gunzip XthRun_S1_L001_R1_00.fastq.gz

Step 2: Overlap the reads with pandaseq v2.5. An example command for this with a minimum overlap of 30 nucleotides is below. This command is appropriate for the V4 amplimers:

pandaseq -g pandaseq_log.txt -T 8 -f Burton-576_S1_L001_R1_001.fastq -r Burton-576_S1_L001_R2_001.fastq -o 30 -w ps_overlapped30.fastq -F &

Step 3: Run the workflow pipeline:

./workflow.sh name 0.97 V4EMB

3.1 extracting out the barcodes and primers associated with a particular samples.txt file. Output is a tabbed format file with the fields: read ID, sample ID primer sequence primer barcode q-score

$BIN/process_miseq_reads.pl $BIN samples.txt reads/overlapped.fastq $primer 8 0 $name T > $rekeyedtabbedfile

3.2 Making a fasta file of all identical sequences (ISU), and an index of those sequences

$BIN/group_gt1.pl $rekeyedtabbedfile $name

3.3 Clustering at 97% identity using usearch (i.e., make OTU), also performs chimera filter singleton reads are excluded

$BIN/usearch -cluster_otus data_$name/groups_uclust.fa -otu_radius_pct 3 -otus data_$name/clustered_otus_usearch.fa

$BIN/usearch -usearch_global $groups_fa_file -db data_$name/clustered_otus_usearch.fa -strand plus -id 0.97 -uc $c95file

3.4 Regenerating the tabbed reads file with each read tagged as to its OTU and ISU group membership

$BIN/map_otu_isu_read_us7.pl $c95file $reads_in_groups_file $rekeyedtabbedfile > $mappedfile

3.5 Making two tables of counts in the analysis directory for OSU and ISU sequences gather the seed sequences for each OTU. Transpose the dataset for ease of import into QIIME

$BIN/get_tag_pair_counts_ps.pl $mappedfile $CUTOFF $name

$BIN/get_seed_otus_uc7.pl $c95file $groups_fa_file analysis_$name/OTU_tag_mapped.txt > analysis_$name/OTU_seed_seqs.fa

Rscript $BIN/OTU_to_QIIME.R analysis_$name

3.6 Using mothur to annotate the OTU sequences against the Silva database (v119)

mothur "#classify.seqs(fasta=analysis_$name/OTU_seed_seqs.fa, template=$TEMPLATE, taxonomy=$TAXONOMY, cutoff=70, probs=T, processors=4)"

$BIN/add_taxonomy_mothur.pl $TAX_FILE analysis_$name/td_OTU_tag_mapped.txt > analysis_$name/td_OTU_tag_mapped_lineage.txt

Appendix C: Recipe of microbiological culture media used in the thesis

Columbia Blood Agar (CBA)
The CBA was made with Difco™ CBA Base, contents were showed in g/L

Pancreatic Digest of Casein	10.0
Proteose Peptone No.3	5.0
Yeast Extract	5.0
Beef Heart, Infusion from 500g	3.0
Corn Starch	1.0
Sodium Chloride	5.0
Agar	15.0

5% volume of sterile defibrinated sheep blood were added to the base after autoclave

Columbia CNA Agar, g/L (Final pH 7.3 +/- 0.2 at 25 °C)

Pancreatic Digest of Casein	12.0
Peptic Digest of Animal Tissue	5.0
Yeast Extract	3.0
Beef Extract	3.0
Corn Starch	1.0
Sodium Chloride	5.0
Agar	13.5
Colistin	0.01
Nalidixic Acid	0.01

5% volume of sterile defibrinated sheep blood were added to the base after autoclave

Chocolate Agar ingredients, g/L (Final pH 7.2 +/- 0.2 at 25 °C)

Proteose Peptone	20.0
Sodium Chloride	5.0
Dipotassium Phosphate	0.50
Dextros	0.50
Agar	15.0

Anaerobic Blood Agar ingredients, g/L

Casein Peptone	15
Sodium Chloride	5
Soy Peptone	5
Yeast Extract	5
L-Cystine	0.4
Vitamin K	0.01
Hemin	0.005
Agar	20

MacConkey Agar ingredients, g/L

Peptone	17.0
Proteose peptone	3.0
Lactose	10.0
NaCl	5.0
Crystal Violet	0.001
Neutral Red	0.03
Bile Salts	1.5
Agar	13.5

Anaerobic Phenylethyl Alcohol (PEA) Agar, g/L

Pancreatic and Enzymatic Digests of Casein	15.0
Papaic Digest of Soybean Meal	5.0g
Sodium Chloride	5.0g
Yeast Extract	5.0g
Phenylethanol	2.5g
Reducing Agents/Peroxide Inhibitors	1.5g
Vitamin K	0.01
Hemin	0.005
Cystine	0.0004
Agar	15.0

5% volume of sterile defibrinated sheep blood were added to the base after autoclave

Curriculum Vitae

Name:	Yige Bao
Post-secondary Education and Degrees:	Sichuan University Chengdu, Sichuan, P. R. China 2005-2013 M.D.
Honours and Awards:	National Scientific Foundation of China, Grant for Young Scientist 81500522, 2016-2018 University of Massachusetts Medical School, Visiting Scholar 2010 May - July
Related Work Experience	Attending Physician West China Hospital, Sichuan University, Chengdu, P. R. China 2013 - present

Publications:

Bao, Y., Al, K. F., Chanyi, R. M., Whiteside, S., Dewar, M., Razvi, H., ... & Burton, J. P. (2017). Questions and challenges associated with studying the microbiome of the urinary tract. Annals of translational medicine, 5(2).

Bao, Y., Welk, B., Reid, G., & Burton, J. P. (2014). Role of the microbiome in recurrent urinary tract infection. Novel Insights into Urinary Tract Infections and their Management, 49.

Di Bella, J. M., Bao, Y., Gloor, G. B., Burton, J. P., & Reid, G. (2013). High throughput sequencing methods and analysis for microbiome research. Journal of microbiological methods, 95(3), 401-414.

Chen, L., Bao, Y., Denstedt, J., & Zhang, J. (2016). Nanostructured bioluminescent sensor for rapidly detecting thrombin. Biosensors and Bioelectronics, 77, 83-89.

Fiuk, J., Bao, Y., Calleary, J. G., Schwartz, B. F., & Denstedt, J. D. (2015). The use of internal stents in chronic ureteral obstruction. The Journal of Urology, 193(4), 1092-1100.

Bao, Y., Bian, H., Qiu, S., Huang, C., Liu, L., Liu, Z., ... & Wei, Q. (2017). Efficacy and safety of medical expulsion therapy for ureteral calculi: a systematic review and network meta-analysis. The Lancet, 390, S104.

Qiu, S., Yang, L., Deng, L., Bao, Y., Tu, X., Li, K., & Wei, Q. (2017). Psychosocial interventions to improve quality of life for men with prostate cancer: a network meta-analysis of 31 randomised controlled trials. The Lancet, 390, S103.

Yang, L., Qiu, S., Deng, L., Bian, H., Tobias, I. C., Huang, C., ... & Wei, Q. (2017). Comparative diagnostic efficacy of biopsy strategies for prostate cancer detection: a Bayesian network meta-analysis of 24 randomised controlled trials. The Lancet, 390, S28.

Chen, W., Bao, Y., Chen, X., Burton, J., Gong, X., Gu, D., ... & Bao, L. (2016). Mycobacterium tuberculosis PE25/PPE41 protein complex induces activation and maturation of dendritic cells and drives Th2-biased immune responses. Medical microbiology and immunology, 205(2), 119-131.