Fall 2022

Polyomavirus JCPyV and Its Role in the Urinary Tract

Rita Lauren Mormando

LOYOLA UNIVERSITY CHICAGO

POLYOMAVIRUS JCPyV AND ITS ROLE IN THE URINARY TRACT

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN BIOINFORMATICS

BY

RITA L. MORMANDO

CHICAGO, IL

AUGUST 2022

ACKNOWLEDGEMENTS

I would like to thank Dr. Catherine Putonti for her incredible guidance and support throughout the past five years and for being the reason I fell in love with viruses and research. I would like to thank the members of the Putonti Lab for their assistance and immense support. I want to thank Genevieve Johnson for her invaluable advice and support throughout the duration of my time in lab and in working on this thesis. I would like to thank the members of my committee, Dr. Alan Wolfe, and Dr. Swarnali Banerjee, for lending their knowledge and assistance to my project. Finally, I would like to thank my parents and family for their love, advice, and support throughout my academic career and my journey through life. There is nothing short of immense pride I feel to be their daughter, sister, fellow “labbie”, mentee, and friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Polyomaviruses are the smallest closed-circular supercoiled double-stranded viruses found in the human microbiota. The polyomavirus JC (JCPyV) is most commonly found within the urinary tract, and prior studies estimate that 20-80% of older adults carry JCPyV. In very rare cases, JCPyV leaves the kidneys, causing progressive multifocal leukoencephalopathy. However, the role of JCPyV within the urinary tract remains an open question. In a prior study conducted by our group, the bladder microbiota of females with and without overactive bladder symptoms (OAB) were sequenced. Interestingly, JCPyV was only detected in females with OAB; none of the control ("asymptomatic") microbiota contained JCPyV. However, the sample size for this study was small (n=30). This thesis is a multidisciplinary approach to explore the presence and prevalence of polyomaviruses in the urinary microbiome (urobiome). In a bioinformatic investigation of JCPyV and BKPyV, a closely related polyomavirus to JCPyV, 165 publicly available urinary virome and urinary metagenome data sets were mined for these two polyomaviruses. Sequence diversity between JCPyV and BKPyV genomes was explored to design a new primer pair to uniquely identify JCPyV in urobiome samples. Using these ultra-specific JCPyV primers, 190 urine samples, including 99 from females with OAB, 33 from females with UTI, and 58 from females without lower urinary tract symptoms, were screened for JCPyV to assess the prevalence of the virus as well as to assess the association of JCPyV with symptom status, age, and race/ethnicity. Additionally, the urobiome of JCPyV+ individuals was sequenced in an effort to identify associations between JCPyV and bacterial taxa. We found no

associations between JCPyV presence or abundance and any of the factors when tested individually. However, some associations were found when some of the factors were considered together in predicting JCPyV prevalence. Additionally, both our bioinformatic and molecular survey suggests that JCPyV is less prevalent in the female population than previously thought.

CHAPTER ONE

INTRODUCTION

Urinary Microbiome

The human microbiota, the community of microorganisms (e.g., bacteria, archaea, and viruses), is vast, diverse, and present in nearly all major organ systems. The discovery of the microbial communities of the human body was enabled by the development of high-throughput sequencing technologies, which has made it possible to catalog the inhabitants of the human microbiota. The genomic sequences of these microbiota, referred to as the microbiome, have provided a glimpse into the diversity and distribution of microorganisms throughout the body (1).

Of the previously characterized organ systems, the urinary tract of asymptomatic ("healthy") individuals was once thought to be "sterile," and the detection of any microorganism was thought to be a clear sign of disease (2). This belief was primarily based on the observed absence of bacterial species in the urine of healthy individuals via standard clinical tests (3). Previous bacterial identification methods relied on culturing the bacteria, but not all bacteria can effectively grow under laboratory conditions. Thus, using high-throughput sequencing technologies, bacterial DNA in the urinary tracts of healthy people was discovered (4, 5). Subsequent metaculturomic studies confirmed that this bacterial DNA was representative of living bacteria in the urinary tract (6–8). Numerous studies have repeatedly found microorganisms within the urinary tract of healthy individuals, the urobiome (9).

While the majority of studies of the urobiome have focused on the bacterial constituents, these communities can also include fungal and archaeal taxa (10–13). Moreover, viruses – including viruses that infect bacterial species (bacteriophage or phage) and viruses that infect human cells – are abundant within the urobiome (14, 15). In fact, it has been known for quite some time that human-infecting viruses are shed into the urine (16). There are four main types of eukaryotic viruses that inhabit the urinary virome (the viral fraction of the urobiome): Adenoviruses, Anelloviruses, Papillomaviruses, and Polyomaviruses (16). Adenoviruses are double-stranded DNA viruses that mostly cause limited, localized infections in the respiratory tract, but can develop into severe and potentially fatal infections in immunocompromised individuals (17, 18). The most common anellovirus found in the urinary tract is Torque teno virus (TTV), a single-stranded DNA virus that has been largely studied in the context of immunodeficiency in transplant recipients (19). Human papillomaviruses (HPV) are mostly double-stranded DNA viruses and are known for infecting the genitourinary tract and causing cervical cancer (20). However, the most prevalent eukaryotic viruses in urine samples are two polyomaviruses: human polyomavirus BK and human polyomavirus JC (21).

Polyomaviruses

Polyomaviruses are the smallest known double-stranded DNA viruses and are abundant in the human microbiota (22). Human polyomavirus BK (BKPyV) and human polyomavirus JC (JCPyV) were first isolated from a urine sample in 1971 (23). JCPyV and BKPyV are believed to be benign members of the urobiome, producing persistent, asymptomatic infections of the kidneys (24, 25). JCPyV prevalence within the population ranges from 20 to 80% (26–28). The variation of reported incidence of JCPyV is undoubtedly reflective of the observed increase in

seroprevalence with age (24). BKPyV is estimated to be far more common than JCPyV, with upwards of 90% of individuals infected by age 10 (29, 30). The majority of research into JCPyV and BKPyV has focused on the rare occurrence in which they cause severe diseases in immunocompromised hosts. Activation of BKPyV leads to nephropathy in renal transplant patients (31), while JCPyV can cause progressive multifocal leukoencephalopathy (PML) (32).

Both polyomaviruses have a genome length of approximately 5100 bp and are 42-45 nm in size with a 72-capsomere icosahedral capsid (33). The genomes of BKPyV and JCPyV share 75% nucleotide sequence similarity and encode for the same six proteins. These proteins include three structural capsid proteins (VP1, VP2, and VP3), as well as three non-capsid regulatory proteins (large tumor antigen (LTag), small tumor antigen (STag), and agnoprotein).

The early coding region in both BKPyV and JCPyV encodes primarily two major proteins: LTag and STag. The LTag protein is the main regulatory protein, as it is required for the replication of the viral genome (34). The main activity for this protein is to bind to two key cellular proteins (pRB and p53 family tumor suppressor proteins) and to block their functions, which effectively prevents the induction of cell death (33, 35). As polyomaviruses do not have their own viral DNA polymerase, the LTag protein will also bind to the host DNA polymerase complex and, along with utilizing its own helicase activity, will help facilitate DNA replication (33). On the other hand, the STag protein is used to effectively bind the catalytic and regulatory subunits of protein phosphatase 2A (PP2A), which inactivates their functions and drives cells into the S-phase of the cell cycle where DNA replication occurs (34, 35). It is worth noting that LTag and STag are universally expressed by all polyomaviruses from their early coding regions (36).

The late coding region in both polyomaviruses encodes the remaining four proteins: VP1, VP2, VP3, and the agnoprotein. The major structural protein, VP1, forms the outer shell of the capsid and facilitates entry into the host cell by attaching the virus to the cell surface receptors, which allows for viral assembly to occur (23, 36). Once the virus has entered the nucleus of the host cells and viral assembly happens, the two other capsid proteins, VP2 and VP3, are added to and arranged on the replicated viral genomes to package the viral DNA into the generated capsids (34, 36). The final shared protein is the agnoprotein. This 71-amino acids protein plays an important role in the regulation of the virus replication cycle by interacting with LTag (35). More importantly, it was recently discovered that the agnoprotein in JCPyV is able to form highly stable dimers and oligomers via its α-helix domain, which is the region that plays a critical role in the function and stability of the protein (37). The study of this region and this protein more extensively could be used for drug discovery purposes against JCPyV-induced progressive multifocal leukoencephalopathy (34).

JCPyV was initially isolated from a patient with progressive multifocal leukoencephalopathy (PML), a rare but often fatal demyelinating disease. Research has determined so far that after primary infection - during childhood - JCPyV becomes dormant but persists in the kidneys and urinary tract. It is also known that JCPyV persists in other tissues (i.e., spleen, tonsils, and bone marrow) (37–40). JCPyV may “reactivate” via genome rearrangement under conditions of immunosuppression (i.e., AIDS patients) and cause PML. PML has been increasingly diagnosed in patients who are receiving immunosuppressive therapies for autoimmune disorders (i.e., multiple sclerosis, systemic lupus erythematosus, and rheumatoid arthritis) (41, 42). Occurrences of PML have also been reported in individuals with MS who

have been treated with natalizumab (Tysabri), dimethyl fumarate (Tecfidera), fingolimod (Gilenya), and ocrelizumab (Ocrevus). As there is no cure for PML, the only treatment includes the immediate termination of the current medication. While the mechanisms for viral replication and disease are well studied (see review (24)), the most common strain of the virus – the "archetype" – detected in the urine is widely considered nonpathogenic (43).

With the exception of the rare cases of nephropathy (BKPyV) and PML (JCPyV), the role that these two polyomaviruses play within the urinary tract is unknown. A previous study showed the presence of JCPyV in females with overactive bladder (OAB) symptoms, indicating that the presence of JCPyV may be associated with lower urinary tract symptoms (LUTS) (15). Two previous case studies have shown associations of BKPyV with LUTS in children. The first study showed the first association of non-hemorrhagic cystitis with BKPyV, as they found that excreted cells contained changes to the nucleus that were suggestive of human polyomavirus infection. The 3½-year-old boy was in generally good health and had no recognized immune impairment, but the data collected and reported in this study showed that BKPyV was the cause of the child's acute cystitis (44). The second case study focused on a healthy 5-year-old boy where, again, a case of BKPyV cystitis was reported. This study also looked at the abnormal urothelial cells that were collected from the child and found that the "large hyperchromatic intranuclear inclusions" were consistent with the other reported case of BKPyV infection (45). In another study, potential associations of JCPyV and LUTS were tested in males and found that JCPyV is "variably associated" with LUTS (46). Thus, the putative role of polyomaviruses in urinary tract health/symptoms remains an open question.

Viral Genomics and Metagenomics

High-throughput sequencing technologies have been instrumental in identifying and cataloging cellular organisms (i.e., bacteria, archaea, fungi) in complex communities, including uncultivated species, through targeted sequencing of gene markers. However, studying viral communities faces unique challenges. First, unlike cellular organisms, there is no universal marker gene(s) in viruses. Thus, shotgun metagenomics, i.e., sequencing all of the DNA present in a sample, is the only way for researchers to study complex viral communities. Second, viral genomes are prone to a high degree of genetic diversity. Sequencing of new viral communities routinely uncovers novel genes and sequences (47). Third, viral sequence databases do not include sufficient representation of viral species. Compared to eukaryotic and prokaryotic organisms, only a small fraction of viral genomes have been sequenced and characterized (48). Fourth, even when sequencing purified viral samples, metagenomic sequencing data often includes non-viral (host) DNA. This is further complicated by the fact that viral genome sequences are often orders of magnitude smaller than host cells or other cellular organisms in the sample.

Shotgun metagenomics has been instrumental in understanding the urinary virome. However, relatively few studies have been conducted to date. There have been four studies that have focused on sequencing the virome specifically. A study conducted by Santiago-Rodriguez et al. (49), analyzed whether the urinary virome had an association, or was affected by, urinary tract infections. Two studies, Sigdel et al. (50) and Rani et al. (51) aimed to characterize the urinary virome in kidney transplant patients to understand the differences in viral composition in patients after transplantation. Finally, a study conducted by Thomas et al. looked at the

relationship between JCPyV presence and LUTS in males by metagenomically sequencing urine samples (46). The diversity of viral species within the urinary tract has also been studied by sequencing the entire urobiome. Moustafa et al. published a study in which the urinary microbiome was sequenced in order to gain insight into the urinary flora of individuals with UTIs, including viral and cellular species (52). Another study, conducted by Garretto et al. (15), aimed to reconstruct the genomes of viral populations within the low-biomass bladder microbiota of females with and without OAB symptoms. There are also several studies that have metagenomic sequences publicly available but details about the analyses have yet to be published. While unfortunate for reproducing studies, it demonstrates there are efforts being made to make data readily available for the public. Interestingly, all of the published studies have detected the presence of polyomaviruses within the urinary virome/metagenome (15, 49–53), which brings us to the current state of the field.

Scope of Thesis

The work presented here is a multidisciplinary approach to exploring the presence and abundance of polyomaviruses in the urobiome. Chapter 2 presents a bioinformatic investigation of JCPyV and BKPyV. 165 publicly available urinary virome and urinary metagenome data sets were mined for polyomaviruses. Sequence diversity between JCPyV and BKPyV genomes was explored to design a new primer pair for uniquely identifying JCPyV in urobiome samples. In Chapter 3, 190 urine samples from female participants were screened for JCPyV in order to assess the prevalence of the viruses as well as assess the association of JCPyV with LUTS, age, and race/ethnicity. Conclusions are presented in Chapter 4.

CHAPTER TWO

DETECTING POLYOMAVIRUSES IN METAGENOMES

Introduction

While early human microbiome studies were focused on the bacterial members of these communities, recent focus has shifted to viral members, which have been found to play a role in human health. For example, viruses have been studied in the context of potential treatments for various ailments, such as inflammatory gut diseases (i.e., IBD). A recent study isolated and analyzed gut virome samples from people with Crohn's disease or ulcerative colitis to see whether changes in gut viruses contribute to inflammation or are collateral damage caused by inflammation (54). They found that transplanting viruses from people with healthy guts into tissue cultures containing viruses from people with IBD suppresses the inflammation effect of the disease. Thus, this study – among many others – emphasizes the notion that our viromes are an important contributor to human health. As such, characterizing the taxonomic and functional diversity of these microbial communities is achieved by using metagenomic sequencing.

In contrast to many organ systems, the urinary tract of healthy individuals was once thought to be "sterile," a belief primarily based on the observed absence of bacterial species and recently debunked using these new sequencing techniques and new culture methods (55). While the urinary tract was often referred to as sterile, it was well known that viruses were shed into urine. Over 50 years ago, polyomaviruses were reported within the urinary tract (56).

Polyomaviruses are the smallest known double-stranded DNA viruses and are abundant in the human microbiota (22). Within the urinary tract, two polyomaviruses have been identified: JC virus (JCPyV) and BK virus (BKPyV). The genomes of these two viruses are 75% similar and they encode for the same six proteins (LTag, Stag, VP1, VP2, VP3, and agnoprotein). Specifics of these proteins are discussed in Chapter 1. In most cases, JCPyV and BKPyV are benign members of the urinary microbiota, producing persistent, asymptomatic infections of the kidneys (23).

Estimates of JCPyV prevalence within the population range from 20 to 80% (26–28). Studies have found that the incidence of JCPyV is low in younger populations and high in the elderly (21, 26–28). This might explain the varied range of detection as seropositive rates of JCPyV increase with age (21). While JCPyV can be acquired by vertical transmission, from mother to child (57), it is believed that the main route of infection is ingestion of contaminated food and/or water; JCPyV is stable in both urine and sewage (see review: (25, 58)). Contaminated water sources likely explain the observations that individuals in the same geographic region harbor related JCPyV strains (59, 60). BKPyV is estimated to be far more common than JCPyV, with an estimated 65-90% of individuals infected by age 10 (29, 30).

Estimates of the prevalence of these two polyomaviruses are frequently based upon amplification-based surveys, and numerous assays/protocols have been designed to detect JCPyV and BKPyV, including methods distinguishing between the two (61–84). Although currently there are just a handful of metagenomic studies of the urobiome, JCPyV and/or BKPyV have frequently been reported (15, 49, 51–53). Given the disparate prior estimates of JCPyV incidence and the known similarity of the JCPyV and BKPyV genomes, we initiated an

investigation into the methods used for JCPyV detection. All publicly available JCPyV genomes were examined, identifying sequences that can serve as genomic signatures for this viral species. Based upon this bioinformatic work, we developed and tested a new PCR-based assay for JCPyV detection. Furthermore, we revisited all publicly available urobiome data sets to specifically hunt for JCPyV. Together, our results suggest that JCPyV is far less prevalent than previously thought.

Methods

Metagenomes: Data Collection

A total of 165 publicly available raw reads from metagenomic sequencing of the urinary microbiome were identified from available literature and retrieved from NCBI's Short Read Archive (SRA) or iMicrobe (49, 51, 52). Sixty-five of the samples were identified in SRA, although no corresponding paper in the literature was found. Four samples were from portable urinals. A summary of these samples is shown in **Table 1**.

Study	Accession No.	# Samples	# Reads Total
Kidney Transplant Virome (51)	PRJEB28510[1]	27	157.5 M[3]
Pulmonary Tuberculosis Urinary Microbiome (unpublished)	PRJNA431965[1]	3	9.5 M[4]
Microbial Metagenome of UTI (52)	PRJNA385350[1]	49	532.0 M[4]
Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	PRJNA587166[1]	62	70.0 M[3]
Virome in Association with UTI (49)	cobian9680[2]	20	14.5 M[3]
Portable Urinal Microbiome (unpublished)	PRJNA399057[1]	4	228.3 M[4]

Table 1. Summary of raw read data sets evaluated in this study. (1) Accession number for NCBI BioProject, (2) Accession number for iMicrobe, (3) Sequencing performed for viral community only, (4) Sequencing performed for the bacterial and viral community.

Metagenomes: Mapping

Each raw read data set was mapped to reference genomes using Bowtie2 (v2.3.2) and visualized through Geneious Prime (Biomatters, Ltd., Auckland, NZ) (85). Reference genomes included the JCPyV RefSeq (accession no. NC_001699.1) and the BKPyV RefSeq (accession no. NC_001538.1). Geneious Prime provided details with regard to the number of reads mapped per data set. For each sample, evenness and completeness of coverage were assessed. Mapped genomes that exhibited even and complete coverages to either JCPyV or BKPyV were considered high confidence matches; the data indicated that the reads did not map to only part of the genome or only a specific gene, rather the entire polyomavirus genome **(Figure 1)**.

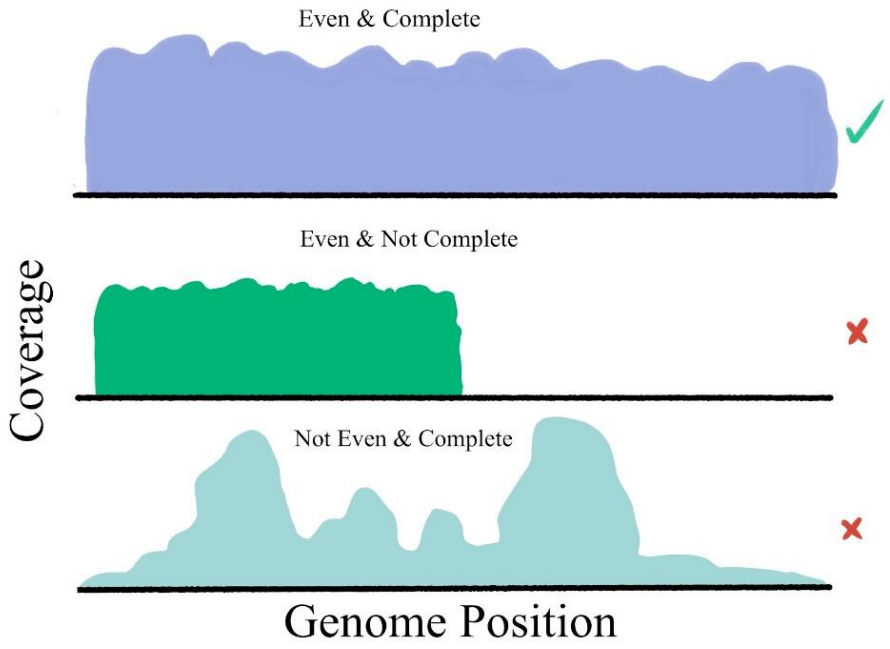

Figure 1. Representation of coverage graphs

JCPyV PCR Primer Design

JCPyV-specific primers were designed to target this virus after conducting a bioinformatic study of the evolution of this viral taxon. All publicly available, complete JCPyV

genomes were retrieved from GenBank and aligned via Clustal-Omega (86). JCPyV genomes deposited in an alternative orientation and/or different start site position were manually corrected. Partial genomes (listed as complete genomes) were identified and manually removed. In total, 620 sequences were retained (**Appendix A**). The final set of sequences were aligned again using Clustal-Omega; a tree was derived using FastTree v2.1.11 through Geneious Prime v2019.1.1 (Biomatters Ltd., Auckland, NZ) and visualized using iTOL v4 (87, 88). From the JCPyV genome alignment, regions of sequence conservation were identified using Geneious Prime. These regions were further investigated as possible hybridization sites for JCPyV-specific primers. The regions were queried against the nr/nt database via blastn to determine their specificity to JCPyV and sensitivity to JCPyV genome variation. The sequence regions with highest specificity were imported into Primer-BLAST to assess their amplicon length, similarity of melting temperature, and their likelihood to either self-hybridize or form a primer-dimer (89).

As JCPyV and BKPyV have a 75% sequence similarity, we wanted to verify that the primers created for JCPyV would not also amplify any BKPyV region. To do this, the primer sequences we created were queried against BKPyV sequences in nr/nt (specifying the organism parameter to “Human polyomavirus 1” [BKPyV]). Primer sequences that did not match perfectly to any BKPyV sequence or were several mismatches away, particularly at the 3’ end, from any given BKPyV sequence were considered further. One primer pair was identified meeting all the aforementioned criteria, JCPyV-2l: 5’-CAG GAA AGT CTT TAG GGT C-3’, and JCPyV-2r: 5’-CCC TGT TTA ATG TGC ATG-3’.

Collecting Primer Sequences from the Literature

PCR primers and Taqman probes used to detect JCPyV DNA were collected from a review of literature (27, 62–84, 90–94). **Appendix B** lists the primer and probe sequences, as well as the papers from which they came. Each primer/probe sequence was independently mapped to the JCPyV RefSeq genome (Accession no. NC_001699.1) via Bowtie2 v.2.3.5 through Geneious Prime.

PCR Detection & Sequencing the Positive Control

Additionally, we sequenced the urinary microbiome of one urine sample in our collection, which was confirmed to be PCR-positive for JCPyV. This urine was collected via transurethral catheterization from a female with overactive bladder (OAB) as part of a previous IRB-approved study (95). DNA was extracted using a phenol-chloroform protocol with a starting volume of 500uL urine. Briefly, a 1:1 ratio of urine to phenol-chloroform was vortexed and then centrifuged for 1 minute. The top layer of the mixture was then pipetted into a new microcentrifuge tube and centrifuged again. This process was repeated until no protein was visible. An equal volume of chloroform was added to the tube, vortexed, and centrifuged. The top layer was then removed and sodium acetate (1:10 by volume) and 100% ethanol (2:1 by volume) were added. The solution was vortexed and put on ice for 30 minutes. The sample was then centrifuged for 10 minutes, the supernatant was decanted, and 750uL of 70% ethanol was added. The solution was then centrifuged for 2 minutes, after which the supernatant was decanted leaving a pellet to dry at room temperature. 50uL of TE buffer was added and the pellet was resuspended by gentle vortexing.

PCR detection was performed with the designed primer pair as listed above (JCPyV-2l and JCPyV-2r) using the thermal cycler conditions listed in **Chapter 3 Table 5**. The 50uL PCR reaction consisted of 0.5uL of each primer (10mM), 5uL of extracted DNA, 19uL of nuclease-free water, and 25uL of Promega GoTaq® Master Mix. Primers were synthesized by Eurofins (Huntsville, AL).

The extracted DNA was sent to the Microbial Genome Sequencing Center (Pittsburgh, PA, USA). Libraries were prepared using a method based upon the Illumina Nextera kit (Illumina, Inc., San Diego, CA, USA) and sequenced on the Illumina NextSeq 550 platform. Sequencing produced 93.5 M reads of read length 150 nucleotides. Raw reads have been deposited in SRA; accession no. SRR13199001.

Assessing Specificity of JCPyV Reads and Primers in Metagenomes

Predicted JCPyV reads were queried against the nr/nt database using blastn as well as against the nr/nt database, limiting the search to “Human Polyomavirus 1” (BKPyV), via discontiguous blast. Five urinary microbiome samples (ERR926151, ERR926109, ERR926113, ERR926116, ERR926117), which we previously reported as JCPyV positive (15), were also screened; primers were aligned against the consensus sequences derived from these microbiomes.

Results

Mining for JCPyV using Publicly Available Metagenomes

To understand the frequency at which JCPyV is found in the urinary microbiome, we retrieved 165 publicly available urinary metagenomic data sets from NCBI’s SRA database and iMicrobe; these were all the publicly available urinary metagenomic data sets as of June 2019.

Prior analyses of some of these metagenomes indicated that JCPyV was present (49, 51, 52). Using these raw read data sets, we specifically looked for JCPyV using the JCPyV RefSeq sequence (5,130 bp). The methods for doing so are detailed in the Methods section. Briefly, raw reads were mapped to the RefSeq and evaluated for their completeness and evenness of coverage of the RefSeq (**Figure 1**). Of the 165 data sets examined, raw reads mapped to the JCPyV RefSeq sequence for 59 of the samples (minimum number of reads mapped: 1; the maximum number of reads mapped: 336,004; the average number of reads mapped: 36,734). Details can be found in **Appendix C**.

Looking closer at the mapped reads, however, we saw that the genome coverages of these reads were unequal; some genes were mapped with significantly higher coverage values, whereas other genes had few or no reads mapped to them (**Figure 2A**). Given this observation, we restricted our analysis to samples in which the mapped reads were uniformly distributed across the JCPyV genome (**Figure 2B**). This stipulation necessitated that the JCPyV genome must be present in its entirety; removing from consideration samples that only contained evidence of part of the genome and/or a subset of genes. Thus, we can determine the presence of JCPyV in these samples with high confidence. From the 165 metagenomes examined, we concluded with high confidence that JCPyV was present in only three samples: accession numbers ERR2798125, ERR2798126, and SRR6519218 (**Table 2**). The first two samples were from the Kidney Transplant Virome data set and the third was from the Pulmonary Tuberculosis Urinary Microbiome data set, which has not been published (51).

The remaining 56 data sets that failed the uniformity test when mapped to JCPyV were then mapped against the BKPyV RefSeq sequence (5,153 bp). Here we found that all 56 reads

mapped to the BKPyV genome with uniform coverage (**Figure 2C**), indicating that BKPyV was present in the sample rather than JCPyV. Mapping these data sets against the BKPyV RefSeq was an important way to accurately distinguish between the two polyomaviruses. After performing this step, we found instances within these data sets where the literature stated the detection of JCPyV within their metagenomes but it was actually the detection of BKPyV (49, 51, 52), according to the evenness and completeness of the mapped genome on the reference genome.

Accession No.	Total # of reads	# Reads mapped to JCPyV	% Reads mapped to JCPyV
ERR2798125	13,710,307	90,043	0.66
ERR2798126	15,725,590	319,807	2.03
SRR6519218	3,304,590	21,724	0.66

Table 2. Data sets that fully mapped to the JCPyV reference genome.

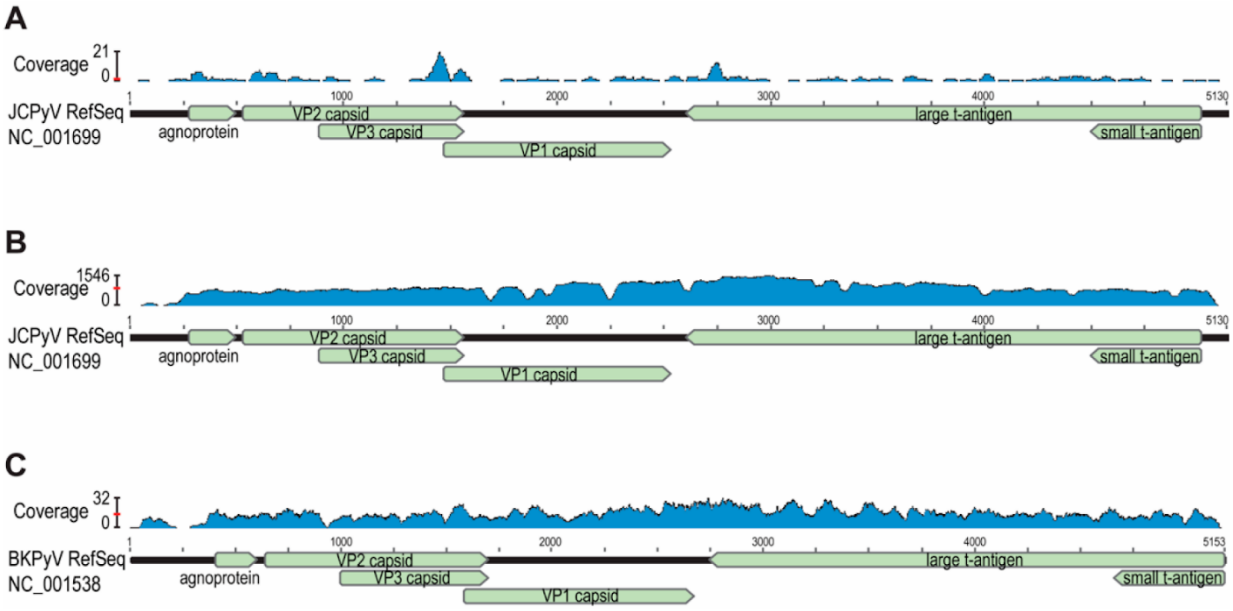

Figure 2. Visualization of the raw read coverage of the JCPyV and BKPyV reference genomes. (A) Coverage of the JCPyV genome for sample ERR2798128 reads; average coverage (indicated by a red bar) was ~1 reads/nucleotide and was concentrated on genomic regions conserved between JCPyV and BKPyV. (B) Coverage of the JCPyV genome for sample ERR2798125 reads; average coverage (indicated by red bar) ~900 reads/nucleotide. (C) Coverage of the

BKPyV genome for sample ERR2798128 reads; average coverage (indicated by red bar) was ~15 reads/nucleotide.

Uniquely Identifying JCPyV

Publicly available JCPyV genomes (n=620) were retrieved from GenBank and aligned to derive a phylogenomic tree (**Figure 3**). Sequence similarity amongst the 620 genomes ranged from 89.8% to 100% nucleotide sequence identity. SNPs were identified in all six of the JCPyV coding regions with 192 bases exhibiting a minimum variant frequency of 25%. These variable sites were distributed among all six coding regions. The large t-antigen coding region contained the most variable sites: 116 major (>25% frequency) SNPs (5.59% of the coding region).

Figure 3. Phylogenomic tree of 620 JCPyV complete genomes.

From this alignment, conserved regions in the JCPyV genome were identified for PCR primer design. Each putative primer sequence was queried against NCBI's nr/nt database and

manually inspected to ensure specificity (100% nucleotide sequence identity) to JCPyV. Given the relatedness of the two species and the frequency in which both are detected in urine samples, specificity was further assessed by querying each putative primer sequence against BKPyV sequences. Primer sequences passing this specificity requirement were then examined for melting temperature, size of amplicon, and likelihood to self- or cross-hybridize. After applying these filters to our search, one primer pair was selected: JCPyV-2l and JCPyV-2r. The sequences for this primer are listed in the Methods. JCPyV-2l hybridizes within Jvgp5, the LTag, while JCPyV-2r hybridizes within Jvgp6, the STag (**Figure 4**, blue bar). The location of these primers was compared to previously published JCPyV primer/probe sequences. While the region targeted by JCPyV-2l is in the same vicinity as other JCPyV primer/probe sequences, the region targeted by JCPyV-2r has not been used as a primer/probe sequence. Generally speaking, very few primers/probes used by other studies have targeted the STag coding region (**Figure 4**).

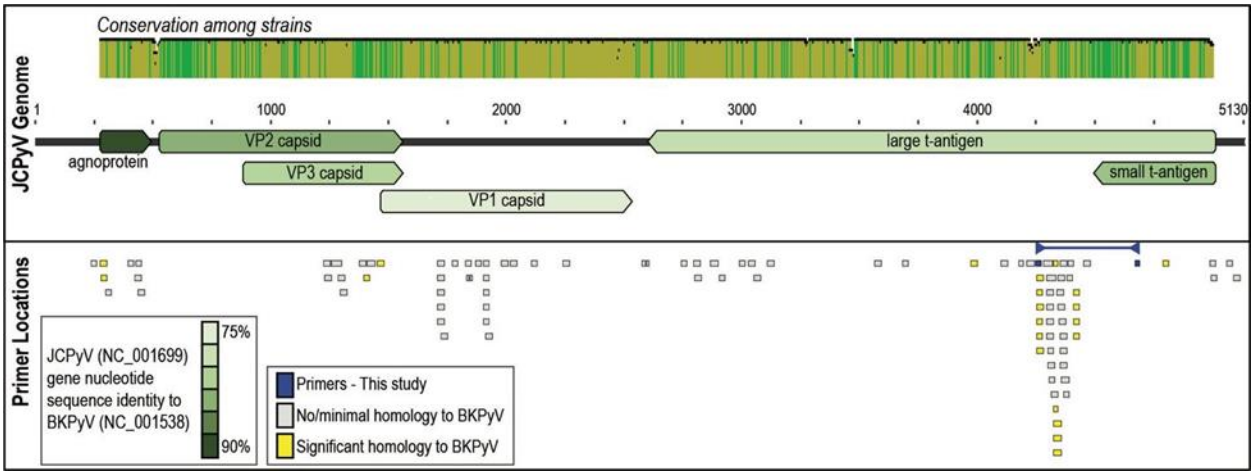

Figure 4. JCPyV genome similarity among variants and PCR primer/probe locations for JCPyV detection. The top panel shows the coding regions within the JCPyV genomes and the conservation of the nucleotide sequences among the 620 JCPyV genomes. Each coding region is color-coded according to their nucleotide sequence similarity with the related polyomavirus BKPyV (see bottom left legend: dark green = most conserved; light green = less conserved). The bottom panel indicates the placement of primers/probes described in other studies (gray or yellow boxes) and the primers described in this study (blue, with the amplified region

highlighted in blue). Primer/probe sequences from other studies are colored gray if they are not expected to hybridize to BKPyV or yellow if they are expected to hybridize with BKPyV.

Evaluating the Sensitivity of JCPyV-Specific Primers

We conducted shotgun sequencing for one PCR-positive sample to confirm that JCPyV was in fact present. One pair of reads (of the 46.7 million pairs) mapped to the JCPyV RefSeq. These reads were then queried against the nr/nt database via blastn and had 100% query coverage (length 150 bp) and 100% sequence identity to JCPyV sequences. Both reads mapped to the large t-antigen coding region. These reads were also compared to BKPyV sequences. The two reads had 99% query coverage and 75% sequence identity and 60% query coverage and 80.22% sequence identity, respectively. This suggests that these reads are in fact representative of JCPyV and not BKPyV. To further evaluate these primers, we mapped the two primer sequences to *de novo* JCPyV assemblies from 5 urinary microbiome samples and found that both primers were identical matches to the JCPyV genomes.

Discussion

This study provides two key insights into the detection of JCPyV in urine samples: 1) the challenges of definitively identifying JCPyV in urine amidst other polyomaviruses, namely BKPyV, and 2) the difficulty in distinguishing between polyomaviruses in metagenomic data sets. Both challenges must be addressed if we are to accurately determine the prevalence of JCPyV in the population and begin to explore what role (if any) JCPyV plays in urinary health.

Metagenomic studies of the urinary microbiome have repeatedly suggested that JCPyV and BKPyV are both present (49, 51, 52). Our analysis of these data sets finds 3 samples containing JCPyV and 56 samples containing BKPyV. BKPyV has been well documented as more prevalent than JCPyV, with seroprevalence rates up to 90% (see review: (32)). As

mentioned in Chapter 1, the current estimates of JCPyV vary widely, from 20 to 80% (26–28). However, this estimated range comes from PCR-based or serotype-based assays where the protocol may not be adequately optimized for specific JCPyV detection. Furthermore, the studies contributing to this estimated range includes surveys with differing populations (e.g., age, sex). These factors may skew the observable detection of JCPyV as there are multiple studies that suggest certain demographics have a higher chance of JCPyV presence than others (24, 27, 51, 59, 61, 73). The data sets examined in this chapter are also collected from varying populations.

Detection of JCPyV in the metagenomic data sets examined here is significantly lower (~1.8%) than prior estimates. It is important to note that the data sets examined here were not explicitly looking for JCPyV. Similarly, in our own prior urobiome study (15), we were not focused on identifying JCPyV. Because polyomavirus detection was not the focus of these sequencing efforts, protocols for DNA extraction may not have been optimized for polyomavirus DNA extraction. This is one of the limitations of conducting meta-analyses on publicly available data sets; several confounding variables (i.e., extraction methods) may skew interpretation. Thus, meta-analyses necessitate the availability of detailed metadata to draw conclusions. With regards to urobiome data sets, there are very few available to the public and three of the six studies do not have accompanying literature. Overall, there is insufficient metadata information (particularly with respect to individual sex, age, and health status). Thus, we cannot confidently say that our observed incidence of ~1.8% is representative of the human population.

Bioinformatic tools designed for analysis of high-throughput studies often lack the sensitivity required to differentiate between closely related species. Tools designed to recognize viral sequences in metagenomic data sets rely on the presence of hallmark genes or best BLAST

hits to make taxonomic calls (96–100). As was mentioned in Chapter 1, there are no genes conserved among all viruses and viruses themselves are vastly underrepresented in sequence databases. Yet, given the high sequence homology between the two polyomavirus genomes, BKPyV and JCPyV, miscalls are not surprising. Here, we have shown that these distinctions can only be reliably made by considering each species independently, examining the evenness of read coverage as well as coverage of, and homology to, unique regions. For instance, the urinary metagenome produced as part of this study, which we found to be PCR-positive for JCPyV, only included two reads that mapped to JCPyV. These reads exhibit greater homology to JCPyV than to BKPyV sequences. Thus, while metagenomic sequencing has significant potential for virus detection, rigorous bioinformatic interrogation is necessary to correctly identify and distinguish between closely related species.

Due to the genomic similarities between BKPyV and JCPyV, PCR-based assays designed to detect JCPyV in urine are only reliable if they are unlikely to hybridize to BKPyV or other DNA in the sample. Thus, selecting primer sequences with a high degree of specificity to JCPyV (i.e., the STag coding region) is critical, particularly given the high incidence of BKPyV in urine (101). As **Figure 4** shows, several primer/probes were expected to hybridize with BKPyV, although it is worth noting that on several occasions, an approach was taken such that both BKPyV and JCPyV DNA were amplified and then distinguished via a nested PCR, probe, or enzymatic digest, e.g. (63). Nevertheless, simple BLAST queries showed that some primers were likely to hybridize, tolerant of mismatches. By examining the divergence within the JCPyV genome, as represented by more than 600 genomes examined here, we have targeted regions of the JCPyV genome that are conserved and unique, as well as sensitive to the variation that exists

within the species. The challenge of designing primers that are both sensitive and specific to their target is well known (see review: (102)). Simple BLAST queries do not suffice in testing such primers, a fact acknowledged by BLAST itself (89). We were able to detect the presence of JCPyV in urines from our collection using our PCR primer pair, even when JCPyV was in very low abundance.

While the urine sample sequenced here was PCR positive, shotgun metagenomic sequencing only produced two JCPyV reads. Thus, it would not pass our threshold for a high confidence call for JCPyV presence. While the coverage requirements introduced here can produce high confidence predictions, samples with JCPyV may be missed, as would be the case with our own sample. By using specific and sensitive primers, PCR can detect low-levels of DNA. Thus, when testing for specific viruses within a complex community, such as the urobiome, PCR provides a more accurate diagnostic than shotgun metagenomics. In a previous study investigating JCPyV prevalence in male urobiomes, a statistically significant association between JCPyV and urinary symptoms was identified by metagenomic sequencing (46). However, when a PCR-based assay was employed, this association was no longer supported; PCR detected JCPyV in samples that were deemed JCPyV negative by metagenomic analysis (46).

Thus, the bioinformatic approach presented here is best suited for detecting a virus that is abundant within a sample. Our further inspection of the JCPyV reads, however, provides confidence that the PCR signal is due to the presence of JCPyV and not BKPyV or other viral species. Even with deeper metagenomic sequencing of the sample sequenced here, it is likely that the complete genome would not be reconstructed. Since our bioinformatic approach requires

complete genomes, it would not characterize this sample as JCPyV+. This is important for future studies to keep in mind; low level taxa may not be sufficiently abundant to be accurately detected via shotgun metagenomics, even when deep-sequencing is performed. This can have immediate implications for diagnostic applications of shotgun metagenomic sequencing.

CHAPTER THREE

INVESTIGATING CAUSE & CONSEQUENCE OF JCPyV IN URINE

Introduction

As seen in Chapter 2, the misidentification of the two urinary polyomaviruses is common due to their significant sequence similarity. This may be contributing to the range of estimates in the literature regarding the prevalence of JCPyV in the population (24). In clinical practice, JCPyV is primarily detected in urine using PCR-based assays. These include "traditional" PCR as well as nested PCR and Taq-Man PCR. As shown in **Chapter 2 Figure 4**, the large-T antigen is a frequent target of PCR primers. Also shown in this figure, very few primer/probe sequences exhibit sequence specificity to JCPyV, and some primers (or primer/probe sets) are designed to amplify both BKPyV and JCPyV.

While the majority of research into JCPyV has focused on its association with PML, very few studies have explored the effects of the chronic form, which is far more prevalent (43, 101), on human health. In a study of patients with systemic lupus erythematosus, JCPyV was found to be associated with arthritis/arthralgia in these individuals (103). While JCPyV has been ruled out as contributing to gastric cancer and other gastroduodenal diseases (104), an association between it and tumor development has been debated (see review: (105)). Nevertheless, JCPyV has been shown to be associated with male infertility (106), as well as reduced rates of nondiabetic chronic kidney disease in African Americans (75).

Using shotgun metagenomics, two recent studies suggest associations between JCPyV and lower urinary tract symptoms (LUTS). In the first study, catheterized urine samples from 20 females with overactive bladder (OAB) symptoms and 10 females without LUTS were sequenced (15). In five of the OAB+ individuals, the full genome of JCPyV was reconstructed. JCPyV was not detected in any of the samples sequenced from females without LUTS. Although the study was not directly looking for JCPyV, it suggested a potential association between JCPyV and OAB in females (15). In the second study, metagenomic sequencing of urine samples from males with (n=29) and without (n=9) LUTS was performed (46). From the metagenomic data, the authors found a statistically significant association between a diagnosis of LUTS and the presence of JCPyV. However, when using a semi-nested PCR test to detect JCPyV, no statistical significance was found between JCPyV presence and symptom states. PCR provided greater sensitivity in detecting JCPyV than shotgun metagenomics.

Given the greater sensitivity for viral detection possible via PCR, we opted to use PCR to ascertain if an association between LUTS and JCPyV existed in females. Using our ultra-specific primers designed in Chapter 2, we screened 190 catheterized urine samples from females with or without LUTS. Two LUTS cohorts were considered: OAB and UTI. The sample size far exceeds those from prior investigations (15, 46). Select samples were also subject to high-throughput sequencing. Numerous statistical analyses were performed on the resulting data to identify associations between JCPyV presence/abundance and symptom status, as well as participant demographics (age and race/ethnicity), and urobiome composition.

Methods

Participant Recruitment and Sample Collection

Urine was collected as part of prior IRB-approved studies (IRB #: 207102, 204195, 206449, 207152, 209545, 204133) by members of the Loyola Urinary Education and Research Collaborative (LUEREC) at Loyola's Health Sciences Campus (107–109). All participants were female, and the participants were given a verbal and written consent form for demographic information (i.e., race/ethnicity, age) and urine collection. All the gathered information and specimens were analyzed specifically for research purposes. Urine was collected aseptically through a transurethral catheter and then placed in a BD Vacutainer Plus C&S preservative tube for culturing. This technique bypasses the vulva, vagina, and urethra, which results in samples from the bladder specifically. Once the urine was collected, AssayAssure was added (10% / v) and the sample was then stored at -80°C.

Participant and Clinical Variables	Total Cohort (N = 190)	NoLUTS Controls (N = 58)	OAB (N = 99)	UTI (N = 33)
Age (years): mean (SD)	60 (15)	48 (13)	67 (11)	65 (16)
Race/Ethnicity				
• White	130 (68%)	34 (58%)	75 (75%)	21 (64%)
• Hispanic	23 (12%)	9 (15%)	9 (9.1%)	5 (15%)
• Black	29 (15%)	11 (19%)	12 (12%)	6 (18%)
• Asain	5 (2.6%)	3 (5.1%)	1 (1.0%)	1 (3.0%)
• Other	3 (1.5%)	1 (1.7%)	2 (2.0%)	0 (0%)

Table 3. Participant details for urine samples screened for JCPyV. Participants belong to one of three symptom cohorts: NoLUTS (controls), OAB, and UTI.

DNA Extraction

Urine samples were removed from the -80°C freezer and thawed to room temperature. DNA was extracted using the Norgen Urine DNA Isolation Kit following the manufacturer's protocol with one exception. We started with 500uL of urine (rather than 1.75mL) and adjusted

the volume of the Binding Solution used accordingly. DNA concentration was quantified using the Qubit fluorometer.

Screening Urines for JCPyV

Extracted DNA from each urine sample was processed as follows. Using the JCPyV's primers listed in **Table 4** (designed in Chapter 2), a 50uL PCR reaction consisted of 0.5uL of each primer (100mM), 5uL of extracted DNA, 19uL of nuclease-free water, and 25uL of Promega GoTaq® Master Mix. To accurately determine the presence of JCPyV, a positive and negative control were used. DNA from the sequenced individual described in Chapter 2 was used as a positive control, and the negative control did not have extracted DNA in the reaction. The negative control PCR reaction then consisted of 0.5uL of each primer (100mM), 24uL of nuclease-free water, and 25uL of Promega GoTaq® Master Mix. Thermal cycler settings are detailed in **Table 5**.

To determine if any microbial DNA was extracted, thus serving as a proxy for a positive control for the success of extraction, 16S rRNA gene sequence amplification was performed using the 63f/1387r primer pair listed in **Table 4** (110). The PCR reaction included the same volumes of components as described for the JCPyV PCR, however, 10mM of each primer was used. Thermal cycling settings are listed in **Table 6**. All primers were synthesized by Eurofins Genomics LLC (Louisville, KY). Amplicons were visualized using a 1.2% agarose gel.

Strain	Primers	Amplicon Size
JCPyV	Forward (JCPyV-2l): CAGGAAAGTCTTTAGGGTC Reverse (JCPyV-2r): CCCTGTTTAATGTGCATG	481
16S	Forward (63f): CAGGCCTAACACATGCAAGTC Reverse (1387r): GGGCGGWGTGTACAAGGC	1325

Table 4. PCR primers for JCPyV and 16S rRNA gene amplification. All primer sequences are listed from the 5' to 3' end.

Step	Temperature (°C)	Time
Initial Denaturation	95	5 min
Amplification		
Denaturation	95	30 sec
Annealing	50	30 sec
Extension	72	30 sec
Cycles: 30		
Final Extension	72	5 min

Table 5. Thermal cycling conditions for JCPyV amplification.

Step	Temperature (°C)	Time
Initial Denaturation	95	5 min
Amplification		
Denaturation	95	1 min
Annealing	55	1 min
Extension	72	1 min 30 sec
Cycles: 35		
Final Extension	72	5 min

Table 6. Thermal cycling conditions for 16S rRNA gene sequence amplification.

qPCR Screening

Quantification was performed in a 20uL reaction, containing: 10uL Invitrogen Power Sybr Green, 6uL nuclease-free water, 1uL of each primer (25mM), and 2uL extracted DNA. Primer sequences are listed in **Table 7** and the thermal cycler settings for JCPyV quantification are listed in **Table 8**. Primer sequences were designed using Primer3 (111). For the reverse qPCR primer for JCPyV, we used a modified version of the original PCR primer (JCPyV-2r in **Table 4**), in which bases were added to raise the melting temperature (112). Primers were synthesized by Eurofins Genomics LLC (Louisville, KY).

Strain	Primers	Amplicon Size
JCPyV	Forward (JCPyV-q): AGATCCCTGTAGGGGGTGTCTCC Reverse (JCPyV-2Rq): GCCCCTGTTTAATGTGCATGC	184
16S	Forward (U16SRT-F): ACTCCTACGGGAGGCAGCAGT Reverse (U16SRT-R): TATTACCGCGGCTGCTGGC	180

Table 7. Primers for qPCR for JCPyV and 16S rRNA gene quantification.

Step	Temperature (°C)	Time
Initial Denaturation	95	20 sec
Denaturation	95	3 sec
Annealing, Extension, and Read Fluorescence	60	30 sec
Cycles: 40		
Melt Curve	95	15 sec
	60	1 min
	95	15 sec

Table 8. Thermal cycling conditions for qPCR.

qPCR Analysis

Using the data generated by the qPCR machine, the relative amounts of JCPyV signal and bacterial signal were compared. The Ct values, generated by the qPCR machine, were used to obtain a ratio of JCPyV signal over 16S bacterial signal. All qPCR reactions were performed with replicates for each sample. Following this analysis, a separate investigation was performed on the Age-Race/Ethnicity Match pairs to compare the relative abundance of bacterial signals in each sample.

Participant samples were paired, with one sample JCPyV positive and one sample JCPyV negative, based on their symptom status, age, and race/ethnicity. Each pair had to have the same symptom status (OAB, UTI, or noLUTS), the same race/ethnicity (White/Caucasian, Black/African American, or Hispanic/Latina), and the same age or < 10 years difference when necessary.

Statistical Analysis

The data used in the statistical analyses came from the participant-provided demographic information (i.e., race/ethnic background, age), the presence or absence of JCPyV (PCR results), and the quantities of JCPyV and bacteria present (qPCR results). All statistical analyses were

performed using RStudio (v3.6.2) (113). Packages used included ‘MASS’ and ‘ggplot2’ (114, 115). Univariate models were performed on individual features using ‘glm’ to fit generalized linear models. Two-way interaction models were performed also using ‘glm’. The full model of all the features was also performed and that model was then reduced using the ‘MASS’ library and the ‘stepAIC’ method (116). Visualization was done with ‘ggplot2’ and ‘Likert’ (117).

From the Age-Race/Ethnicity matched pairs, the Ct values generated from the 16S rRNA gene sequence qPCR quantification for each pair were used. The difference in bacterial abundance between the pairs (JCPyV+ and JCPyV-) was calculated and a paired t-test was performed, using R, to calculate any statistical significance between those differences in bacterial abundance. The bacterial abundance of all the Age-Race/Ethnicity matched pairs relative to their symptom status (OAB+, UTI+, noLUTS) was also analyzed. A subsequent paired t-test was used on this data to derive the significance of the difference in bacteria for each symptom category. Shapiro-Wilks normality tests were used to test the normality of the results from the t-tests. If the normality assumption fails, instead of using the t-test, we used its nonparametric version – a Wilcoxon Signed Rank test.

Metagenome Sequencing

Thirty-three DNA samples (16 OAB+, 2 noLUTS/OAB-, and 15 UTI+) were sent to be sequenced at MIGS (Pittsburgh, PA) (**Table 9**). Libraries were prepared by MIGS using the Illumina DNA Prep kit and IDT 10bp UDI indices, and sequenced on an Illumina NextSeq 2000, producing 2x151bp reads. Demultiplexing, quality control, and adapter trimming were performed with bcl-convert (v3.9.3). Sequencing produced 240M reads with read lengths of 150

nucleotides. Raw sequences were deposited into NCBI's SRA database (Study ID: PRJNA680735) (**Table 9**).

Sample ID	Symptom Status	SRA Accession Number
2720	OAB+	SRR19149285
2981	OAB+	SRR19149284
3000	OAB+	SRR19149273
3001	OAB+	SRR19149262
3002	OAB+	SRR19149258
3022	OAB+	SRR19149257
4462	OAB+	SRR19149256
4511	OAB+	SRR19149255
4578	OAB+	SRR19149254
4821	noLUTS	SRR19149253
5461	noLUTS	SRR19149283
5520	OAB+	SRR19149282
6162	OAB+	SRR19149281
6403	OAB+	SRR19149280
6517	OAB+	SRR19149279
6578	OAB+	SRR19149278
6901	OAB+	SRR19149277
7264	OAB+	SRR19149276
7531	UTI+	SRR19149275
7651	UTI+	SRR19149274
7654	UTI+	SRR19149272
7672	UTI+	SRR19149271
7676	UTI+	SRR19149270
7707	UTI+	SRR19149269
7714	UTI+	SRR19149268
7720	UTI+	SRR19149267
7771	UTI+	SRR19149266
7772	UTI+	SRR19149265
7775	UTI+	SRR19149264
7785	UTI+	SRR19149263
7786	UTI+	SRR19149261
7791	UTI+	SRR19149260
7803	UTI+	SRR19149259

Table 9. List of Sample IDs and their SRA Accession Numbers.

Sequenced Metagenomes Mapping

The resulting sequences were uploaded and paired in Geneious v2022.0.2 (Biomatters, Ltd., Auckland, New Zealand). The sequences were then mapped to the JCPyV (accession no. NC_001699.1) and BKPyV (accession no. NC_001538.1) reference genomes, using the Bowtie2 plug-in (85) as described in Chapter 2 Methods.

MG-RAST Analysis

The raw sequence read fastq files were uploaded to MG-RAST as paired reads. To best fit the data, the pipeline was configured with the following parameters: replication, screening for *Homo sapiens*, and dynamic trimming. Replication removed artificial replicated sequences produced by sequencing artifacts (118). Screening for *H. sapiens* (NCBI v36) removed any human-specific sequences from the analysis by using DNA level matching with Bowtie (119). Dynamic trimming removed low-quality sequences using a modified DynamicTrim, this step was configured with the following specifications: the lowest phred score counted as a high-quality base was 15, and each sequence would be trimmed to contain at most 5 bases below 15 (120).

Urobiome Metagenome Analysis

Using the data generated from the MG-RAST analysis, an analysis was performed to determine whether bacterial taxa in the urobiome differs with symptom status. To do this, a 2-way ANOVA model was conducted, along with subsequent Shapiro-Wilks and Levene tests. The taxonomic order used in this analysis went down to genus (Phylum, Class, Family, Order and Genus).

Results

Assessing JCPyV Prevalence in Urine

DNA was extracted from 190 urine samples collected from participants from three symptom groups: OAB, UTI, and lack of lower urinary tract symptoms (noLUTS). **Table 10** lists the symptom status for the samples tested. In total, 52 of the samples tested positive for JCPyV. This was determined through PCR with the JCPyV-specific designed primers (**Table 4**). JCPyV samples were found in all three symptom cohorts, and no association between symptom status and JCPyV presence was found (OAB+ p-value: 0.771810; UTI+ p-value: 0.405519).

	# Samples	# positive	% of positives
noLUTS	58	14	24.137931
OAB+	99	27	27.2727273
UTI+	33	11	33.3333333
TOTALS	190	52	27.3684211

Table 10. Results from screening 190 frozen urine samples for JCPyV.

To verify the presence of JCPyV in the sample, 33 of the 52 positive samples were sent for shotgun metagenomic sequencing (**Table 9**). These 33 samples were selected as they had sufficient DNA (>1 ng/ul) for sequencing. For each metagenome, I mined for JCPyV using the methods described in Chapter 2.

Of the 33 metagenomes, we were able to reconstruct the full JCPyV genome in 18 of the samples, signifying that JCPyV was likely present in abundance in the urine samples (**Table 11**). Five additional samples had “partial coverage,” indicating that some reads mapped to the JCPyV reference genome but did not cover enough of the genome to reconstruct the full JCPyV genome (**Table 12**). The remaining ten metagenomes from JCPyV PCR positive samples did not actually produce any contigs that mapped to the JCPyV reference genome (**Table 13**).

Sample ID	Coverage of JCPyV Genome	Depth of JCPyV Coverage (x)	Size of Metagenome (bp)
2720	93.3%	3.5	1,035,594,728
2981	97.1%	368.3	777,919,682
3000	96.7%	57.7	814,807,775
3001	96.8%	133.8	914,256,943
3002	96.8%	26.7	989,357,294
3022	97.9%	213.1	1,065,264,347
4462	88.7%	3.1	1,312,737,432
4821	96.6%	12.4	973,263,519
5461	96.6%	13.0	1,530,615,878
6162	98.3%	475.1	1,372,417,896
6403	96.6%	54.1	1,397,423,635
6517	90.9%	3.6	862,608,832
6578	97.0%	529.7	606,619,855
7531	96.6%	20.9	1,425,736,841
7672	93.0%	6.1	1,240,442,055
7676	75.0%	2.8	1,271,227,832
7772	78.7%	2.8	1,623,832,233
7786	96.6%	14.2	1,856,253,332

Table 11. Metagenomes producing a full coverage JCPyV genome. The coverage of JCPyV column lists the percentage of the RefSeq genome mapped to by metagenome reads. Sequence Depth is the mean number read coverage of the RefSeq genome. The final column is the total number of sequenced nucleotides for each sample.

Sample ID	Coverage of JCPyV Genome	Depth of JCPyV Coverage (x)	Size of Metagenome (bp)
4511	15.1%	5.8	1,047,203,156
4578	35.0%	0.7	1,274,697,938
7264	25.3%	0.5	1,253,519,499
7714	5.9%	0.1	1,239,087,612
7785	39.8%	0.7	1,786,261,288

Table 12. Metagenomes producing a partial coverage JCPyV genome. The coverage of JCPyV column lists the percentage of the RefSeq genome mapped to by metagenome reads. Sequence Depth is the mean read coverage of the RefSeq genome. The final column is the total number of sequenced nucleotides for each sample.

Sample ID	Size of Metagenome (bp)
5520	1,116,748,348
6901	977,113,518
7651	706,138,317
7654	579,479,747

7707	644,718,918
7720	493,689,843
7771	628,231,521
7775	792,313,184
7791	671,789,553
7803	783,217,677

Table 13. Metagenomes producing no JCPyV genome reads. These samples did not map any contigs to the RefSeq JCPyV genome. The second column lists the total number of sequenced nucleotides for each sample.

All 33 samples were also mapped to the BKPyV reference genome. Only one of the samples (Sample ID 3000) mapped with full coverage to the BKPyV reference genome as well as to the JCPyV reference genome. Both coverage graphs are represented in **Figure 5**. None of the other 32 metagenomes had reads that mapped to the BKPyV reference genome.

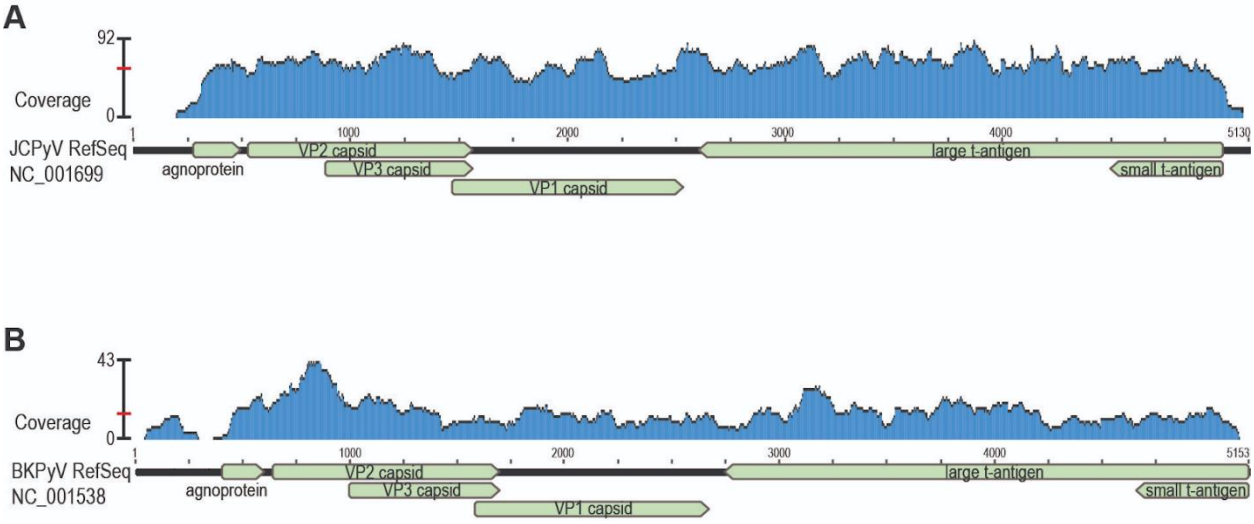

Figure 5. Raw read coverage of Sample 3000 to the (A) JCPyV RefSeq genome and (B) BKPyV RefSeq genome. The average coverage (indicated by the red bar) was ~57.7 reads/nucleotide for JCPyV and ~13.8 reads/nucleotide for BKPyV.

Investigating Association between JCPyV Presence and Demographics

Next, JCPyV PCR screening results were examined to evaluate possible associations with available demographic data: age and race/ethnicity. Three samples were removed from our demographic analysis as they either had an ambiguous race/ethnicity listed (“Other”) or more

than one race/ethnicity. Additionally, we removed data from females who self-identified as "Asian;" we only had five samples and thus not enough data for statistical analyses. Thus, our analysis focused on 182 samples from three race/ethnic groups: White/Caucasian, Black/African American, and Hispanic/Latina. JCPyV was detected via PCR in the urine samples of females from all age groups and all races/ethnicities assayed (**Table 14**).

	White/Caucasian	Black/African American	Hispanic/Latina	TOTAL
<30	0	3	2	5
30-39	7	3	4	14
40-49	12	4	6	22
50-59	28	8	4	40
60-69	29	5	5	39
70-79	36	5	2	43
80+	18	1	0	19
TOTAL	130	29	23	182

Table 14. Breakdown of the 182 samples by age and race/ethnicity. The "TOTAL" row and column is a sum of the participants from each cohort.

The first set of statistical analyses conducted were univariate models on age, race/ethnicity, and symptom status. We modeled the presence of JCPyV in a logistic regression model with age, race/ethnicity, and participant symptom status as covariates. Since all these covariates were categorical, a baseline category had to be established. For age, individuals who belonged to the < 40 age group were used as the baseline comparison category. For race/ethnicity, individuals who belonged to the "Black/African American" group were selected as the baseline comparison category. Finally, for symptom status, individuals who belonged to the noLUTS group were used as the baseline comparison category. None of the age groups, race/ethnicity, or symptom groups were found to be statistically significant (all p-values > 0.05; **Table 15**). However, the White/Caucasian p-value is very close to the boundary, which is an

indication that JCPyV incidence may be significantly different for White/Caucasian females when compared to Black/African American females.

Univariate Models	p-value
Age Group	
(40,50]	0.09091
(50,60]	0.37855
(60,70]	0.93284
(70,100]	0.17560
Race/Ethnicity Group	
Hispanic/Latina	0.8696
White/Caucasian	0.0585 *
Symptom Group	
OAB+	0.771810
UTI+	0.405519

Table 15. The p-value for each separate univariate model and their respective cohorts. These numbers demonstrate that there is no significant association between age, race/ethnicity, or symptom status and JCPyV presence. The asterisk (*) is used to denote a p-value that is close to the 0.05 boundary for a variable.

The expected log odds of JCPyV presence were then calculated for the age groups (**Figure 6**), race/ethnicity groups (**Figure 7**), and symptom groups (**Figure 8**). If the estimated log odds of JCPyV (the white circle) is above 0 (i.e., the expected odds are above 1), then JCPyV presence is more likely to occur in that age (**Figure 6**), race/ethnicity (**Figure 7**), or symptom category (**Figure 8**) than in the baseline category (females under 40, Black/African American females, or noLUTS symptoms), and vice versa. The age graph (**Figure 6**) shows all four of the blue lines crossing the red line, indicating that the age groups are not significant for JCPyV presence. The race/ethnicity graph (**Figure 7**) shows both blue lines crossing the red line, indicating that the race/ethnicity groups are not significant for JCPyV presence. Finally, the symptom graph (**Figure 8**) shows both blue lines crossing the red line, indicating that the symptom groups also are not significant for JCPyV presence.

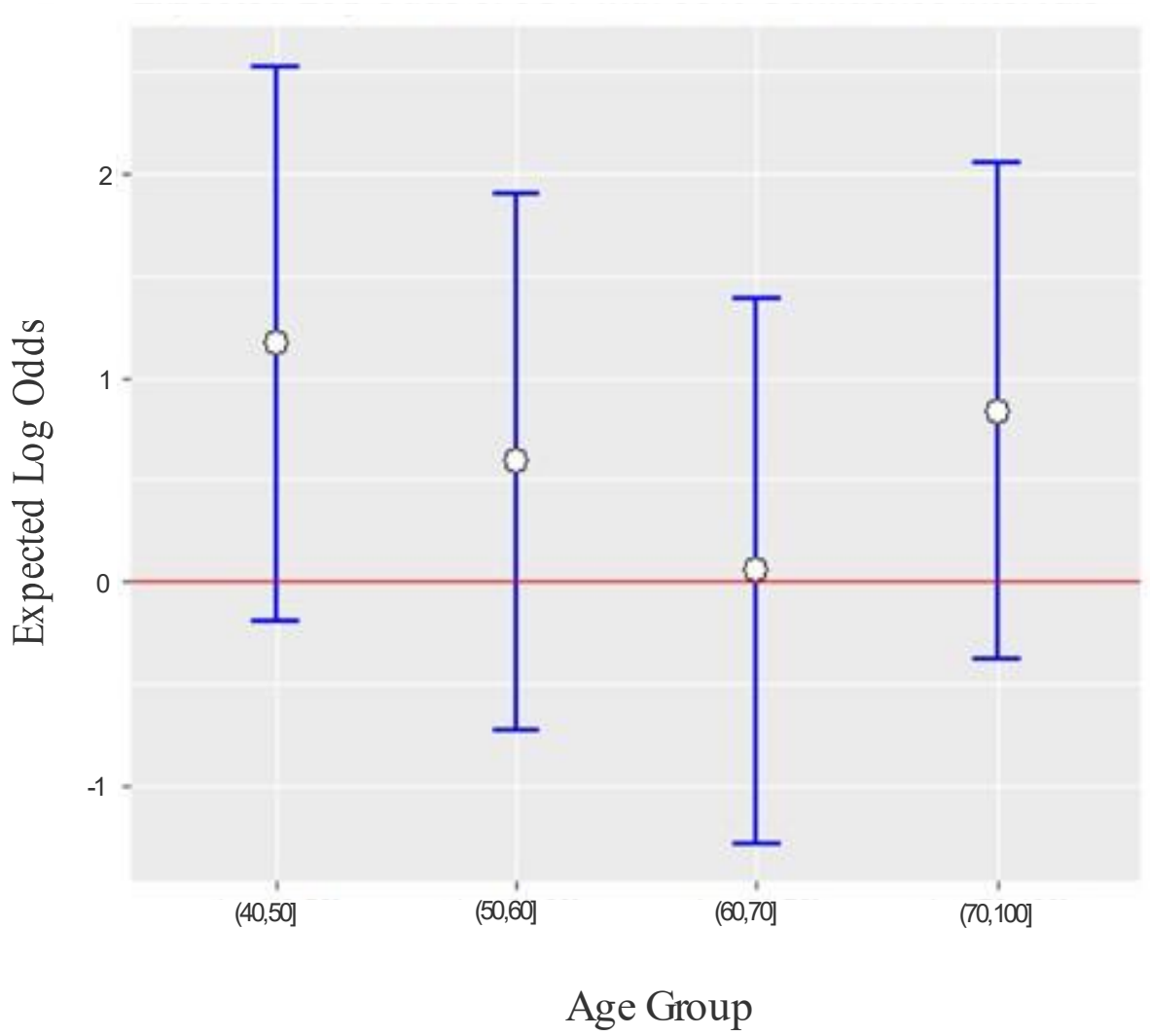

Figure 6. The expected log odds ratio of JCPyV presence for age groups with respect to the baseline category - the < 40 age group. The white circle represents the estimated log odds of JCPyV of a participant in the age category with respect to a participant who is 40 or younger in age. The blue lines are the 95% confidence interval for each estimate. The red line indicates a log odds of 0, in other words an expected odds of 1, which indicates no significant difference with respect to the baseline category. If the blue lines do not cross the red line, then they are statistically significant.

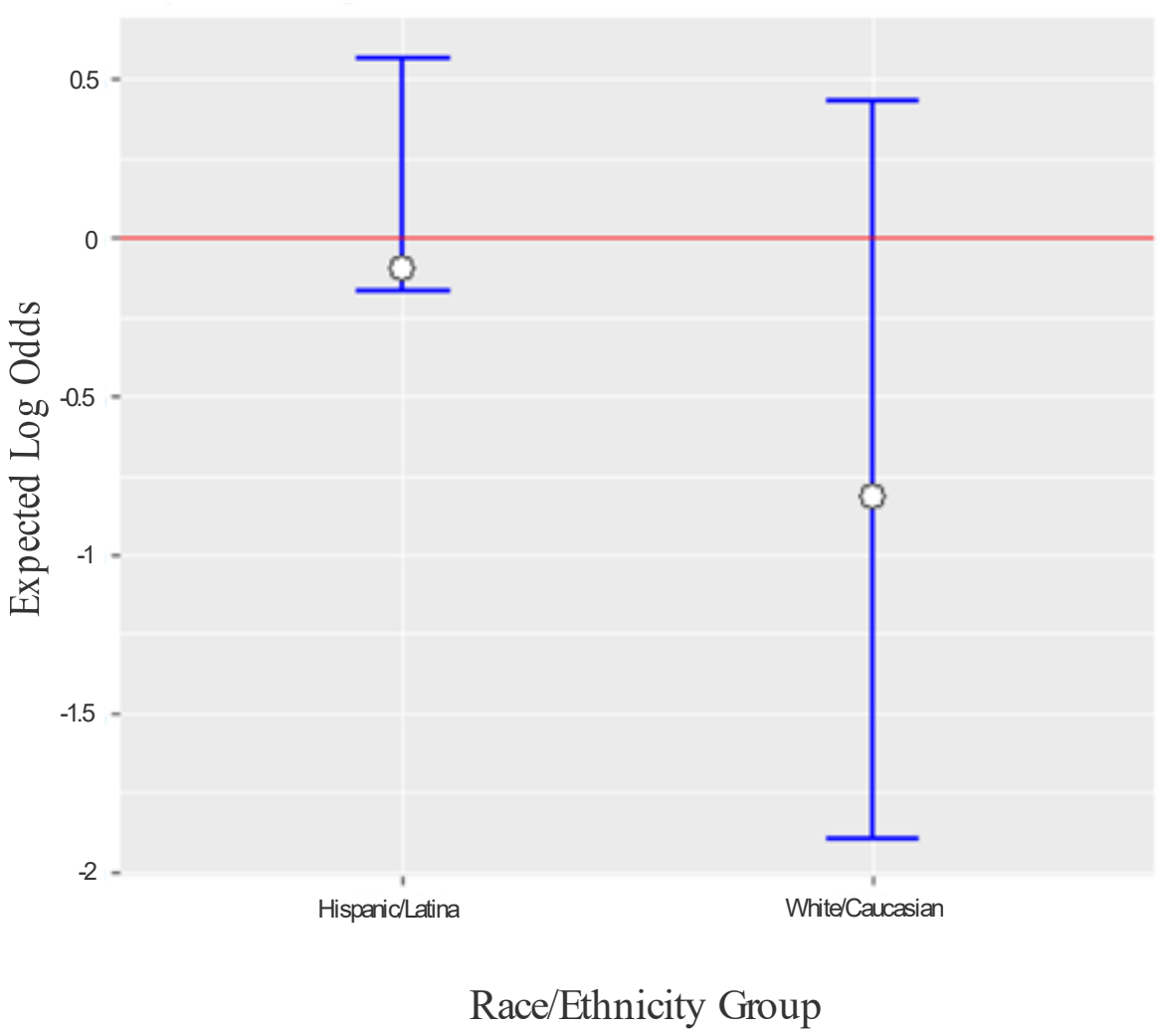

Figure 7. The expected log odds ratio of JCPyV presence for race/ethnicity groups with respect to the baseline category - the “Black/African American” group. The white circle represents the estimated log odds of JCPyV of a participant in the race/ethnicity category with respect to a participant who is Black/African American. The blue lines are the 95% confidence interval for each estimate. The red line indicates a log odds of 0, in other words an expected odds of 1, which indicates no significant difference with respect to the baseline category. If the blue lines do not cross the red line, then they are statistically significant

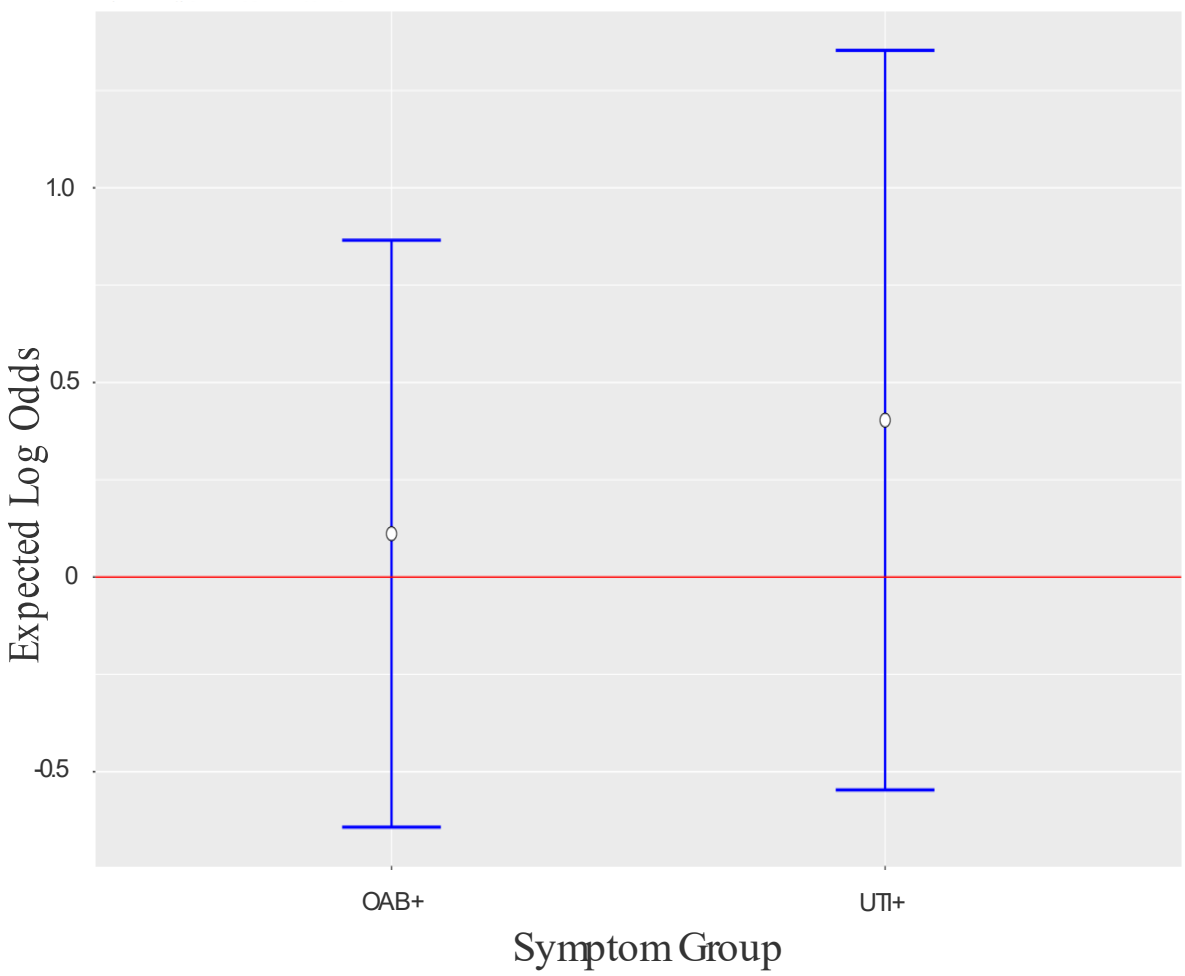

Figure 8. The expected log odds ratio of JCPyV presence for symptom groups with respect to the baseline category - the “noLUTS” group. The white circle represents the estimated log odds of JCPyV of a participant in the symptom category with respect to participants in the noLUTS group. The blue lines are the 95% confidence interval for each estimate. The red line indicates a log odds of 0, in other words an expected odds of 1, which indicates no significant difference with respect to the baseline category. If the blue lines do not cross the red line, then they are statistically significant

Next, we considered two-way interaction logistic regression models to explore if these covariates have a joint effect on JCPyV prevalence. The age and symptom interaction model showed no significance (all p-values > 0.05). The age and race/ethnicity model showed that females aged 70 years and older (from all races/ethnicities) had a significantly higher prevalence of JCPyV compared to females younger than 40 (p-value = 0.0409). The estimated coefficient

was 3.1781, meaning females 70 years and older were $e^{3.1781} = 24.011 \approx 24$ times more likely to have JCPyV than females under 40. Otherwise, none of the other categories for age and race/ethnicity were statistically significant. Finally, the race/ethnicity and symptom interaction model showed no statistical significance. However, some of the p-values were very close to the 0.05 threshold and are worth future inquiry (**Appendix D**).

The full model looked at the interactions between the three categories. However, this does not mean we are looking at the three-way interaction between age, race/ethnicity, and symptoms as there is not enough data to look at three-way interaction terms. Instead, this model includes all two-way interactions and individual effects for age, race/ethnicity and symptoms. The model was then reduced to identify significant features. In search of a parsimonious model with high accuracy we used Akaike Information Criterion (*AIC*). The general ruling for this assessment is the lower the *AIC*, the better the model. One can obtain the best model (the one with the lowest *AIC*) by reducing the model only with categories where at least one level is significant. The full model had an $AIC = 245.84$, which was reduced using the StepAIC function in R. The final reduced model had an $AIC = 217.9$. None of the interaction terms and symptom categories were significant and dropped from the model. Both age and race/ethnicity had significant categories and were therefore retained in the reduced model. Once the model was reduced there were several instances of suggested significance in predicting the prevalence of JCPyV. **Table 16** shows the output including the estimated odds of JCPyV prevalence for our final reduced model.

These results were then visualized to show their significance as seen in **Figure 9**. Females in the (40,50] age group were 4.4 times and females in the (70,100] age group were 4.3

times more likely to have JCPyV than females under 40 (p-values found in **Table 16**). White/Caucasian females had a p-value of 0.0154, and with an estimated coefficient value of -1.13; thus, White/Caucasian females are about 3 ($e^{1.13}$) times less likely to have JCPyV than Black/African American females (**Table 16**). In other words, our investigation finds that White/Caucasian females were about 70% less likely to have JCPyV than Black/African American females.

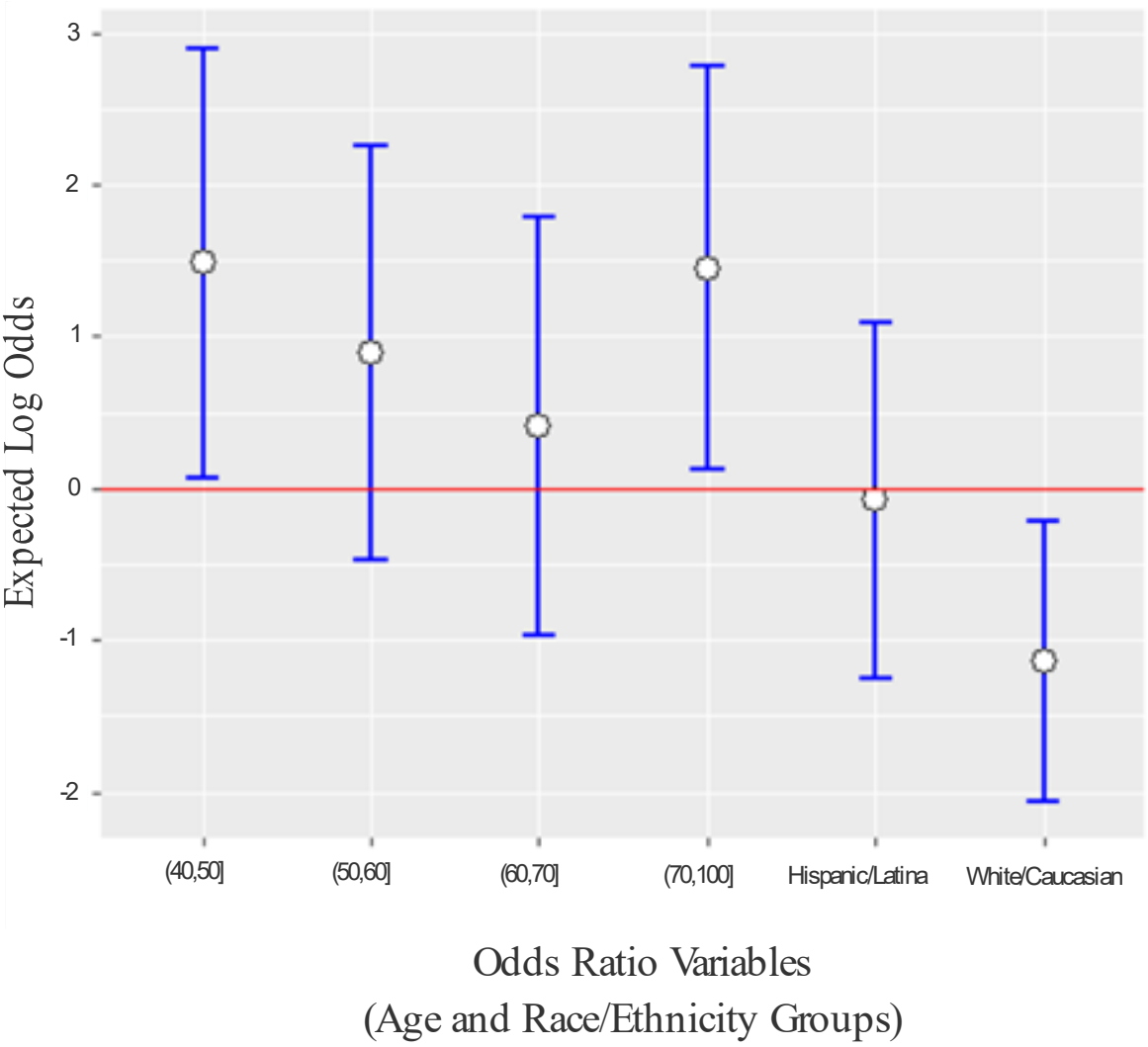

Figure 9. Visual representation of the expected log odds ratios for the final reduced model. The "Black/African American" race/ethnicity group and the < 40 age group were the baseline categories. The white circle represents the estimated log odds ratio. The blue lines are the 95%

confidence interval for each estimate. The red line indicates an expected log odds equal to 0 (or expected odds of 1), and if the blue lines do not cross the red line, then they are statistically significant.

	Estimate Coefficient	Expected Odds	Standard Error	p-value
(40,50]	1.48497	4.41483296	0.72218	0.0398 **
(50,60]	0.89235	2.44085893	0.69457	0.1989
(60,70]	0.40842	1.50443889	0.70382	0.5617
(70,100]	1.45190	4.27122213	0.67816	0.0323 **
Hispanic/Latina	-0.07978	0.92331945	0.60043	0.8943
White/Caucasian	-1.13630	0.32100454	0.46922	0.0154 **

Table 16. Results from the final reduced logistic regression model. The first column is the coefficient estimate and its sign is used to understand if the response is more or less likely for a category compared to the baseline. The second column is the expected odds of a response (likelihood that JCPyV is to be present). The third column is the calculated standard error and the last column is the calculated p-value. The double asterisk (**) is used to denote a significant p-value for a variable.

Is JCPyV Abundance Associated with Symptom Status?

Next, qPCR was performed to investigate associations between symptom groups and the relative abundance of JCPyV and bacteria in each sample. Urobiome samples that were PCR-positive for JCPyV were assayed using both the JCPyV and 16S primers listed in **Table 7**. Ct values were used as a proxy for DNA abundance and recorded for both the JCPyV and 16S rRNA reactions (**Appendix E**). Using these two values, the relative abundance of JCPyV relative to the number of 16S rRNA gene sequences was computed (**Figure 10**).

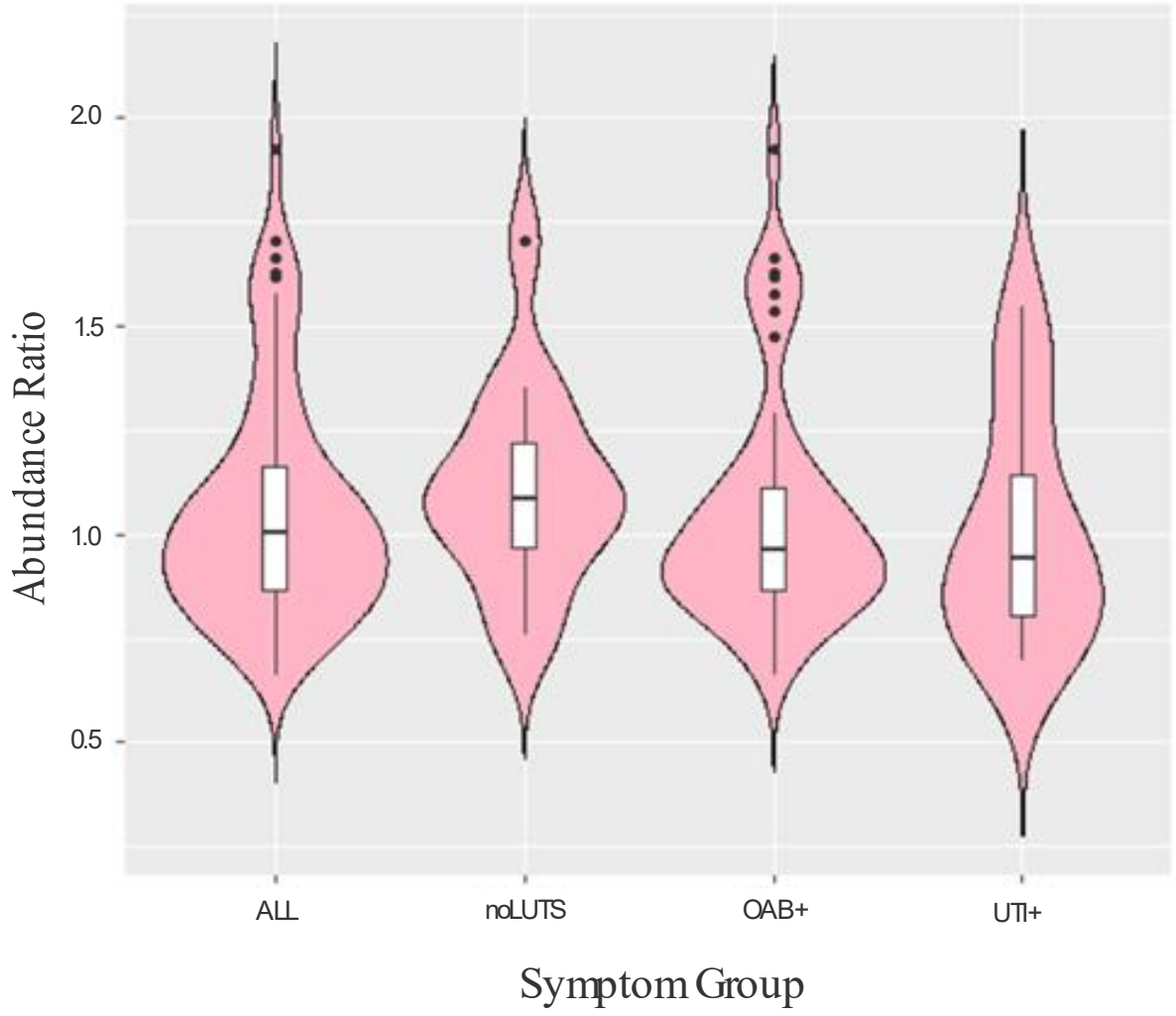

Figure 10. Violin plots of the relative abundance of JCPyV to bacteria Ct values within 4 groups. ALL represents all the data, regardless of symptom status; noLUTS includes only JCPyV+ individuals from the control group; OAB+ includes females that are JCPyV PCR-positive and OAB+; UTI+ includes females who are JCPyV PCR-positive and UTI+.

As seen in **Figure 10**, the violin plots show the relationship of symptom status to the relative abundance ratio. The box plot elements show that the median values of the abundance ratios for symptom status are all relatively similar to each other. The OAB+ group had a mean value of 1.053661. The UTI+ group had a mean value of 1.010301. The noLUTS group had a mean value of 1.113096. Finally, the "ALL" group, which was the ratio calculated from all samples regardless of symptom status, had a mean value of 1.058734. The closeness of these

mean values indicates that each sample contains relatively the same number of JCPyV virions relative to their bacterial community size, regardless of symptom status. The shape of the distribution (extremely skinny on each end and wide in the middle) in each of the plots indicates the ratio values for each symptom status are highly concentrated near the median value. Overall, this analysis showed that there was no indication of significant associations between the symptom status and relative amounts of JCPyV to bacteria in a sample. However, it is worth it to consider the outliers present in the data, as denoted by the black dots in **Figure 10**. While the ALL, noLUTS, and OAB+ plots have similar shapes – they are all stretched out due to the presence of outliers. These outliers indicate that, for some of the samples, there was almost double the amount of JCPyV in comparison to the relative amount of bacteria. The UTI+ plot is not as similar in shape as it does not have any outliers.

Is JCPyV Presence/Abundance Associated with Urobiome Biomass?

To ascertain if JCPyV presence was due to a higher urobiome biomass, qPCR was performed using the 16S rRNA gene primers for JCPyV PCR-negative samples that had matched symptom status, age, and race/ethnicity to JCPyV PCR-positive samples (**Appendix F**). The difference was computed between the total bacterial abundance from each matched pair. No significant difference (p-value = 0.4497) was found in the total bacteria between the urobiomes of individuals with or without JCPyV (**Figure 11**). From **Figure 11** we also see that the 95% confidence interval for the difference in total bacteria between the JCPyV+ and JCPyV- participants includes zero, which confirms that there is no significant difference between these populations. **Figure 12** is a 0-centered stacked barplot of the bacterial abundance for each paired sample in the Age-Race/Ethnicity Match cohort. The color of the bar represents the JCPyV+

individual (dark blue) and the JCPyV- individual (light blue) within a single pair, and the length of the bar represents the relative amount of bacteria found in each sample. The majority of the dark blue and paired light blue bars have the same length, or amount of bacteria, thus, indicating no significant difference between JCPyV presence or absence and bacterial abundance.

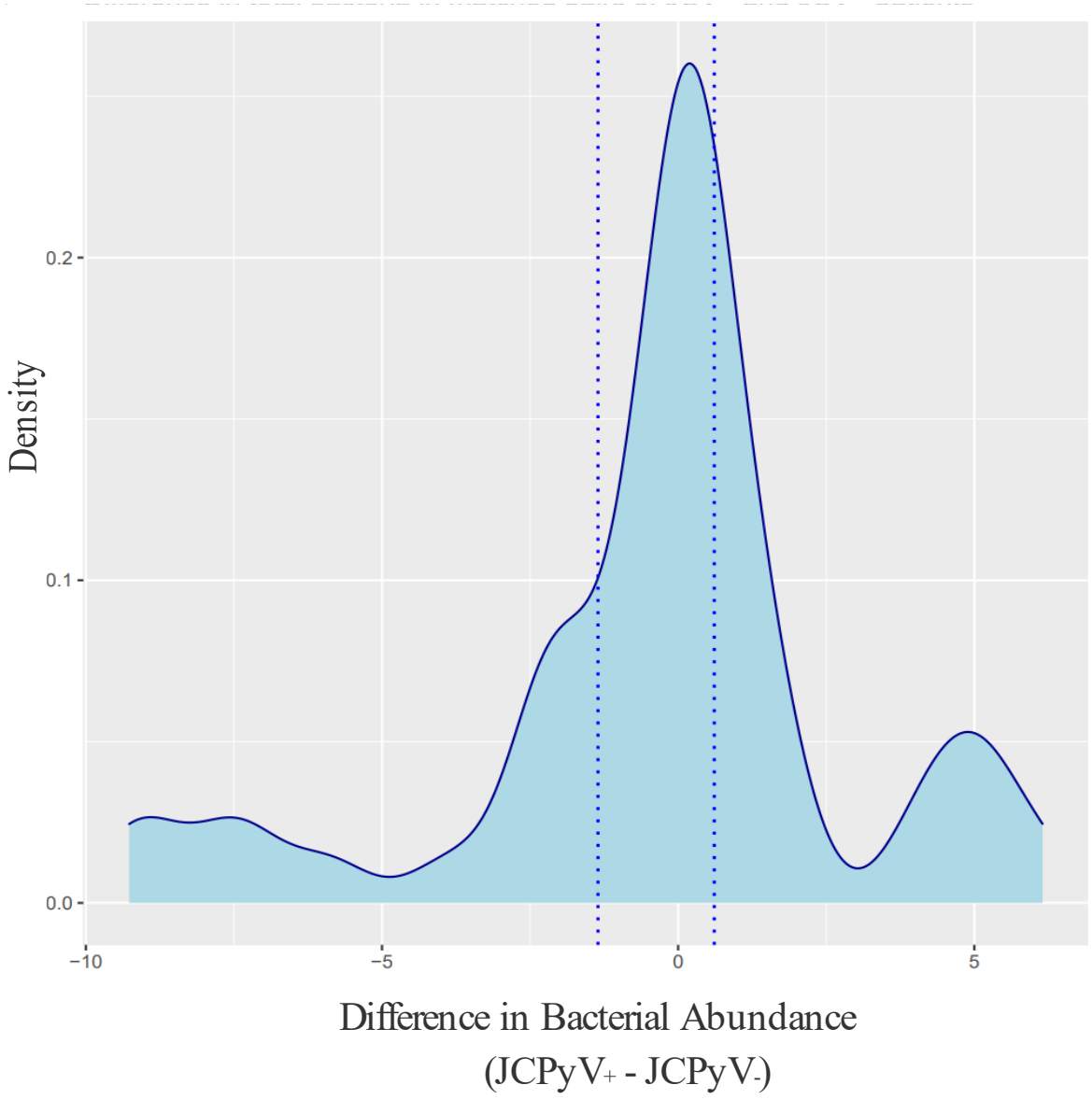

Figure 11. Density plot of difference counts of total bacteria between JCPyV+ and JCPyV- participants from the Age-Race/Ethnicity matched paired data. The dotted lines signify the 95% confidence interval for the difference in total bacteria between the matched pairs.

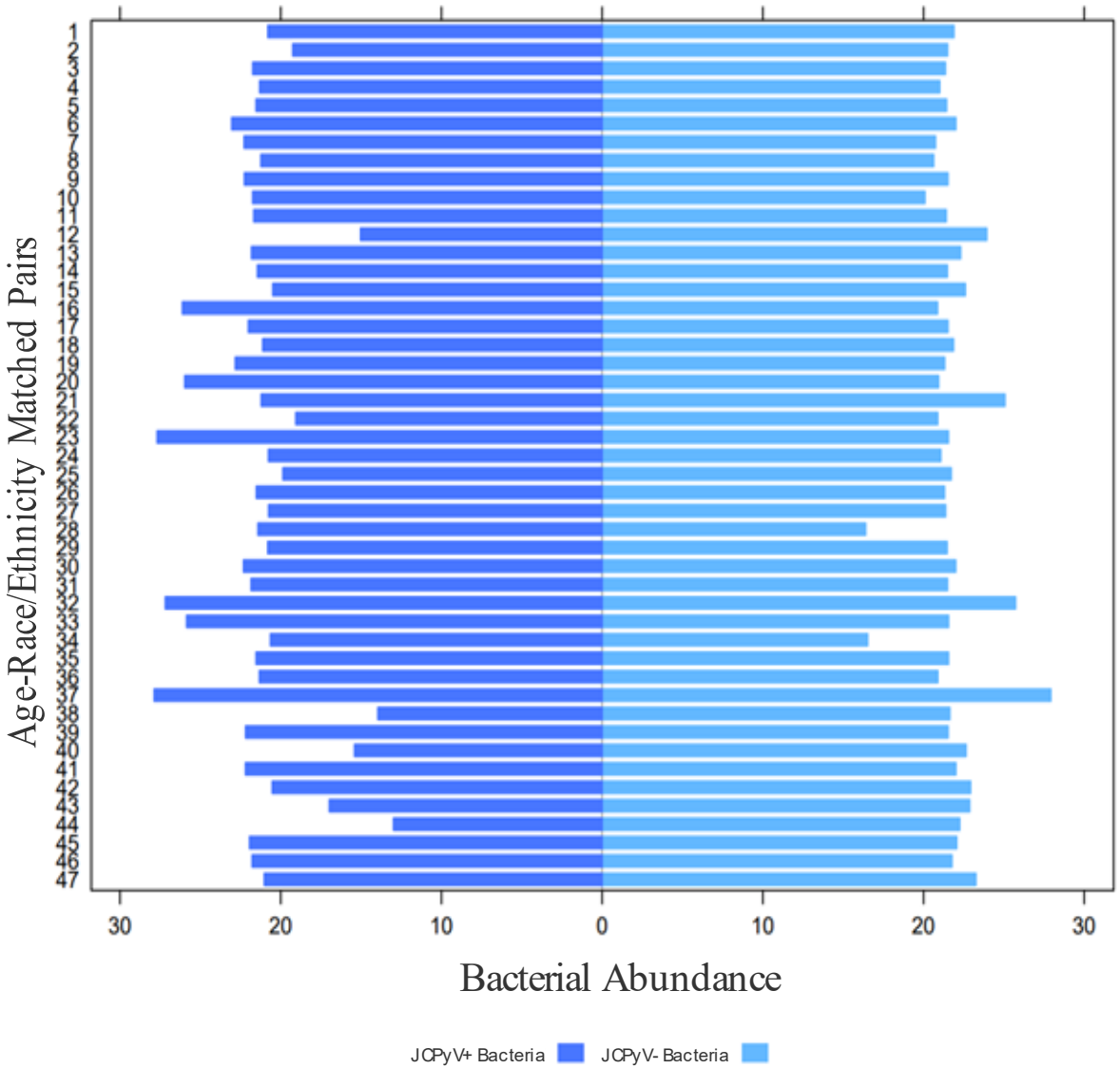

Figure 12. Comparison of bacterial abundance between JCPyV PCR-positive samples (dark blue) and their Age-Race/Ethnicity matched JCPyV PCR-negative sample. The dark blue indicates the relative amount of bacteria (Ct value) for the JCPyV+ sample and the light blue indicates the relative amount of bacteria (Ct-value) for the JCPyV- sample.

We also looked at the bacterial abundance in these matched pairs specific to symptom status. In all, there were 14 pairs of noLUTS females, 25 pairs of OAB+ females, and 8 pairs of UTI+ females. JCPyV+ and JCPyV- urobiome bacterial abundance showed no significant differences for both the noLUTS (p-value=0.6257) and OAB+ groups (p-value=0.327). Since the

noLUTS group had only 14 JCPyV PCR-positive samples and the normality assumption was not met (Shapiro-Wilks normality test p-value=0.0001), we used the Wilcoxon Signed Rank test here. In contrast, the 8 pairs of UTI+ females had a p-value of 0.0346 (**Table 17**, **Figure 13**). For the OAB+ (Shapiro Wilks normality test p-value$= 0.0826$) and UTI+ (Shapiro Wilks normality test p-value$= 0.1784$) group, the normality assumption was met and hence the paired t-test was used.

	Mean	Standard Deviation	95% CI
JCPyV+	19.1091	3.5076	(16.1767, 22.0416)
JCPyV-	22.4797	0.5254	(22.0404, 22.9189)

Table 17. Total bacteria in JCPyV+ and JCPyV- participants in the UTI+ group.

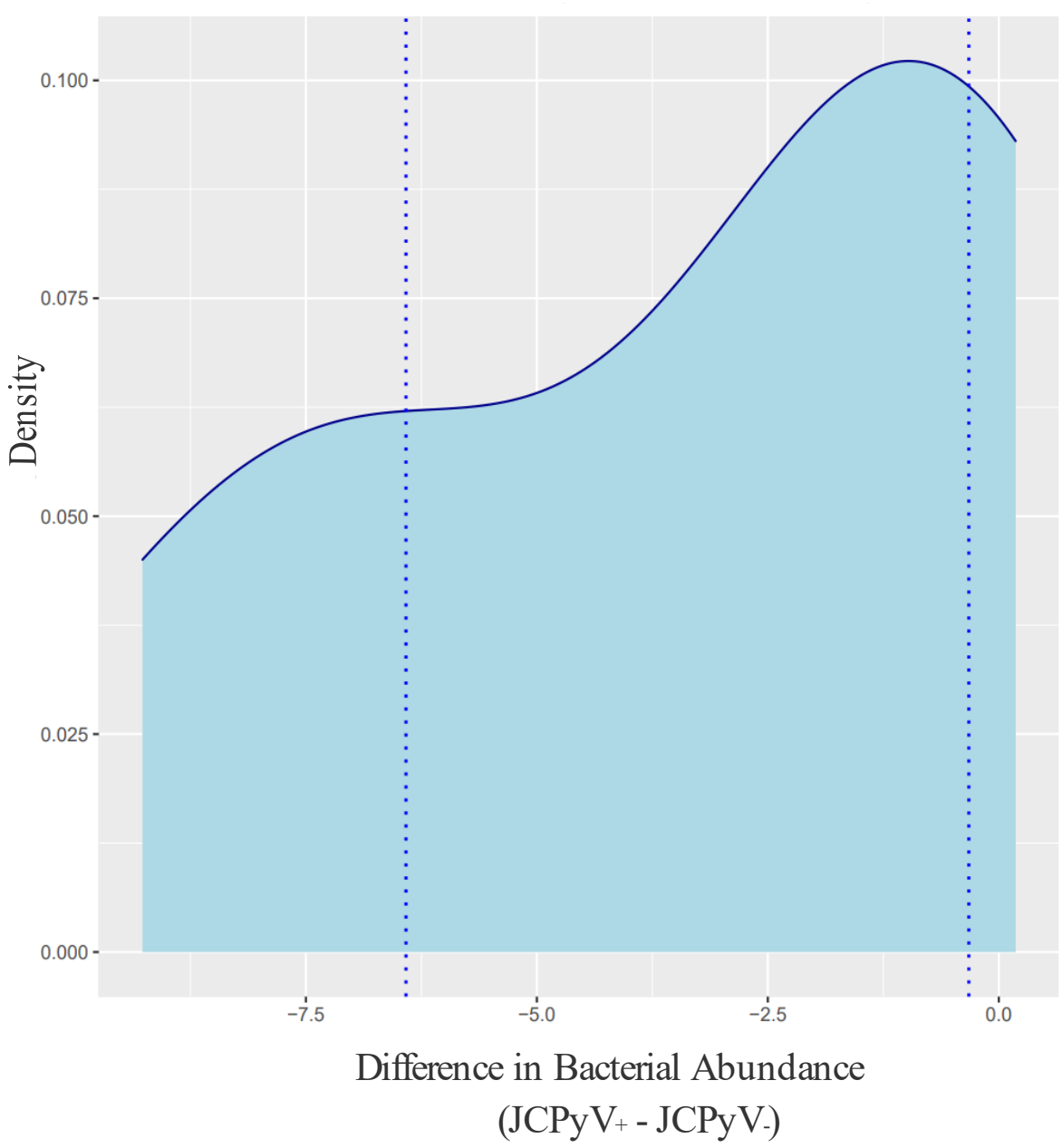

Figure 13. Density plot of difference counts of total bacteria between JCPyV+ and JCPyV- participants from the Age-Race/Ethnicity Match paired data for UTI+ participants only.

For the 8 UTI+ paired data sets, there was a significant difference in total bacteria between JCPyV+ and JCPyV- participants. As **Figure 13** shows, the 95% CI for the difference in total bacteria count is (-6.4168, 0.3243) indicating that there were fewer bacteria in samples that

were JCPyV+ participants than in samples that were JCPyV-. However, the JCPyV PCR-positive samples from the UTI+ cohort had more variation than their matched controls (**Table 17**).

Is JCPyV Presence Associated with Urobiome Composition?

To explore if JCPyV prevalence is associated with urobiome composition, we checked if there was any association between bacterial taxa and participants' symptoms for the JCPyV+ participants only. In this test, the complete urobiome of 25 JCPyV+ participants were sequenced; 16 were UTI+, 7 were OAB+, and there were 2 noLUTS controls. For each level of taxa (Phylum, Class, Family, Order and Genus) we checked for significant association between taxa and participant symptoms (UTI+, OAB+, noLUTS) using a two-way ANOVA model. For each model we checked assumptions of normality and homoscedasticity using the Shapiro-Wilks and the Levene' test, respectively.

We did not find any significant interaction with participant symptoms for any of the taxa level (Phylum, Class, Family, Order and Genus) among the JCPyV+ urobiomes. However, we found that by ignoring symptom status, Genus, Family and Order were found significant for the 25 JCPyV+ participants. This implies that these 25 participants showed significant difference in the distribution of bacterial taxa at the Genus (p-value = 0), Family (p-value = 0.0351) and Order (p-value = 0.0068) levels.

Discussion

A previous study found repeated instances of JCPyV in females who were OAB+, and only in those who were OAB+ (15). This led to the original hypothesis of this study as the presence of JCPyV may be correlated, or be associated with, OAB symptoms in females. This prior study, however, had a small sample size: just 30 samples (20 from OAB+ and 10 from

OAB-/noLUTS). Our investigation gathered a larger sample size of 190 samples. Looking at this larger cohort, we found no statistical significance between symptom status and JCPyV presence. While the prior study did not find JCPyV in the 10 noLUTS samples sequenced, our PCR-based assay identified 14 out of the 58 samples from females without LUTS to be positive for JCPyV (**Table 10**). Furthermore, when two of these JCPyV+ noLUTS samples (IDs 4821 and 5461) were sequenced, the complete genome of JCPyV could be assembled (**Table 11**), indicative of its presence in abundance. Thus, with the large cohort considered here, we can state that JCPyV can be present in individuals without LUTS. With ~27% of the urine samples screened here found to be JCPyV positive, our results suggest that JCPyV prevalence in the female population is ~27%, the low end of current estimates in the literature (24).

While the presence of JCPyV does not seem to be associated with LUTS in females (explicitly in females with OAB or UTIs), our analysis did show that the primers are specific to JCPyV. Shotgun metagenomic sequencing confirmed JCPyV presence in most (~54%) PCR-positive samples. After mapping the 33 metagenomes to the BKPyV reference genome, only one indicated the presence of BKPyV (in addition to JCPyV), Sample ID 3000 (**Figure 5**). This result directly speaks to the specificity of the primers designed in Chapter 2. Overall, we find that shotgun metagenomic sequencing provides a means of effectively discriminating between closely related viral species, such as BKPyV and JCPyV, if they are abundant within the sample.

While we were unable to detect JCPyV sequences from 10 of metagenomic samples that were JCPyV PCR-positive, we are confident that JCPyV was present but in low abundance. PCR can detect low levels of DNA (46). PCR products can detect and amplify very low quantities of a viral species, which is why there may be a positive PCR result but no coverage when mining the

reads of the sample. Thus, for targeted detection of a particular virus of interest, PCR is far more sensitive than shotgun metagenomics. In the study of Thomas et al., which looked at JCPyV in males, the authors similarly found samples that were JCPyV- via metagenomic sequencing but JCPyV+ via PCR (46).

While our study's results refute our initial hypothesis of an association between JCPyV and OAB, we explored additional factors that may be associated with JCPyV presence, including age, race/ethnicity, and/or symptom status. A previous study looked at the incidence of JCPyV viruria and seropositive rates in the general population in Taiwan across various age groups, ranging from 0 to 94 years old (21). From 1012 samples, the authors found via PCR assays that the incidence of JCPyV urinary shedding was 27.6% for those 31 to 40 years old, 32.7% for those 41 to 50 years old, 41.9% for those 51 to 60 years old, 47.3% for those 61 to 70 years old, and 65.5% for those older than 70 years, leading them to conclude that JCPyV incidence in the urine increases with age (21). Furthermore, they found that the JCPyV seropositive rate was approximately 73% in all subjects older than 20 years (21). While we did not find any statistically significant association between JCPyV and age or race/ethnicity when considered individually, when both were considered in a two-way interaction model, we found that females in the 70+ age group were 24 times more likely to be JCPyV positive than the under 40 female population. The final reduced model also identified that females 40-49 were more likely to be JCPyV+ than females < 40, but strikingly, this same observation was not made for the 50-59- and 60–69-year-old cohorts. Thus, our study found that the incidence of JCPyV in the urine does not increase with age (**Table 16**), as previously suggested (21).

Another study looked at the prevalence of JCPyV antibodies in patients with MS across 10 countries and 4 races (White, Black, Asian, and Other) (121). Most relevant to our own work here, they found that seroprevalence across the four race types were generally consistent. Our final reduced model found that the White/Caucasian race/ethnicity group is less likely to have JCPyV than the Black/African American race/ethnicity group. This differs from a prior study that found a JCPyV antibody seroprevalence percentage of 57% for White/Caucasian participants and 56% for Black/African American participants (121). However, this trend was not observed in our univariant model (**Figure 7**), when race/ethnicity was considered alone. We consistently saw a trend in the data of White/Caucasian females displaying a significantly lower prevalence of JCPyV than females of other races/ethnicities, but it could be because this group had a larger sample size in our data set.

Because the biomass of the urobiome is known to be quite low (9), we next tested the hypothesis that JCPyV was only being detected in urobiomes with a greater biomass. Using qPCR of the bacterial 16S rRNA gene sequences as a proxy for biomass, we examined the relative abundance of JCPyV virons to bacterial biomass. Even for the samples from UTI+ females, which are known to have a greater bacterial abundance than individuals without LUTS (91), no difference in the relative abundance of JCPyV was observed (**Figure 10**). Thus, we find no association with the abundance of JCPyV and symptom status. To our knowledge, this is the only study to analyze abundance of JCPyV relative to OAB and UTIs specifically. Only one other study has looked at the association of JCPyV presence with regards to LUTS patients (46).

To further explore the contribution of the urobiome biomass and JCPyV presence, we compared the bacterial urobiome “size” between matched JCPyV+ and JCPyV- samples.

Although no statistically significant association was detected, when we looked only at the UTI+ group, we did detect a signal; there were fewer bacterial cells in JCPyV+ samples than in JCPyV- samples (**Figure 13**). This would suggest that JCPyV+ females with UTI symptoms had less bacteria in their urinary tract than JCPyV- females with UTI symptoms. It is important to note, however, that there were only 8 pairs evaluated here and there was considerable variation between JCPyV+/UTI+ samples (**Table 17**). In total, only 33 samples from UTI+ females were included in this study. Moreover, most of these samples were from White/Caucasian individuals (64%). Future investigations are needed to ascertain if JCPyV is associated with urobiome biomass in UTI+ females. No other studies that have specifically looked at the association between JCPyV presence and UTIs.

Just like before with age, race/ethnicity, and symptom we explored associations of taxa with JCPyV incidence using a two-way ANOVA. For instance, at the Phylum level, we found 28 unique levels of Phylum for individuals in our data and tested for significance of interaction effect between Phylum and JCPyV using a two-way ANOVA model. Our test had a p-value of 1 indicating that this interaction effect was not statistically significant. In other words, bacterial taxa at Phylum level had no significant association with presence/absence of JCPyV. We ran similar tests at every other level of taxa (i.e., Family, Order, and Genus). None of those p-values were significant or even close to the level 0.05. As there are no other studies that have specifically looked at the association between JCPyV presence and bacterial composition of the bladder, every discovery made here is novel.

While our data cannot support an association between JCPyV and LUTS, we cannot definitively prove that one does not exist. Here we evaluated JCPyV in the bladder, using

catheterized urine samples. We know that JCPyV resides in the kidneys, and thus may not be shed in all individuals. The unequal sample size of race/ethnicity in the data may be insufficient to see significance - obtaining a larger sample size and one with equal distribution of age and race/ethnicity would be necessary for further research. It is also worth noting that while utilizing a p-value of 0.05 for analyzing statistically significance is generally accepted, it is not the only factor that can prove interactions between variables. We have seen several instances where the p-value came close to the 0.05 significance threshold, and that may be an indication that there is something worth looking into even though it cannot be statistically stated as significant. Therefore, it is imperative while considering these results to acknowledge that the p-value is a tool to use in understanding complex data and there may be interactions between JCPyV and age, race/ethnicity, and/or symptom status regardless of their statistical significance.

CHAPTER FOUR

CONCLUSIONS

The microbiota of the urinary tract are understudied. It was only recently that bacteria were identified in "healthy" bladders (2), but even less is known about the viruses that persist within the urinary tract. This is despite the fact that we have known for over five decades that polyomaviruses are found in urine. My study was driven by our lab's prior observation that JCPyV was only found in the urine samples of females with OAB (15). However, this prior observation was based on a small number of OAB (n=20) and noLUTS (n=10) samples. To date, very few studies have explored JCPyV's contribution to lower urinary tract symptoms (46).

Our new bioinformatic method for detecting and distinguishing between closely related viral species in metagenomes found a low level of JCPyV (~1.8%) in 165 urinary shotgun metagenomes. This is far lower than the 20-80% reported in the literature, which were determined via PCR-based or serology-based tests. In tandem, we assessed PCR primers used in the literature and designed a new pair of ultra-specific and sensitive JCPyV primers. By conducting shotgun metagenomic sequencing of a JCPyV PCR-positive sample, we found that the sample would in fact fail our bioinformatic test for JCPyV detection as it was not sufficiently abundant in the sample that we sequenced. Furthermore, we confirmed that BKPyV was not present in the sample, demonstrating that the PCR primers were accurately detecting JCPyV. Thus, we conclude that metagenomic sequencing can be effective in detecting JCPyV when it is abundant in an individual's urine and, in such cases, can be used to reconstruct JCPyV genomes.

More importantly, this effort confirms that PCR has the potential to accurately detect JCPyV, even at low levels, and distinguish between BKPyV when the primers used are ultra-specific and sensitive. The primers that we have developed are ideal candidates for future PCR-based tests of JCPyV in urine.

Armed with our JCPyV-specific primers, we tested 190 urine samples of females with and without LUTS (OAB and UTIs) to explore JCPyV's potential association with LUTS. Among our 190 urine samples tested, we only identified 52 positive samples for JCPyV (~27%). This falls at the lower end of estimates of its presence in the human population, although our samples were all from females. JCPyV presence was not found to be significantly associated with symptom status, age, or race/ethnicity when considered individually. Although, White/Caucasian females were seen to have a p-value that was close to the boundary in the race/ethnicity univariate model. In the age and race/ethnicity two-way interaction model, females aged 70 years and older had a significantly higher prevalence of JCPyV compared to females younger than 40. While none of the other interactions models found anything of significance, there were instances in the symptom and race/ethnicity interaction model where the p-values came close to the threshold. Finally, the final reduced model again found significance with females aged 70 years and older; significance was also found with females aged 40-50 and White/Caucasian females. Furthermore, there was no clear difference between the symptom status of individuals and the relative abundance of JCPyV or bacteria in their urinary tract. However, the presence of outliers in the data indicate there is an influx of JCPyV present as compared to bacterial load in certain samples and are worth further investigation. There also was no association between symptom status and urobiome composition in JCPyV positive females.

Overall, this study looked at the relationship of two closely related polyomaviruses and made efforts to distinguish detection between them. It is known that JCPyV persists in other anatomical sites (i.e., spleen, tonsils, bone marrow) (37–40), and future studies may explore if JCPyV plays a role in these microbiota. While we did not find that JCPyV presence or abundance has a significant association with LUTS, a more nuanced association may exist. Future studies that include a larger sample size and more aspects of demographic data (beyond just age and race/ethnicity) may reveal an association. In the meantime, the contribution of JCPyV to the urinary tract microbiome remains an open question.

APPENDIX A

METAGENOME DATA

Accession Number
AB038249.1
AB038250.1
AB038251.1
AB038252.1
AB038253.1
AB038254.1
AB038255.1
AB048545.1
AB048546.1
AB048547.1
AB048548.1
AB048549.1
AB048550.1
AB048551.1
AB048552.1
AB048553.1
AB048554.1
AB048555.1
AB048556.1
AB048557.1
AB048558.1
AB048559.1
AB048560.1
AB048561.1
AB048562.1
AB048563.1
AB048564.1
AB048565.1
AB048566.1
AB048567.1
AB048568.1
AB048569.1
AB048570.1
AB048571.1
AB048572.1
AB048573.1
AB048574.1

AB048575.1
AB048576.1
AB048577.1
AB048578.1
AB048579.1
AB048580.1
AB048581.1
AB048582.1
AB074575.1
AB074576.1
AB074577.1
AB074578.1
AB074579.1
AB074580.1
AB074581.1
AB074582.1
AB074583.1
AB074584.1
AB074585.1
AB074586.1
AB074587.1
AB074588.1
AB074589.1
AB074590.1
AB074591.1
AB077855.1
AB077856.1
AB077857.1
AB077858.1
AB077859.1
AB077860.1
AB077861.1
AB077862.1
AB077863.1
AB077864.1
AB077865.1
AB077866.1
AB077867.1

AB077868.1
AB077869.1
AB077870.1
AB077871.1
AB077872.1
AB077873.1
AB077874.1
AB077875.1
AB077876.1
AB077877.1
AB077878.1
AB077879.1
AB081005.1
AB081006.1
AB081007.1
AB081008.1
AB081009.1
AB081010.1
AB081011.1
AB081012.1
AB081013.1
AB081014.1
AB081015.1
AB081016.1
AB081017.1
AB081018.1
AB081019.1
AB081020.1
AB081021.1
AB081022.1
AB081023.1
AB081024.1
AB081025.1
AB081026.1
AB081027.1
AB081028.1
AB081029.1
AB081030.1

AB081600.1
AB081601.1
AB081602.1
AB081603.1
AB081604.1
AB081605.1
AB081606.1
AB081607.1
AB081608.1
AB081609.1
AB081610.1
AB081611.1
AB081612.1
AB081613.1
AB081614.1
AB081615.1
AB081616.1
AB081617.1
AB081618.1
AB081654.1
AB092578.1
AB092579.1
AB092580.1
AB092581.1
AB092582.1
AB092583.1
AB092584.1
AB092585.1
AB092586.1
AB092587.1
AB103387.1
AB103402.1
AB103403.1
AB103404.1
AB103405.1
AB103406.1
AB103407.1
AB103408.1

AB103409.1
AB103410.1
AB103411.1
AB103412.1
AB103413.1
AB103414.1
AB103415.1
AB103416.1
AB103417.1
AB103418.1
AB103419.1
AB103420.1
AB103421.1
AB103422.1
AB103423.1
AB104487.1
AB113118.1
AB113119.1
AB113120.1
AB113121.1
AB113122.1
AB113123.1
AB113124.1
AB113125.1
AB113126.1
AB113127.1
AB113128.1
AB113129.1
AB113130.1
AB113131.1
AB113132.1
AB113133.1
AB113134.1
AB113135.1
AB113136.1
AB113137.1
AB113138.1
AB113139.1

AB113140.1
AB113141.1
AB113142.1
AB113143.1
AB113144.1
AB113145.1
AB113216.1
AB113217.1
AB118231.1
AB118232.1
AB118233.1
AB118234.1
AB118235.1
AB118651.1
AB118652.1
AB118653.1
AB118654.1
AB118655.1
AB118656.1
AB118657.1
AB118658.1
AB118659.1
AB126981.1
AB126982.1
AB126983.1
AB126984.1
AB126985.1
AB126986.1
AB126987.1
AB126988.1
AB126989.1
AB126990.1
AB126991.1
AB126992.1
AB126993.1
AB126994.1
AB126995.1
AB126996.1

AB126997.1
AB126998.1
AB126999.1
AB127000.1
AB127001.1
AB127002.1
AB127003.1
AB127004.1
AB127005.1
AB127006.1
AB127007.1
AB127008.1
AB127009.1
AB127010.1
AB127011.1
AB127012.1
AB127013.1
AB127014.1
AB127015.1
AB127016.1
AB127017.1
AB127018.1
AB127019.1
AB127020.1
AB127021.1
AB127022.1
AB127023.1
AB127024.1
AB127025.1
AB127026.1
AB127027.1
AB127342.1
AB127343.2
AB127344.1
AB127345.2
AB127346.1
AB127347.1
AB127348.1

AB127349.1
AB127350.2
AB127351.2
AB127352.1
AB127353.1
AB183152.1
AB185020.1
AB195639.1
AB195640.1
AB198940.1
AB198941.1
AB198942.1
AB198943.1
AB198944.1
AB198945.1
AB198946.1
AB198947.1
AB198948.1
AB198949.1
AB198950.1
AB198951.1
AB198952.1
AB198953.1
AB198954.1
AB212952.1
AB212953.1
AB212954.1
AB220939.1
AB220940.1
AB220941.1
AB220942.1
AB220943.1
AB262396.1
AB262397.1
AB262398.1
AB262399.1
AB262400.1
AB262401.1

AB262402.1
AB262403.1
AB262404.1
AB262405.1
AB262406.1
AB262407.1
AB262408.1
AB262409.1
AB262410.1
AB262411.1
AB262412.1
AB262413.1
AB362351.1
AB362352.1
AB362353.1
AB362354.1
AB362355.1
AB362356.1
AB362357.1
AB362358.1
AB362359.1
AB362360.1
AB362361.1
AB362362.1
AB362363.1
AB362364.1
AB362365.1
AB362366.1
AB372036.1
AB372037.1
AB372038.1
AF004349.1
AF004350.1
AF030085.1
AF295733.1
AF295734.1
AF295736.1
AF300945.1

AF300946.1
AF300947.1
AF300948.1
AF300949.1
AF300950.1
AF300951.1
AF300952.1
AF300953.1
AF300954.1
AF300955.1
AF300956.1
AF300957.1
AF300958.1
AF300959.1
AF300960.1
AF300961.1
AF300962.1
AF300963.1
AF300964.1
AF300965.1
AF300966.1
AF300967.1
AF363830.1
AF363831.1
AF363832.1
AF363833.1
AF363834.1
AY121907.1
AY121908.1
AY121909.1
AY121910.1
AY121911.1
AY121912.1
AY121913.1
AY121914.1
AY121915.1
AY328376.1
AY342299.1

AY349147.1
AY356539.1
AY364314.1
AY366359.1
AY373463.1
AY376828.1
AY376829.1
AY376830.1
AY376831.1
AY378084.1
AY378085.1
AY378086.1
AY378087.1
AY382184.1
AY382185.1
AY382186.1
AY382187.1
AY382188.1
AY386373.1
AY386374.1
AY386375.1
AY386376.1
AY386377.1
AY386378.1
AY536239.1
AY536240.1
AY536241.1
AY536242.1
AY536243.1
DQ875211.1
DQ875212.1
HB980181.1
HG764413.1
J02226.1
JA916531.1
JB982623.1
JB982624.1
JF424834.1

JF424835.1
JF424836.1
JF424837.1
JF424838.1
JF424839.1
JF424840.1
JF424841.1
JF424842.1
JF424843.1
JF424844.1
JF424845.1
JF424846.1
JF424847.1
JF424848.1
JF424849.1
JF424850.1
JF424851.1
JF424852.1
JF424853.1
JF424854.1
JF424855.1
JF424856.1
JF424857.1
JF424858.1
JF424859.1
JF424860.1
JF424861.1
JF424862.1
JF424863.1
JF424864.1
JF424865.1
JF424866.1
JF424867.1
JF424868.1
JF424869.1
JF424870.1
JF424871.1
JF424872.1

JF424873.1
JF424874.1
JF424875.1
JF424877.1
JF424878.1
JF424879.1
JF424880.1
JF424881.1
JF424882.1
JF424883.1
JF424884.1
JF424885.1
JF424886.1
JF424887.1
JF424888.1
JF424889.1
JF424890.1
JF424891.1
JF424892.1
JF424893.1
JF424894.1
JF424895.1
JF424896.1
JF424897.1
JF424898.1
JF424899.1
JF424900.1
JF424901.1
JF424902.1
JF424903.1
JF424904.1
JF424905.1
JF424906.1
JF424907.1
JF424908.1
JF424909.1
JF424910.1
JF424911.1

JF424912.1
JF424913.1
JF424914.1
JF424915.1
JF424916.1
JF424917.1
JF424918.1
JF424919.1
JF424920.1
JF424921.1
JF424922.1
JF424923.1
JF424924.1
JF424925.1
JF424926.1
JF424927.1
JF424928.1
JF424929.1
JF424930.1
JF424931.1
JF424932.1
JF424933.1
JF424934.1
JF424935.1
JF424936.1
JF424937.1
JF424938.1
JF424939.1
JF424940.1
JF424941.1
JF424942.1
JF424943.1
JF424944.1
JF424945.1
JF424946.1
JF424947.1
JF424948.1
JF424949.1

JF424950.1
JF424951.1
JF424952.1
JF424953.1
JF424954.1
JF424955.1
JF424956.1
JF424957.1
JF424958.1
JF424959.1
JF424960.1
JF424961.1
JF424962.1
JF425488.1
JF425489.1
JF425490.1
JF425491.1
JF425492.1
JF425493.1
JF425494.1
JF425495.1
JF425496.1
JF425497.1
JF425498.1
JF425499.1
JF425500.1
JF425501.1
JF425502.1
JF425503.1
JF425504.1
JF425551.1
JF425552.1
JF425553.1
JF425554.1
JF425555.1
JF425556.1
JQ237146.1
JX273163.1

KJ659286.1
KJ659287.1
KJ659288.1
KJ659289.1
KM225765.1
LC164349.1
LC164350.1
LC164351.1
LC164352.1
LC164353.1
LC164354.1
LC422956.1
LT615219.1
LT615220.1
LT615221.1
LT615222.1
LT615223.1
MF662180.1
MF662181.1
MF662182.1
MF662183.1
MF662184.1
MF662185.1
MF662186.1
MF662187.1
MF662188.1
MF662189.1
MF662190.1
MF662191.1
MF662192.1
MF662193.1
MF662194.1
MF662195.1
MF662196.1
MF662197.1
MF662198.1
MF662199.1
MF662200.1

MF662201.1
MF662202.1
MF662203.1
MF662204.1
MF662205.1
MF662206.1
MF662207.1
MF662208.1
NC_001699.1
U61771.1
U73500.1
U73501.1
U73502.1

APPENDIX B

PRIMER SEQUENCES

First author	Primers & Probe Sequences	Predicted to Hybridize to BKPyV	Link to Paper
Agostini	ACAGTGTGGCCAGAATTCCACTACC (JLP15 primer), TAAAGCCTCCCCCCCAACAGAAA (JLP16 primer)	JLP15	https://pubmed.ncbi.nlm.nih.gov/11297697/
Baksh	Set 2 consisted of oligonucleotide primers specific for the V-T intergenic region of JCV only (nucleotides 2537 to 2703; 167 bp; MAD-1 numbering (refr: 46 -- Frisque RJ, Bream GL, Cannella MT: Human polyomavirus JCV genome. J Virol 51:458-469, 1984)	-	https://pubmed.ncbi.nlm.nih.gov/11479162/
Bogdanovic	Nested PCR: GTATACACAGCAAAGGAAGC (P-3) and GCTCATCAGCCTGATTTTGG (P-4) [outer]; AGTCTTTAGGGTCTTCTACC (P-1) and GGTGCCAACCTATGGAACAG (P-2) [inner]; differentiate by digestion with BamHI	P3, P4, P1, P2	https://pubmed.ncbi.nlm.nih.gov/15566767/
Cayres-Vallinoto	ACAGTGTGGCCAGAATTCCACTACC (JLP15 primer), TAAAGCCTCCCCCCCAACAGAAA (JLP16 primer); TTTTGGGACACTAACAGGAGG (P-1-F) and AGCAGAAGACTCTGGACATGG (P-2-R)	JLP15, P2R	https://pubmed.ncbi.nlm.nih.gov/23071582/
da Silva Nali	ACAGTGTGGCCAGAATTCCACTACC (JLP15 primer), TAAAGCCTCCCCCCCAACAGAAA (JLP16 primer)	JLP15	https://pubmed.ncbi.nlm.nih.gov/22850991/
de Souza	CCAAAGGGAGGGAACCTATATT (JCf), GGCAACATCCATTGAGGAG (BKJCr)	BKJCr	https://pubmed.ncbi.nlm.nih.gov/30231168/
Del Valle	AGTCTTTAGGGTCTTCTACC (primer, Pep1), GGTGCCAACCTATGGAACAG (primer, Pep2), GTTGGGATCCTGTGTTTTCATC (probe)	Pep1, Pep2	https://pubmed.ncbi.nlm.nih.gov/11358858/
Dumonceaux	AGAGTGTTGGGATCCTGTGTTTT (primer 1), GAGAAGTGGGATGAAGACCTGTTT (primer 2), TCATCACTGGCAAACAT (probe 1), TCATCACTGGCAAACATTTCTTCATGGC (probe 2)	Probe 1, Probe 2	https://pubmed.ncbi.nlm.nih.gov/18614652/

Dumoulin	CTAAACACAGCTTGACTGAGGAATG (JC-Forward), CATTTAATGAGAAGTGGGATGAAGAC (JC-Reverse), TAGAGTGTTGGGATCCTGTGTTTTCATCATCACT (JC-Probe)	JCForward, JCReverse	https://pubmed.ncbi.nlm.nih.gov/21325560/
Elfaitouri*	GGG GAC CTA RTT GCY AST GT (BJS-FP), GCA ASR GAT GCA AKT TSMAC (BJS-RP), ACW GGA TTT TCA GTR GCT GAA ATT GCT GCT GG (Probe)	BJS-Probe	https://pubmed.ncbi.nlm.nih.gov/16677718/
Funahashi	AGAGTGTTGGGATCCTGTGTTTT (primer), GAGAAGTGGGATGAAGACCTGTTT (primer), TCATCACTGGCAAACATTTCTTCATGGC (probe)	Probe	https://pubmed.ncbi.nlm.nih.gov/20053854/
Giovannelli	AGTGTTGGGATCCTGTGTTTTCA (JCT-3F forward primer), GTGGGATGAAGACCTGTTTTGC (JCT-4R reverse primer), and CATCACTGGCAAACAT (TaqMan MGB JCT-1.2 probe); GGAGCCCTGGCTGCAT (JRR-1F forward primer), TGTGATTAAGGACTATGGGAGG (JRR-2R reverse primer), and CTGGCAGTTATAGTGAAACC (MGB JRR-1.1 probe)	JCT-4R; JCT-1.2 Probe	https://pubmed.ncbi.nlm.nih.gov/27454232/
Haghighi	AAGTCTTTAGGGTCTTCTAC (primer) and GAATCCTGGTGGAATACA (primer)	primer1	https://pubmed.ncbi.nlm.nih.gov/30807732/
Hori	AGT CTT TAG GGT CTT CTA (JCT-1F) and GGT GCC AAC CTA TGG AAC AG (JCT-1R); nested primers JCT-1F and TGA AGA CCT GTT TTG CCA TG (JCT-2R)	JCT1F, JCT1R	https://pubmed.ncbi.nlm.nih.gov/16021515/
Hori (others)	TGT GCA CTC TAA TGG GCA AGC (VP Primer1), CTA GGT ACG CCT TGT GCT CTG (VP Primer 2), GAT TGC ACT GTG GCA TTC TTT GG (VP Primer 3; nested with VP Primer 1), AGC CAG TGC AGG GCA CCA GC (VP Probe); GTC TGC TCA GTC AAA CCA CTG (Agnoprotein Primer 1), GTT CTT CGC CAG CTG TCA C (Agnoprotein Primer 2), GCA CAG GTG AAG ACA GTG TAG (Agnoprotein Primer 3; nested with	Agno.2, Agno.3	https://pubmed.ncbi.nlm.nih.gov/16021515/

	Agnoprotein Primer 1), AAA GAC AGA GAC ACA GTG GTT (Probe)		
Hu	ACAGTGTGGCCAGAATTCCACTACC (JLP15 primer), TAAAGCCTCCCCCCCAACAGAAA (JLP16 primer)	JLP15	https://pubmed.ncbi.nlm.nih.gov/29322824/
Hussain	AGAGTGTTGGGATCCTGTGTTTT (forward primer), TTGCAGGGCATTTTGTTTTTTAC (reverse primer)	-	https://pubmed.ncbi.nlm.nih.gov/28438210/
Karalic	ttccactacccaatctaaatgagg (P13 primer) and gtttgtaaacatgccacagacatc (M5 primer); semi-nested M5 and ctcatgtgggaggctgtgacct (JLP1). P13 and M5 from Lafon et al. 1998; JLP1 from Agostini et al. 1996	P13, JLP1	https://pubmed.ncbi.nlm.nih.gov/24123117/
Kruzel-Davila	GAGTGTTGGGATCCTGTGTTTT (forward), AGAAGTGGGATGAAGACCTGTTT (reverse)	-	https://pubmed.ncbi.nlm.nih.gov/30715336/
Makvandi	ACAGCATTCAAGAAGTTACCCA (F3J), CCCCTGTAATTCTAAAGCCTCC (R3J), ACTCTAATGGTCAAGCAACTC (F4J), GCAACTGTAAAGTAAAGCTGG (R4J)	F3J	https://pubmed.ncbi.nlm.nih.gov/31364012/
McNees	ttcttcatggcaaaacaggtctt (primer), ttccaccaggattcccattc (primer - need rc), ccacttctcattaaatg (probe)	primer 1, primer 2	https://pubmed.ncbi.nlm.nih.gov/16087125/
Merlino	AAGTATTCCTTATTCACACC (primer), AACTTTTATAAGTAGACATGGT (primer)	primer1	https://pubmed.ncbi.nlm.nih.gov/15805571/
Mou	AGT CTT TAG GGT CTT CTA (JCT-1F) and GGT GCC AAC CTA TGG AAC AG (JCT-1R); nested primers JCT-1F and TGA AGA CCT GTT TTG CCA TG (JCT-2R)	JCT1F, JCT1R	https://pubmed.ncbi.nlm.nih.gov/22606241/
Pal	JE3 -- ATGTTTGCCAGTGATGATGAAAA (primer), GGAAAGTCTTTAGGGTCTTCTACCTTT (primer), AGGATCCCAACACTCTACCCCACCTAAAAAGA (probe); JL1 -- AAGGGAGGGAACCTATATTTCTTTTG (primer), TCTAGCCTTTGGGTAACTTCTTGAA (primer), CTCATACACCCAAAGTATAGATGATGCAGACAGC	JE31, JE32, JE3Probe, JL1.2, JL1.Probe, JL4.2, JL4.Probe	https://pubmed.ncbi.nlm.nih.gov/16527364/

	A (probe); JL2 -- GGTTTAGGCCAGTTGCTGACTT (primer), GTCTCCCCATACCAACATTAGCTT (primer), TCTTTCCACTGCACAATCCTCTCATGAATG (probe); JL4 -- GAGGTGCAAATCAAAGATCTGCT (primer), GGGCCATCTTCATATGCTTCA (primer), AGTCCCGTACAACCCTAAAAGTAAAGGCAACA (probe)		
Rafique	CCCTATTCAGCACTTTGTCC (JR1) and CAAACCACTGTGTCTCTGTC (JR2); nested: GGGAATTTCCCTGGCCTCCT-(JR3) and ACTTTCACAGAAGCCTTACG (JR4)	-	https://pubmed.ncbi.nlm.nih.gov/18475011/
Rencic	TTCCTCCCTATTCAGCACTT (primer pair 1), AAAACAGCTCTGGCTCGCAA (primer pair 1), CCCCATACCAACATTAGCTTTC (primer pair 2), CCAGATTTGTAAGGCAGATAG (primer pair 2)	Primer Pair 2 reverse	https://pubmed.ncbi.nlm.nih.gov/8692997/
Ryschkewitsch	AGAGTGTTGGGATCCTGTGTTTT (JCT-1, primer), GAGAAGTGGGATGAAGACCTGTTT (JCT-2, primer), TCATCACTGGCAAACATTTCTTCATGGC (JCT, probe); GCAGCTTAGTGATTTTCTCAGG (JTP-1, primer), CACCAAAACAAAAGAACACAGG (JTP-2, primer), CTGTAAAGTTCTAGGCACTGAATAT (JTP, probe)	JCT-Probe; JTP-2	https://pubmed.ncbi.nlm.nih.gov/15381359/
Sehbani	AAGAAATTAAACCTTTCAACTAAC (JCV-ABI-F1), CTGCAAAAATTTGGGCATTATA (JCV-UCL-R1)	-	https://pubmed.ncbi.nlm.nih.gov/16542870/
Tsai*	JBR1 (58-CCTCCACGCCCTTACTACTTCTGAG-38) and JBR2 (58-GTGACAGCTGGCGAAGAAC-CATGGC-	JBR1,JBR2	https://pubmed.ncbi.nlm.nih.gov/9210032/
Urbano	ACAGTGTGGCCAGAATTCCACTACC (JLP15 primer), TAAAGCCTCCCCCCCAACAGAAA (JLP16 primer)	JLP15	https://pubmed.ncbi.nlm.nih.gov/26147595/

APPENDIX C

JCPyV MAPPING

SRA Number	Reference to Paper	total # of reads	mapped to JCPyV?	# mapped reads	% reads mapped	good consensus?
ERR2798113	Kidney Transplant Virome (Rani et al.)	7,849,916	YES	243,930	3.11	NO
ERR2798114	Kidney Transplant Virome (Rani et al.)	7,329,554	YES	213,032	2.91	NO
ERR2798115	Kidney Transplant Virome (Rani et al.)	10,901,881	YES	117,781	1.08	NO
ERR2798116	Kidney Transplant Virome (Rani et al.)	9,410,822	YES	280,672	2.98	NO
ERR2798117	Kidney Transplant Virome (Rani et al.)	6,826,287	YES	213,582	3.13	NO
ERR2798118	Kidney Transplant Virome (Rani et al.)	11,521	YES	323	2.80	NO
ERR2798119	Kidney Transplant Virome (Rani et al.)	12,785,361	YES	336,004	2.63	NO
ERR2798120	Kidney Transplant Virome (Rani et al.)	8,509,165	YES	125,852	1.48	NO
ERR2798121	Kidney Transplant Virome (Rani et al.)	2,049,492	YES	157	0.01	NO
ERR2798122	Kidney Transplant Virome (Rani et al.)	3,139,995	YES	24,000	0.76	NO
ERR2798123	Kidney Transplant Virome (Rani et al.)	5,832,624	YES	116,070	1.99	NO
ERR2798124	Kidney Transplant Virome (Rani et al.)	15,636,176	YES	21,553	0.14	NO
ERR2798125	Kidney Transplant Virome (Rani et al.)	13,710,307	YES	90,043	0.66	YES
ERR2798126	Kidney Transplant Virome (Rani et al.)	15,725,590	YES	319,807	2.03	YES
ERR2798127	Kidney Transplant Virome (Rani et al.)	14,285,402	YES	97	0.00	NO

ERR2798128	Kidney Transplant Virome (Rani et al.)	17,578,195	YES	188	0.00	NO
ERR2798131	Kidney Transplant Virome (Rani et al.)	281,816	YES	2	0.00	NO
ERR2798139	Kidney Transplant Virome (Rani et al.)	295,740	NO	--	--	--
ERR2798138	Kidney Transplant Virome (Rani et al.)	563,269	NO	--	--	--
ERR2798137	Kidney Transplant Virome (Rani et al.)	446,857	NO	--	--	--
ERR2798136	Kidney Transplant Virome (Rani et al.)	881,827	NO	--	--	--
ERR2798135	Kidney Transplant Virome (Rani et al.)	689,472	NO	--	--	--
ERR2798134	Kidney Transplant Virome (Rani et al.)	580,114	NO	--	--	--
ERR2798133	Kidney Transplant Virome (Rani et al.)	399,222	NO	--	--	--
ERR2798132	Kidney Transplant Virome (Rani et al.)	779,563	NO	--	--	--
ERR2798130	Kidney Transplant Virome (Rani et al.)	392,172	NO	--	--	--
ERR2798129	Kidney Transplant Virome (Rani et al.)	642,591	NO	--	--	--
SRR6519218	Pulmonary Tuberculosis Urinary Microbiome (unpublished)	3,304,590	YES	21,724	0.66	YES
SRR6519219	Pulmonary Tuberculosis Urinary Microbiome (unpublished)	4,190,598	YES	1	0.00	NO
SRR6519220	Pulmonary Tuberculosis Urinary Microbiome (unpublished)	2,001,846	NO	--	--	--

SRR5535770	Microbial Metagenome of UTI (Moustafa et al.)	35,752,252	YES	12	0.00	NO
SRR5535725	Microbial Metagenome of UTI (Moustafa et al.)	2,551,900	YES	2	0.00	NO
SRR5535727	Microbial Metagenome of UTI (Moustafa et al.)	1,014,752	YES	8	0.00	NO
SRR5535728	Microbial Metagenome of UTI (Moustafa et al.)	1,133,228	YES	2	0.00	NO
SRR5535745	Microbial Metagenome of UTI (Moustafa et al.)	10,367,022	YES	2	0.00	NO
SRR5535752	Microbial Metagenome of UTI (Moustafa et al.)	12,426,018	YES	313	0.00	NO
SRR5535757	Microbial Metagenome of UTI (Moustafa et al.)	38,132,956	YES	10	0.00	NO
SRR5535763	Microbial Metagenome of UTI (Moustafa et al.)	4,919,760	YES	73	0.00	NO
SRR5535768	Microbial Metagenome of UTI (Moustafa et al.)	1,701,162	YES	4	0.00	NO
SRR5535772	Microbial Metagenome of UTI (Moustafa et al.)	18,391,406	NO	--	--	--
SRR5535771	Microbial Metagenome of UTI (Moustafa et al.)	860,248	NO	--	--	--
SRR5535769	Microbial Metagenome of UTI (Moustafa et al.)	2,054,042	NO	--	--	--
SRR5535767	Microbial Metagenome of UTI (Moustafa et al.)	24,638,130	NO	--	--	--
SRR5535766	Microbial Metagenome of UTI (Moustafa et al.)	3,714,246	NO	--	--	--
SRR5535765	Microbial Metagenome of UTI (Moustafa et al.)	750,482	NO	--	--	--

SRR5535764	Microbial Metagenome of UTI (Moustafa et al.)	881,546	NO	--	--	--
SRR5535762	Microbial Metagenome of UTI (Moustafa et al.)	19,306,142	NO	--	--	--
SRR5535761	Microbial Metagenome of UTI (Moustafa et al.)	7,075,952	NO	--	--	--
SRR5535760	Microbial Metagenome of UTI (Moustafa et al.)	2,733,440	NO	--	--	--
SRR5535759	Microbial Metagenome of UTI (Moustafa et al.)	1,216,950	NO	--	--	--
SRR5535758	Microbial Metagenome of UTI (Moustafa et al.)	1,427,178	NO	--	--	--
SRR5535756	Microbial Metagenome of UTI (Moustafa et al.)	6,845,858	NO	--	--	--
SRR5535755	Microbial Metagenome of UTI (Moustafa et al.)	3,499,650	NO	--	--	--
SRR5535754	Microbial Metagenome of UTI (Moustafa et al.)	7,572,096	NO	--	--	--
SRR5535753	Microbial Metagenome of UTI (Moustafa et al.)	12,409,778	NO	--	--	--
SRR5535751	Microbial Metagenome of UTI (Moustafa et al.)	52,260,132	NO	--	--	--
SRR5535750	Microbial Metagenome of UTI (Moustafa et al.)	3,906,042	NO	--	--	--
SRR5535749	Microbial Metagenome of UTI (Moustafa et al.)	66,782,998	NO	--	--	--
SRR5535748	Microbial Metagenome of UTI (Moustafa et al.)	2,811,688	NO	--	--	--
SRR5535747	Microbial Metagenome of UTI (Moustafa et al.)	2,166,950	NO	--	--	--

SRR5535746	Microbial Metagenome of UTI (Moustafa et al.)	30,653,920	NO	--	--	--
SRR5535744	Microbial Metagenome of UTI (Moustafa et al.)	12,009,642	NO	--	--	--
SRR5535743	Microbial Metagenome of UTI (Moustafa et al.)	11,063,948	NO	--	--	--
SRR5535742	Microbial Metagenome of UTI (Moustafa et al.)	9,581,612	NO	--	--	--
SRR5535741	Microbial Metagenome of UTI (Moustafa et al.)	972,308	NO	--	--	--
SRR5535740	Microbial Metagenome of UTI (Moustafa et al.)	6,031,910	NO	--	--	--
SRR5535739	Microbial Metagenome of UTI (Moustafa et al.)	37,260,038	NO	--	--	--
SRR5535738	Microbial Metagenome of UTI (Moustafa et al.)	11,553,214	NO	--	--	--
SRR5535737	Microbial Metagenome of UTI (Moustafa et al.)	8,628,166	NO	--	--	--
SRR5535736	Microbial Metagenome of UTI (Moustafa et al.)	16,679,908	NO	--	--	--
SRR5535735	Microbial Metagenome of UTI (Moustafa et al.)	1,616,642	NO	--	--	--
SRR5535734	Microbial Metagenome of UTI (Moustafa et al.)	3,108,646	NO	--	--	--
SRR5535733	Microbial Metagenome of UTI (Moustafa et al.)	1,106,132	NO	--	--	--
SRR5535732	Microbial Metagenome of UTI (Moustafa et al.)	22,811,252	NO	--	--	--
SRR5535731	Microbial Metagenome of UTI (Moustafa et al.)	663,178	NO	--	--	--

SRR5535730	Microbial Metagenome of UTI (Moustafa et al.)	1,051,130	NO	--	--	--
SRR5535729	Microbial Metagenome of UTI (Moustafa et al.)	1,332,986	NO	--	--	--
SRR5535726	Microbial Metagenome of UTI (Moustafa et al.)	943,980	NO	--	--	--
SRR5535724	Microbial Metagenome of UTI (Moustafa et al.)	5,618,780	NO	--	--	--
SRR1038 7480	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,652,248	YES	1	0.00	NO
SRR1038 7483	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	2,376,972	YES	5,501	0.23	NO
SRR1038 7484	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	293,222	YES	714	0.24	NO
SRR1038 7486	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,867,774	YES	4,269	0.23	NO
SRR1038 7488	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	2,991,850	YES	5,125	0.17	NO
SRR1038 7489	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	486,290	YES	104	0.02	NO
SRR1038 7490	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	624,780	YES	231	0.04	NO
SRR1038 7552	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,864,536	YES	3,937	0.21	NO
SRR1038 7554	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,052,844	YES	1	0.00	NO
SRR1038 7555	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	938,294	YES	11	0.00	NO
SRR1038 7556	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	867,976	YES	574	0.07	NO

SRR1038 7558	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	91,862	YES	152	0.17	NO
SRR1038 7559	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	260,104	YES	157	0.06	NO
SRR1038 7566	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	442,614	YES	100	0.02	NO
SRR1038 7567	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	87,314	YES	126	0.14	NO
SRR1038 7568	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	728,368	YES	207	0.03	NO
SRR1038 7573	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,527,010	YES	1	0.00	NO
SRR1038 7576	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,115,988	YES	2,860	0.26	NO
SRR1038 7579	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,283,206	YES	1	0.00	NO
SRR1038 7583	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,305,680	YES	2	0.00	NO
SRR1038 7586	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	18,164	YES	40	0.22	NO
SRR1038 7589	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,496,056	YES	3,995	0.27	NO
SRR1038 7592	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,395,500	YES	5,674	0.41	NO
SRR1038 7593	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,114,760	YES	7	0.00	NO
SRR1038 7594	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,253,736	YES	3,837	0.31	NO
SRR1038 7596	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,084,862	YES	1	0.00	NO

SRR1038 7602	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,083,164	YES	3,605	0.33	NO
SRR1038 7605	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,164,426	NO	--	--	--
SRR1038 7604	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,167,366	NO	--	--	--
SRR1038 7603	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,066,084	NO	--	--	--
SRR1038 7601	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,283,714	NO	--	--	--
SRR1038 7600	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	2,022,602	NO	--	--	--
SRR1038 7599	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,763,842	NO	--	--	--
SRR1038 7598	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,259,724	NO	--	--	--
SRR1038 7591	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,112,520	NO	--	--	--
SRR1038 7590	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,007,546	NO	--	--	--
SRR1038 7588	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	442,762	NO	--	--	--
SRR1038 7587	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	17,004	NO	--	--	--
SRR1038 7582	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,231,892	NO	--	--	--
SRR1038 7581	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,188,690	NO	--	--	--
SRR1038 7580	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	982,456	NO	--	--	--

SRR1038 7578	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,371,862	NO	--	--	--
SRR1038 7577	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,218,612	NO	--	--	--
SRR1038 7575	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,596,834	NO	--	--	--
SRR1038 7572	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,834,876	NO	--	--	--
SRR1038 7571	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	478,212	NO	--	--	--
SRR1038 7570	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	3,345,314	NO	--	--	--
SRR1038 7569	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	578,760	NO	--	--	--
SRR1038 7565	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	305,948	NO	--	--	--
SRR1038 7564	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	319,214	NO	--	--	--
SRR1038 7562	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	835,388	NO	--	--	--
SRR1038 7561	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	617,474	NO	--	--	--
SRR1038 7560	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	583,092	NO	--	--	--
SRR1038 7557	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	935,144	NO	--	--	--
SRR1038 7553	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,300,096	NO	--	--	--
SRR1038 7550	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,287,206	NO	--	--	--

SRR1038 7549	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,316,196	NO	--	--	--
SRR1038 7482	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	572,498	NO	--	--	--
SRR1038 7481	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,905,960	NO	--	--	--
SRR1038 7479	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,087,456	NO	--	--	--
SRR1038 7478	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,887,036	NO	--	--	--
SRR1038 7477	Virome in Healthy and BK Disease of Kidney Transplant (unpublished)	1,600,496	NO	--	--	--
URP14	Virome in Association with UTI (Santiago-Rodriguez et al.)	724,661	NO	--	--	--
URP12	Virome in Association with UTI (Santiago-Rodriguez et al.)	691,906	NO	--	--	--
URP10	Virome in Association with UTI (Santiago-Rodriguez et al.)	624,144	NO	--	--	--
URP9	Virome in Association with UTI (Santiago-Rodriguez et al.)	768,502	NO	--	--	--
URP7	Virome in Association with UTI (Santiago-Rodriguez et al.)	785,652	NO	--	--	--
URP6	Virome in Association with UTI (Santiago-Rodriguez et al.)	564,278	NO	--	--	--
URP4	Virome in Association with UTI (Santiago-Rodriguez et al.)	766,694	NO	--	--	--
URP3	Virome in Association with UTI (Santiago-Rodriguez et al.)	514,004	NO	--	--	--
URP2	Virome in Association with UTI (Santiago-Rodriguez et al.)	632,903	NO	--	--	--

URP1	Virome in Association with UTI (Santiago-Rodriguez et al.)	662,120	NO	--	--	--
URN16	Virome in Association with UTI (Santiago-Rodriguez et al.)	857,887	NO	--	--	--
URN15	Virome in Association with UTI (Santiago-Rodriguez et al.)	631,819	NO	--	--	--
URN13	Virome in Association with UTI (Santiago-Rodriguez et al.)	729,718	NO	--	--	--
URN12	Virome in Association with UTI (Santiago-Rodriguez et al.)	1,116,705	NO	--	--	--
URN11	Virome in Association with UTI (Santiago-Rodriguez et al.)	1,003,850	NO	--	--	--
URN10	Virome in Association with UTI (Santiago-Rodriguez et al.)	650,536	NO	--	--	--
URN9	Virome in Association with UTI (Santiago-Rodriguez et al.)	658,048	NO	--	--	--
URN6	Virome in Association with UTI (Santiago-Rodriguez et al.)	765,829	NO	--	--	--
URN2	Virome in Association with UTI (Santiago-Rodriguez et al.)	858,681	NO	--	--	--
URN1	Virome in Association with UTI (Santiago-Rodriguez et al.)	534,412	NO	--	--	--
SRR7716742	Portable Urinal Microbiome (unpublished)	18,776,064	YES	21	0.00	NO
SRR7716743	Portable Urinal Microbiome (unpublished)	79,224,022	YES	304	0.00	NO
SRR7716744	Portable Urinal Microbiome (unpublished)	144,274,626	YES	54	0.00	NO
SRR7716745	Portable Urinal Microbiome (unpublished)	96,000,000	YES	472	0.00	NO

APPENDIX D

STATISICAL MODELS

	Estimate	Expected Odds	Standard Error	p-value
Age Univariate Model				
(40,50]	1.167661	3.214465205	0.69064	0.09091
(50,60]	0.58779	1.800006003	0.6675	0.37855
(60,70]	0.05716	1.058825209	0.67821	0.93284
(70,100]	0.83625	2.307696867	0.61741	0.1756
Race Univariate Model				
Hispanic/Latina	-0.09353	0.910710696	0.56982	0.8696
White/Caucasian	-0.81283	0.443600898	0.42955	0.0585
Symptom Univariate Model				
OAB+	0.1116	1.118065545	0.3847	0.77181
UTI+	0.4032	1.496606183	0.4847	0.405519
Age and Symptom Interaction Model				
(40,50]	1.2528	3.500129613	0.9335	0.18
(50,60]	1.0296	2.799945632	0.9599	0.283
(60,70]	1.0296	2.799945632	1.1276	0.361
(70,100]	1.9459	6.999928957	1.6036	0.225
OAB+	-13.6202	1.21569E-06	1455.3977	0.993
UTI+	1.5404	4.66645648	1.1852	0.194
(40,50] & OAB+	14.8241	2741730.916	1455.398	0.992
(50,60] & OAB+	13.5809	790878.7911	1455.3979	0.993
(60,70] & OAB+	12.9678	428394.5947	1455.3981	0.993
(70,100] & OAB+	12.7729	352533.1714	1455.3985	0.993
(40,50] & UTI+	-16.4134	7.44305E-08	1455.3981	0.991
(50,60] & UTI+	-1.3173	0.267857542	1.8041	0.465
(60,70] & UTI+	-2.2336	0.107142024	1.8179	0.219
(70,100] & UTI+	-1.8971	0.150002998	1.9099	0.321
Age and Race Interaction Model				
(40,50]	1.0986	2.999963134	1.633	0.5011
(50,60]	1.7918	6.00024319	1.291	0.1652
(60,70]	1.0986	2.999963134	1.3844	0.4275
(70,100]	3.1781	24.0011081	1.5546	0.0409
Hispanic/Latina	0.1823	1.199974132	1.5384	0.9057
White/Caucasian	0.539	1.714291713	1.3452	0.6887
(40,50] & Hispanic/Latina	0.5108	1.666623961	2.1292	0.8104
(50,60] & Hispanic/Latina	-0.1823	0.833351298	1.9664	0.9261
(60,70] & Hispanic/Latina	-0.1823	0.833351298	1.9664	0.9261

(70,100] & Hispanic/Latina	12.9975	441308.7399	882.7454	0.9883
(40,50] & White/Caucasian	-0.2513	0.777789	1.894	0.8944
(50,60] & White/Caucasian	-2.0971	0.122812067	1.6162	0.1944
(60,70] & White/Caucasian	-1.7177	0.179478474	1.6876	0.3088
(70,100] & White/Caucasian	-2.8548	0.057567333	1.7755	0.1079
Race and Symptom Interaction Model				
Hispanic/Latina	0.25131	1.285708592	1.11981	0.8224
White/Caucasian	0.62861	1.875002514	0.86763	0.4687
OAB+	1.50408	4.500011715	0.97183	0.1217
UTI+	2.19722	8.999958804	1.16667	0.0597
Hispanic/Latina & OAB+	-0.02817	0.972223075	1.42734	0.9843
White/Caucasian & OAB+	-1.93356	0.14463239	1.07961	0.0733
Hispanic/Latina & UTI+	-1.34993	0.259258408	1.68443	0.4229
White/Caucasian & UTI+	-2.48491	0.083333054	1.32864	0.0614
Full Model				
(40,50]	1.61819	5.043952495	2.41062	0.502
(50,60]	2.37901	10.7942113	1.88291	0.2064
(60,70]	1.81159	6.120170772	2.14931	0.3993
(70,100]	2.75832	15.7733215	2.66817	0.3012
OAB+	-12.59703	3.38205E-06	1455.39838	0.9931
UTI+	3.51244	33.52998121	2.18498	0.1079
Hispanic/Latina	-0.74516	0.474658349	2.76431	0.7875
White/Caucasian	1.59716	4.938985768	1.79936	0.3747
(40,50] & OAB+	15.01558	3320347.486	1455.39813	0.9918
(50,60] & OAB+	13.3117	604223.8374	1455.39814	0.9927
(60,70] & OAB+	12.97992	433618.329	1455.39818	0.9929
(70,100] & OAB+	13.35116	628543.175	1455.39853	0.9927
(40,50] & UTI+	-17.86924	1.73575E-08	1455.40011	0.9902
(50,60] & UTI+	-1.39622	0.247530864	2.36589	0.5551
(60,70] & UTI+	-2.04777	0.129022303	2.24461	0.3616
(70,100] & UTI+	-1.83059	0.160318952	2.3277	0.4316
Hispanic/Latina & OAB+	-0.04791	0.953219573	1.80216	0.9788
Hispanic/Latina & UTI+	-0.75862	0.468312252	2.84881	0.79
White/Caucasian & OAB+	-1.73874	0.175741696	1.35741	0.2002
White/Caucasian & UTI+	-2.61685	0.073032553	2.24375	0.2435
(40,50] & Hispanic/Latina	1.24527	3.473872618	3.30978	0.7067
(50,60] & Hispanic/Latina	0.52118	1.684013614	2.88993	0.8569

(60,70] & Hispanic/Latina	0.72011	2.054659211	2.94135	0.8066
(70,100] & Hispanic/Latina	15.67416	6415066.998	1455.4004	0.9914
(40,50] & White/Caucasian	-1.02852	0.357535722	2.53542	0.685
(50,60] & White/Caucasian	-2.25113	0.105280191	2.065	0.2757
(60,70] & White/Caucasian	-1.45425	0.23357548	2.23108	0.5145
(70,100] & White/Caucasian	-1.52802	0.216964833	2.29222	0.505

APPENDIX E

qPCR DATA

Sample ID	Symptom	JCPyV Signal	Bacterial Signal	Abundance Ratio
2656	OAB+	19.5011692	27.88180542	0.699422756
2662	OAB+	23.46768951	25.86221313	0.907412269
2720	OAB+	21.40681267	25.97675514	0.8240757
2769	OAB+	20.64847565	20.75850868	0.994699377
2981	OAB+	17.95087624	27.18789673	0.66025248
3000	OAB+	21.36740112	27.69773293	0.77144946
3001	OAB+	20.42099953	22.95108223	0.889761943
3002	OAB+	22.89677238	20.6443882	1.109103944
3022	OAB+	18.59108543	22.01903725	0.844318724
3127	OAB+	24.46646118	26.12587738	0.93648381
3131	OAB+	34.77993774	21.52444458	1.615834388
3241	OAB+	22.05724525	21.21721268	1.039592032
3363	OAB+	22.54339027	22.32190132	1.009922495
3765	OAB+	23.10961342	21.35718346	1.082053421
3356	OAB+	19.24667358	21.24588585	0.905901204
3791	OAB+	33.31470108	21.12228394	1.577230056
4021	OAB+	19.64975548	20.76813316	0.94614934
4214	OAB+	16.38001633	20.81211853	0.787042237
4304	noLUTS	24.21574593	20.80507469	1.163934583
4376	OAB+	17.33990479	18.70080376	0.927227782
4388	noLUTS	23.91977692	23.05531502	1.037495124
4417	OAB+	20.56339073	21.33322716	0.963913738
4462	OAB+	20.34226418	19.8664856	1.023948805
4494	OAB+	18.67056465	21.10313606	0.884729388
4496	OAB+	24.96712303	20.50193024	1.217793775
4511	OAB+	21.22506905	22.35007477	0.949664342
4539	noLUTS	20.35417557	21.7481308	0.935904596
4567	OAB+	15.96492386	21.22679138	0.752111969
4578	OAB+	23.98238182	14.71793747	1.629466212
4594	OAB+	17.70431519	21.59204865	0.819946059
4639	OAB+	18.48090744	21.29706764	0.867767702
4645	noLUTS	22.5163517	21.31454277	1.056384457
4646	noLUTS	29.85168648	22.2653904	1.340721449
4647	noLUTS	16.9053421	22.2843399	0.758619828
4668	noLUTS	23.50241661	21.23691368	1.106677598
4692	OAB+	28.16348839	21.85209084	1.288823509
4742	noLUTS	16.86803627	21.754076	0.775396586

4814	noLUTS	36.96261597	21.68200684	1.704759907
4815	noLUTS	23.34731865	21.54071617	1.083869193
4821	noLUTS	23.7624588	21.82576942	1.088734071
4851	noLUTS	20.2883358	21.45603371	0.945577178
4852	OAB+	23.32775307	21.67765808	1.076119615
4979	noLUTS	26.00753212	19.25813293	1.35047007
5461	noLUTS	18.56741524	15.03671265	1.234805484
5468	OAB+	24.90854454	21.42233849	1.162736951
5520	OAB+	26.84179497	13.95525742	1.923418118
6115	OAB+	19.31811905	19.06550407	1.013249845
6162	OAB+	14.79330826	16.46980095	0.898208078
6206	OAB+	19.93521881	18.11376381	1.100556407
6403	OAB+	17.6954937	15.13295364	1.169335089
6503	OAB+	19.46761131	21.53987885	0.903793909
6517	OAB+	23.86850929	22.19278717	1.075507511
6578	OAB+	17.97973633	22.40927315	0.80233465
6635	OAB+	21.54712868	21.92750549	0.982652983
6901	OAB+	35.27679062	22.98445129	1.534811085
6933	OAB+	18.63139153	22.05608177	0.844728076
7224	OAB+	19.78726768	22.83168983	0.866658045
7250	OAB+	26.11292458	17.73113441	1.472715956
7264	OAB+	23.3297348	14.03883362	1.661800078
7519	UTI+	22.38907623	21.94596863	1.020190843
7531	UTI+	15.95742416	21.02646828	0.758920802
7672	UTI+	19.45863152	15.41108799	1.262638402
7674	UTI+	15.09441853	21.57970619	0.699472847
7676	UTI+	18.27129936	12.9889183	1.406683677
7679	UTI+	18.46508598	21.78534889	0.847591933
7713	UTI+	20.68468857	21.92334557	0.943500549
7714	UTI+	26.27718925	16.98213959	1.547342673
7772	UTI+	18.67436981	22.04235077	0.847204094
7785	UTI+	21.0113678	20.53281975	1.023306494
7786	UTI+	16.79366112	22.20022011	0.756463721

APPENDIX F

AGE-RACE/ETHNICITY PAIRS

				JCPyV+						JCPyV-		
Pair ID	Patient ID	Sample ID	Symptom	Race/Ethnicity	Age	Bacterial Abundance	Patient ID	Sample ID	Symptom	Race/Ethnicity	Age	Bacterial Abundance
1	R01FUM_38	4304	noOAB/noLUTS	Black/African American	54	20.80507469	R01FUM_83	4850	noOAB/noLUTS	Black/African American	48	21.89781761
2	R01FUM_89	4979	noOAB/noLUTS	Black/African American	57	19.25813293	R01FUM_57	4600	noOAB/noLUTS	Black/African American	52	21.5015316
3	R01FUM_53	4539	noOAB/noLUTS	Hispanic/ Latina	48	21.7481308	R01FUM_61	4662	noOAB/noLUTS	Hispanic/ Latina	48	21.36637115
4	R01FUM_58	4645	noOAB/noLUTS	Hispanic/ Latina	57	21.31454277	R01FUM_2	4575	noOAB/noLUTS	Hispanic/ Latina	49	21.01452446
5	R01FUM_76	4815	noOAB/noLUTS	White/Caucasian	35	21.54071617	R01FUM_3	4576	noOAB/noLUTS	White/Caucasian	34	21.43721771
6	R01FUM_49	4388	noOAB/noLUTS	White/Caucasian	38	23.05531502	R01FUM_90	5027	noOAB/noLUTS	White/Caucasian	40	22.01726723
7	R01FUM_60	4647	noOAB/noLUTS	White/Caucasian	41	22.2843399	R01FUM_72	4796	noOAB/noLUTS	White/Caucasian	42	20.74575424
8	R01FUM_63	4668	noOAB/noLUTS	White/Caucasian	48	21.23691368	R01FUM_85	4855	noOAB/noLUTS	White/Caucasian	47	20.63358879
9	R01FUM_59	4646	noOAB/noLUTS	White/Caucasian	50	22.2653904	R01FUM_18	4053	noOAB/noLUTS	White/Caucasian	50	21.52291298

10	R01FUM_65	4742	noOAB/noLUTS	White/Caucasian	50	21.75407606	R01FUM_71	4795	noOAB/noLUTS	White/Caucasian	51	20.09239006
11	R01FUM_75	4814	noOAB/noLUTS	White/Caucasian	54	21.68200684	R01FUM_87	4877	noOAB/noLUTS	White/Caucasian	54	21.41673279
12	R01FUM_100	5461	noOAB/noLUTS	White/Caucasian	65	15.03671265	R21_14	3818	noOAB/noLUTS	White/Caucasian	63	23.93748474
13	R01FUM_80	4821	noOAB/noLUTS	White/Caucasian	67	21.82576942	R01FUM_79	4820	noOAB/noLUTS	White/Caucasian	62	22.31437874
14	R01FUM_84	4851	noOAB/noLUTS	White/Caucasian	83	21.45603371	R01FUM_98	5424	noOAB/noLUTS	White/Caucasian	71	21.48872566
15	MIR_39	4496	OAB+	Black/African American	48	20.50193024	MIR_52	4861	OAB+	Black/African American	55	22.60758018
16	EST_21	3127	OAB+	Black/African American	54	26.12587738	MIR_81	7230	OAB+	Black/African American	54	20.87561989
17	MIR_16	3022	OAB+	Black/African American	59	22.01903725	MIR_71	6513	OAB+	Black/African American	59	21.5242939
18	EST_34	3791	OAB+	Black/African American	67	21.12228394	MIR_65	6212	OAB+	Black/African American	67	21.86877632
19	MIR_80	7224	OAB+	Black/African American	68	22.83168983	MIR_50	4817	OAB+	Black/African American	65	21.31832695

20	MIR_4	2720	OAB+	Black/African American	79	25.97675514	EST_52	4532	OAB+	Black/African American	70	20.94066429
21	MIR_20	3241	OAB+	Hispanic/Latina	46	21.21721268	MIR_7	2785	OAB+	Hispanic/Latina	57	25.06987953
22	MIR_61	6115	OAB+	Hispanic/Latina	47	19.06550407	MIR_37	4326	OAB+	Hispanic/Latina	58	20.88613129
23	MIR_13	3000	OAB+	Hispanic/Latina	69	27.69773293	MIR_45	4660	OAB+	Hispanic/Latina	68	21.54701424
24	MIR_34	4021	OAB+	White/Caucasian	49	20.76813316	MIR_22	3263	OAB+	White/Caucasian	48	21.07706261
25	MIR_41	4462	OAB+	White/Caucasian	50	19.8664856	MIR_24	3349	OAB+	White/Caucasian	50	21.72794151
26	MIR_18	3131	OAB+	White/Caucasian	53	21.52444458	EST_50	4518	OAB+	White/Caucasian	54	21.30768776
27	MIR_6	2769	OAB+	White/Caucasian	54	20.75850868	MIR_68	6426	OAB+	White/Caucasian	55	21.37845802
28	MIR_56	5468	OAB+	White/Caucasian	58	21.42233849	MIR_35	4275	OAB+	White/Caucasian	58	16.40048409
29	MIR_36	4214	OAB+	White/Caucasian	63	20.81211853	MIR_51	4848	OAB+	White/Caucasian	63	21.47080231
30	EST_27	3363	OAB+	White/Caucasian	72	22.32190132	EST_23	3280	OAB+	White/Caucasian	72	22.00371742
31	MIR_42	4692	OAB+	White/Caucasian	72	21.85209084	MIR_26	3458	OAB+	White/Caucasian	71	21.49610901
32	EST_19	2981	OAB+	White/Caucasian	73	27.18789673	MIR_10	2933	OAB+	White/Caucasian	73	25.72686958
33	EST_5	2662	OAB+	White/Caucasian	77	25.86221313	MIR_43	4631	OAB+	White/Caucasian	77	21.56041527
34	MIR_15	3002	OAB+	White/Caucasian	77	20.6443882	MIR_49	4773	OAB+	White/Caucasian	77	16.52857208

35	MIR_69	6503	OAB+	White/Caucasian	78	21.53987885	MIR_70	6504	OAB+	White/Caucasian	79	21.55736351
36	MIR_40	4417	OAB+	White/Caucasian	81	21.33322716	MIR_82	7249	OAB+	White/Caucasian	80	20.89541245
37	EST_4	2656	OAB+	White/Caucasian	82	27.88180542	MIR_8	2864	OAB+	White/Caucasian	80	27.92691994
38	MIR_58	5520	OAB+	White/Caucasian	85	13.95525742	MIR_46	4746	OAB+	White/Caucasian	86	21.64009094
39	MIR_73	6517	OAB+	White/Caucasian	86	22.19278717	MIR_47	4760	OAB+	White/Caucasian	86	21.5287838
40	SvE_204	7672	UTI+	Black/African American	73	15.41108799	SvE_212	7716	UTI+	Black/African American	76	22.64885902
41	SvE_220	7786	UTI+	Hispanic/Latina	32	22.20022011	SvE_224	7805	UTI+	Hispanic/Latina	39	22.01782799
42	SvE_219	7785	UTI+	Hispanic/Latina	54	20.53281975	SvE_199	7653	UTI+	Hispanic/Latina	66	22.93267822
43	SvE_210	7714	UTI+	White/Caucasian	64	16.98213959	SvE_203	7667	UTI+	White/Caucasian	62	22.87467384
44	SvE_206	7676	UTI+	White/Caucasian	74	12.9889183	SvE_201	7659	UTI+	White/Caucasian	74	22.25759315
45	SvE_170	7519	UTI+	White/Caucasian	75	21.94596863	SvE_192	7600	UTI+	White/Caucasian	72	22.06970406
46	SvE_207	7679	UTI+	White/Caucasian	75	21.78534889	SvE_223	7792	UTI+	White/Caucasian	68	21.77348137
47	SvE_172	7531	UTI+	White/Caucasian	76	21.02646828	SvE_171	7528	UTI+	White/Caucasian	78	23.26261711

REFERENCE LIST

1. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The Human Microbiome Project. 7164. Nature 449:804–810.

2. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 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 52:871–876.

3. Thomas-White K, Brady M, Wolfe AJ, Mueller ER. 2016. The Bladder Is Not Sterile: History and Current Discoveries on the Urinary Microbiome. Curr Bladder Dysfunct Rep 11:18–24.

4. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, Fitzgerald M, Mueller ER, Schreckenberger P, Dong Q, Nelson DE, Brubaker L. 2012. Evidence of uncultivated bacteria in the adult female bladder. J Clin Microbiol 50:1376–1383.

5. Fouts DE, Pieper R, Szpakowski S, Pohl H, Knoblach S, Suh M-J, Huang S-T, Ljungberg I, Sprague BM, Lucas SK, Torralba M, Nelson KE, Groah SL. 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 10:174.

6. Lewis D, Brown R, Williams J, White P, Jacobson S, Marchesi J, Drake M. 2013. The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. Frontiers in Cellular and Infection Microbiology 3.

7. Proctor LM. 2011. The Human Microbiome Project in 2011 and Beyond. Cell Host & Microbe 10:287–291.

8. Marrazzo JM. 2011. Interpreting the epidemiology and natural history of bacterial vaginosis: Are we still confused? Anaerobe 17:186–190.

9. Jones J, Murphy CP, Sleator RD, Culligan EP. 2021. The urobiome, urinary tract infections, and the need for alternative therapeutics. Microbial Pathogenesis 161:105295.

10. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014. The Female Urinary Microbiome:

a Comparison of Women with and without Urgency Urinary Incontinence. mBio 5:e01283-14.

11. Ackerman AL, Underhill DM. 2017. The mycobiome of the human urinary tract: potential roles for fungi in urology. 2. Annals of Translational Medicine 5:5–5.

12. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspõllu A, Väin E, Saarma I, Salumets A, Donders GGG, Metsis M. 2013. Characterization of the Vaginal Micro- and Mycobiome in Asymptomatic Reproductive-Age Estonian Women. PLOS ONE 8:e54379.

13. Grine G, Lotte R, Chirio D, Chevalier A, Raoult D, Drancourt M, Ruimy R. 2019. Co-culture of Methanobrevibacter smithii with enterobacteria during urinary infection. eBioMedicine 43:333–337.

14. Miller-Ensminger T, Garretto A, Brenner J, Thomas-White K, Zambom A, Wolfe AJ, Putonti C. 2018. Bacteriophages of the Urinary Microbiome. J Bacteriol 200.

15. Garretto A, Thomas-White K, Wolfe AJ, Putonti C. 2018. Detecting viral genomes in the female urinary microbiome. J Gen Virol 99:1141–1146.

16. Salabura A, Łuniewski A, Kucharska M, Myszak D, Dołęgowska B, Ciechanowski K, Kędzierska-Kapuza K, Wojciuk B. 2021. Urinary Tract Virome as an Urgent Target for Metagenomics. 11. Life 11:1264.

17. Echavarria M, Forman M, Ticehurst J, Dumler JS, Charache P. 1998. PCR Method for Detection of Adenovirus in Urine of Healthy and Human Immunodeficiency Virus-Infected Individuals. Journal of Clinical Microbiology 36:3323–3326.

18. Lion T. 2014. Adenovirus Infections in Immunocompetent and Immunocompromised Patients. Clinical Microbiology Reviews 27:441–462.

19. Tan SK, Relman DA, Pinsky BA. 2017. The Human Virome: Implications for Clinical Practice in Transplantation Medicine. Journal of Clinical Microbiology 55:2884–2893.

20. Pathak N, Dodds J, Zamora J, Khan K. 2014. Accuracy of urinary human papillomavirus testing for presence of cervical HPV: systematic review and meta-analysis. BMJ 349:g5264.

21. Chang H, Wang M, Tsai R-T, Lin H-S, Huan J-S, Wang W-C, Chang D. 2002. High incidence of JC viruria in JC-seropositive older individuals. Journal of NeuroVirology 8:447–451.

22. Payne S. 2017. Chapter 31 - Family Polyomaviridae, p. 247–251. *In* Payne, S (ed.), Viruses. Academic Press.

23. Pinto M, Dobson S. 2014. BK and JC virus: a review. J Infect 68 Suppl 1:S2-8.

24. Shackelton LA, Rambaut A, Pybus OG, Holmes EC. 2006. JC Virus Evolution and Its Association with Human Populations. J Virol 80:9928–9933.

25. Bofill-Mas S, Girones R. 2003. Role of the environment in the transmission of JC virus. J Neurovirol 9 Suppl 1:54–58.

26. Kitamura T, Kunitake T, Guo J, Tominaga T, Kawabe K, Yogo Y. 1994. Transmission of the human polyomavirus JC virus occurs both within the family and outside the family. J Clin Microbiol 32:2359–2363.

27. Agostini HT, Yanagihara R, Davis V, Ryschkewitsch CF, Stoner GL. 1997. Asian genotypes of JC virus in Native Americans and in a Pacific Island population: Markers of viral evolution and human migration. Proc Natl Acad Sci U S A 94:14542–14546.

28. Bofill-Mas S, Girones R. 2001. Excretion and transmission of JCV in human populations. Journal of NeuroVirology 7:345–349.

29. Knowles WA. 2006. Discovery and Epidemiology of the Human Polyomaviruses BK Virus (BKV) and JC Virus (JCV), p. 19–45. *In* Ahsan, N (ed.), Polyomaviruses and Human Diseases. Springer, New York, NY.

30. Kean JM, Rao S, Wang M, Garcea RL. 2009. Seroepidemiology of Human Polyomaviruses. PLOS Pathogens 5:e1000363.

31. Bennett SM, Broekema NM, Imperiale MJ. 2012. BK polyomavirus: emerging pathogen. Microbes and Infection 14:672–683.

32. Atkinson AL, Atwood WJ. 2020. Fifty Years of JC Polyomavirus: A Brief Overview and Remaining Questions. 9. Viruses 12:969.

33. Cubie HA, Cuschieri KS, Tong CYW. 2012. Papillomaviruses and polyomaviruses, p. 452–463. *In* Medical Microbiology. Elsevier.

34. Safak M. 2018. Polyomaviruses of Humans, p. B9780128012383027000. *In* Reference Module in Biomedical Sciences. Elsevier.

35. Barbanti-Brodano G, Sabbioni S, Martini F, Negrini M, Corallini A, Tognon M. 2013. BK Virus, JC Virus and Simian Virus 40 Infection in Humans, and Association with Human TumorsMadame Curie Bioscience Database [Internet]. Landes Bioscience. https://www.ncbi.nlm.nih.gov/books/NBK6100/. Retrieved 31 May 2022.

36. Saribas AS, Coric P, Bouaziz S, Safak M. 2019. Expression of novel proteins by polyomaviruses and recent advances in the structural and functional features of agnoprotein of JC virus, BK virus, and simian virus 40. Journal Cellular Physiology 234:8295–8315.

37. Nakamichi K, Kawamoto M, Ishii J, Saijo M. 2019. Improving detection of JC virus by ultrafiltration of cerebrospinal fluid before polymerase chain reaction for the diagnosis of progressive multifocal leukoencephalopathy. BMC Neurology 19:252.

38. Tan CS, Ellis LC, Wüthrich C, Ngo L, Broge TA, Saint-Aubyn J, Miller JS, Koralnik IJ. 2010. JC Virus Latency in the Brain and Extraneural Organs of Patients with and without Progressive Multifocal Leukoencephalopathy. Journal of Virology 84:9200–9209.

39. Tan CS, Dezube BJ, Bhargava P, Autissier P, Wuthrich C, Miller J, Koralnik IJ. 2009. Detection of JC virus DNA and proteins in bone marrow of HIV-positive and HIV-negative patients. J Infect Dis 199:881–888.

40. Monaco MCG, Jensen PN, Hou J, Durham LC, Major EO. 1998. Detection of JC Virus DNA in Human Tonsil Tissue: Evidence for Site of Initial Viral Infection. J Virol 72:9918–9923.

41. Tan CS, Koralnik IJ. 2010. Progressive multifocal leukoencephalopathy and other disorders caused by JC virus: clinical features and pathogenesis. Lancet Neurol 9:425–437.

42. Weissert R. 2011. Progressive multifocal leukoencephalopathy. Journal of Neuroimmunology 231:73–77.

43. O'Neill FJ, Greenlee JE, Dörries K, Clawson SA, Carney H. 2003. Propagation of archetype and nonarchetype JC virus variants in human fetal brain cultures: Demonstration of interference activity by archetype JC virus. Journal of NeuroVirology 9:567–576.

44. Padgett B. 1983. BK VIRUS AND NON-HAEMORRHAGIC CYSTITIS IN A CHILD. The Lancet 321:770.

45. Saitoh K, Sugae N, Koike N, Akiyama Y, Iwamura Y, Kimura H. 1993. Diagnosis of childhood BK virus cystitis by electron microscopy and PCR. Journal of Clinical Pathology 46:773–775.

46. Thomas S, Dunn CD, Campbell LJ, Strand DW, Vezina CM, Bjorling DE, Penniston KL, Li L, Ricke WA, Goldberg TL. 2021. A multi-omic investigation of male lower urinary tract symptoms: Potential role for JC virus. PLOS ONE 16:e0246266.

47. Sanjuán R, Domingo-Calap P. 2021. Genetic Diversity and Evolution of Viral Populations, p. 53–61. *In* Bamford, DH, Zuckerman, M (eds.), Encyclopedia of Virology (Fourth Edition). Academic Press, Oxford.

48. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AMQ, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz EJ, Nibert ML, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, van der Vlugt RA, Varsani A, Zerbini FM. 2017. Virus taxonomy in the age of metagenomics. 3. Nat Rev Microbiol 15:161–168.

49. Santiago-Rodriguez TM, Ly M, Bonilla N, Pride DT. 2015. The human urine virome in association with urinary tract infections. Frontiers in Microbiology 6.

50. Sigdel TK, Mercer N, Nandoe S, Nicora CD, Burnum-Johnson K, Qian W-J, Sarwal MM. 2018. Urinary Virome Perturbations in Kidney Transplantation. Frontiers in Medicine 5.

51. Rani A, Ranjan R, McGee HS, Metwally A, Hajjiri Z, Brennan DC, Finn PW, Perkins DL. 2016. A diverse virome in kidney transplant patients contains multiple viral subtypes with distinct polymorphisms. 1. Sci Rep 6:33327.

52. Moustafa A, Li W, Singh H, Moncera KJ, Torralba MG, Yu Y, Manuel O, Biggs W, Venter JC, Nelson KE, Pieper R, Telenti A. 2018. Microbial metagenome of urinary tract infection. 1. Sci Rep 8:4333.

53. Carpenter ML, Tan SK, Watson T, Bacher R, Nagesh V, Watts A, Bentley G, Weber J, Huang C, Sahoo MK, Hinterwirth A, Doan T, Carter T, Dong Q, Gourguechon S, Harness E, Kermes S, Radhakrishnan S, Wang G, Quiroz-Zárate A, Ching J, Pinsky BA. 2019. Metagenomic Next-Generation Sequencing for Identification and Quantitation of Transplant-Related DNA Viruses. Journal of Clinical Microbiology 57.

54. Adiliaghdam F, Amatullah H, Digumarthi S, Saunders TL, Rahman R-U, Wong LP, Sadreyev R, Droit L, Paquette J, Goyette P, Rioux JD, Hodin R, Mihindukulasuriya KA, Handley SA, Jeffrey KL. 2022. Human enteric viruses autonomously shape inflammatory bowel disease phenotype through divergent innate immunomodulation. Science Immunology 7:eabn6660.

55. Neugent ML, Hulyalkar NV, Nguyen VH, Zimmern PE, De Nisco NJ. 2020. Advances in Understanding the Human Urinary Microbiome and Its Potential Role in Urinary Tract Infection. mBio 11:e00218-20.

56. Gardner SylviaD, Field AnneM, Coleman DulcieV, Hulme B. 1971. NEW HUMAN PAPOVAVIRUS (B.K.) ISOLATED FROM URINE AFTER RENAL TRANSPLANTATION. The Lancet 297:1253–1257.

57. Boldorini R, Allegrini S, Miglio U, Paganotti A, Cocca N, Zaffaroni M, Riboni F, Monga G, Viscidi R. 2011. Serological evidence of vertical transmission of JC and BK polyomaviruses in humans. J Gen Virol 92:1044–1050.

58. Bofill-Mas S, Formiga-Cruz M, Clemente-Casares P, Calafell F, Girones R. 2001. Potential Transmission of Human Polyomaviruses through the Gastrointestinal Tract after Exposure to Virions or Viral DNA. J Virol 75:10290–10299.

59. Cui X, Wang JC, Deckhut A, Joseph BC, Eberwein P, Cubitt CL, Ryschkewitsch CF, Agostini HT, Stoner GL. 2004. Chinese Strains (Type 7) of JC Virus Are Afro-Asiatic in Origin But Are Phylogenetically Distinct from the Mongolian and Indian Strains (Type 2D) and the Korean and Japanese Strains (Type 2A). J Mol Evol 58:568–583.

60. Pavesi A. 2005. Utility of JC polyomavirus in tracing the pattern of human migrations dating to prehistoric times. Journal of General Virology 86:1315–1326.

61. Agostini HT, Deckhut A, Jobes DV, Girones R, Schlunck G, Prost MG, Frias C, Pérez-Trallero E, Ryschkewitsch CF, Stoner GLY 2001. Genotypes of JC virus in East, Central and Southwest Europe. Journal of General Virology 82:1221–1331.

62. Baksh FK, Finkelstein SD, Swalsky PA, Stoner GL, Ryschkewitsch CF, Randhawa P. 2001. Molecular genotyping of BK and JC viruses in human polyomavirus[ndash]associated interstitial nephritis after renal transplantation. American Journal of Kidney Diseases 38:354–365.

63. Bogdanovic G, Brytting M, Cinque P, Grandien M, Fridell E, Ljungman P, Lönnqvist B, Hammarin A-L. 1994. Nested PCR for detection of BK virus and JC virus DNA. Clinical and Diagnostic Virology 2:211–220.

64. Cayres-Vallinoto IMV, Vallinoto ACR, Azevedo VN, Machado LFA, Ishak M de OG, Ishak R. 2012. Human JCV Infections as a Bio-Anthropological Marker of the Formation of Brazilian Amazonian Populations. PLOS ONE 7:e46523.

65. Souza LM de, Savassi-Ribas F, Almeida SGS de, Silva RNN da, Baez CF, Zalis MG, Guimarães MAAM, Varella RB. 2018. A globally applicable PCR-based detection and discrimination of BK and JC polyomaviruses. Rev Inst Med trop S Paulo 60.

66. Del Valle L, Gordon J, Assimakopoulou M, Enam S, Geddes JF, Varakis JN, Katsetos CD, Croul S, Khalili K. 2001. Detection of JC virus DNA sequences and expression of the viral regulatory protein T-antigen in tumors of the central nervous system. Cancer Res 61:4287–4293.

67. Dumonceaux TJ, Mesa C, Severini A. 2008. Internally Controlled Triplex Quantitative PCR Assay for Human Polyomaviruses JC and BK. Journal of Clinical Microbiology 46:2829–2836.

68. Dumoulin A, Hirsch HH. 2011. Reevaluating and Optimizing Polyomavirus BK and JC Real-Time PCR Assays To Detect Rare Sequence Polymorphisms. Journal of Clinical Microbiology 49:1382–1388.

69. Funahashi Y, Iwata S, Ito Y, Kojima S, Yoshikawa T, Hattori R, Gotoh M, Nishiyama Y, Kimura H. 2010. Multiplex Real-Time PCR Assay for Simultaneous Quantification of BK Polyomavirus, JC Polyomavirus, and Adenovirus DNA. Journal of Clinical Microbiology 48:825–830.

70. Giovannelli I, Ciccone N, Vaggelli G, Malva ND, Torricelli F, Rossolini GM, Giannecchini S. 2016. Utility of droplet digital PCR for the quantitative detection of polyomavirus JC in clinical samples. Journal of Clinical Virology 82:70–75.

71. Haghighi MF, Seyyedi N, Farhadi A, Zare F, Kasraian L, Refiei Dehbidi GR, Ranjbaran R, Behzad-Behbahani A. 2019. Polyomaviruses BK and JC DNA infection in peripheral blood cells from blood donors. The Brazilian Journal of Infectious Diseases 23:22–26.

72. Hori R, Murai Y, Tsuneyama K, Abdel-Aziz HO, Nomoto K, Takahashi H, Cheng C, Kuchina T, Harman BV, Takano Y. 2005. Detection of JC virus DNA sequences in colorectal cancers in Japan. Virchows Arch 447:723–730.

73. Hussain I, Tasneem F, Umer M, Pervaiz A, Raza M, Arshad MI, Shahzad N. 2017. Specific and quantitative detection of Human polyomaviruses BKPyV and JCPyV in the healthy Pakistani population. Virology Journal 14:86.

74. Karalic D, Lazarevic I, Knezevic A, Cupic M, Jevtovic D, Jovanovic T. 2014. Distribution of JC virus genotypes among serbian patients infected with HIV and in healthy donors. Journal of Medical Virology 86:411–418.

75. Kruzel-Davila E, Divers J, Russell GB, Kra-Oz Z, Cohen MS, Langefeld CD, Ma L, Lyles DS, Hicks PJ, Skorecki KL, Freedman BI. 2019. JC Viruria Is Associated With Reduced Risk of Diabetic Kidney Disease. None 104:2286–2294.

76. Makvandi M, Mombeini H, Haghighi SB, Dastoorpoor M, Khodadad N, Babaahmadi MK, Tabasi M, Pirmoradi R. 2020. Molecular epidemiology of JC polyomavirus in HIV-infected patients and healthy individuals from Iran. Braz J Microbiol 51:37–43.

77. McNees AL, White ZS, Zanwar P, Vilchez RA, Butel JS. 2005. Specific and quantitative detection of human polyomaviruses BKV, JCV, and SV40 by real time PCR. Journal of Clinical Virology 34:52–62.

78. Merlino C, Bergallo M, Daniele R, Ponzi AN, Cavallo R. 2005. BKV-DNA and JCV-DNA Co-quantification assay to evaluate viral load in urine and serum. Mol Biotechnol 30:1–8.

79. Pal A, Sirota L, Maudru T, Peden K, Lewis AM. 2006. Real-time, quantitative PCR assays for the detection of virus-specific DNA in samples with mixed populations of polyomaviruses. Journal of Virological Methods 135:32–42.

80. Rafique A, Jiang SC. 2008. Genetic diversity of human polyomavirus JCPyV in Southern California wastewater. Journal of Water and Health 6:533–538.

81. Rencic A, Gordon J, Otte J, Curtis M, Kovatich A, Zoltick P, Khalili K, Andrews D. 1996. Detection of JC virus DNA sequence and expression of the viral oncoprotein, tumor antigen, in brain of immunocompetent patient with oligoastrocytoma. Proceedings of the National Academy of Sciences 93:7352–7357.

82. Ryschkewitsch C, Jensen P, Hou J, Fahle G, Fischer S, Major EO. 2004. Comparison of PCR-southern hybridization and quantitative real-time PCR for the detection of JC and BK

viral nucleotide sequences in urine and cerebrospinal fluid. Journal of Virological Methods 121:217–221.

83. Sehbani L, Kabamba-Mukadi B, Vandenbroucke A-T, Bodéus M, Goubau P. 2006. Specific and quantitative detection of human polyomaviruses BKV and JCV by LightCycler® real-time PCR. Journal of Clinical Virology 36:159–162.

84. Tsai R-T, Wang M, Ou W-C, Lee Y-L, Li S-Y, Fung C-Y, Huang Y-L, Tzeng T-Y, Chen Y, Chang D. 1997. Incidence of JC viruria is higher than that of BK viruria in Taiwan. Journal of Medical Virology 52:253–257.

85. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

86. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539.

87. Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol Biol Evol 26:1641–1650.

88. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128.

89. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134.

90. Hu C, Huang Y, Su J, Wang M, Zhou Q, Zhu B. 2018. Detection and analysis of variants of JC polyomavirus in urine samples from HIV-1-infected patients in China's Zhejiang Province. J Int Med Res 46:1024–1032.

91. Urbano PRP, Oliveira RR, Romano CM, Pannuti CS, Fink MCD da S. 2016. Occurrence, genotypic characterization, and patterns of shedding of human polyomavirus JCPyV and BKPyV in urine samples of healthy individuals in São Paulo, Brazil: JCPyV and BKPyV Shedding in Healthy Individuals. J Med Virol 88:153–158.

92. Nali LH da S, Centrone C de C, Urbano PRP, Penalva-de-Oliveira AC, Vidal JE, Miranda EP, Pannuti CS, Fink MCD da S. 2012. High prevalence of the simultaneous excretion of polyomaviruses JC and BK in the urine of HIV-infected patients without neurological symptoms in São Paulo, Brazil. Rev Inst Med trop S Paulo 54:201–205.

93. Elfaitouri A, Hammarin A-L, Blomberg J. 2006. Quantitative real-time PCR assay for detection of human polyomavirus infection. Journal of Virological Methods 135:207–213.

94. Mou X, Chen L, Liu F, Lin J, Diao P, Wang H, Li Y, Lin J, Teng L, Xiang C. 2012. Prevalence of JC Virus in Chinese Patients with Colorectal Cancer. PLoS ONE 7:e35900.

95. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, Hilt EE, Price TK, Wolfe AJ, Lawley TD. 2018. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. 1. Nat Commun 9:1557.

96. Wommack KE, Bhavsar J, Polson SW, Chen J, Dumas M, Srinivasiah S, Furman M, Jamindar S, Nasko DJ. 2012. VIROME: a standard operating procedure for analysis of viral metagenome sequences. Stand Genomic Sci 6:427–439.

97. Huson DH, Weber N. 2013. Microbial Community Analysis Using MEGAN, p. 465–485. *In* Methods in Enzymology. Elsevier.

98. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985.

99. Roux S, Tournayre J, Mahul A, Debroas D, Enault F. 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC Bioinformatics 15:76.

100. Keegan KP, Glass EM, Meyer F. 2016. MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function, p. 207–233. *In* Martin, F, Uroz, S (eds.), Microbial Environmental Genomics (MEG). Springer New York, New York, NY.

101. Broekema NM, Imperiale MJ. 2012. Efficient Propagation of Archetype BK and JC Polyomaviruses. Virology 422:235–241.

102. Chuang L-Y, Cheng Y-H, Yang C-H. 2013. Specific primer design for the polymerase chain reaction. Biotechnol Lett 35:1541–1549.

103. Lu M-C, Yin W-Y, Liu S-Q, Koo M, Tung C-H, Huang K-Y, Lai N-S. 2015. Increased prevalence of JC polyomavirus viruria was associated with arthritis/arthralgia in patients with systemic lupus erythematosus. Lupus 24:687–694.

104. Yazdani Cherati A, Yahyapour Y, Ranaee M, Rajabnia M, Shokri Shirvani J, Hajiahmadi M, Sadeghi F. 2018. No Evidence for an Association between JC Polyomavirus Infection and Gastroduodenal Diseases. Gastrointest Tumors 5:47–53.

105. Delbue S, Comar M, Ferrante P. 2017. Review on the role of the human Polyomavirus JC in the development of tumors. Infect Agents Cancer 12:10.

106. Comar M, Zanotta N, Croci E, Murru I, Marci R, Pancaldi C, Dolcet O, Luppi S, Martinelli M, Giolo E, Ricci G, Tognon M. 2012. Association between the JC Polyomavirus Infection and Male Infertility. PLoS ONE 7:e42880.

107. Thomas-White K, Taege S, Limeira R, Brincat C, Joyce C, Hilt EE, Mac-Daniel L, Radek KA, Brubaker L, Mueller ER, Wolfe AJ. 2020. Vaginal estrogen therapy is associated with increased Lactobacillus in the urine of postmenopausal women with overactive bladder symptoms. American Journal of Obstetrics and Gynecology 223:727.e1-727.e11.

108. Price TK, Lin H, Gao X, Thomas-White KJ, Hilt EE, Mueller ER, Wolfe AJ, Dong Q, Brubaker L. 2020. Bladder bacterial diversity differs in continent and incontinent women: a cross-sectional study. American Journal of Obstetrics and Gynecology 223:729.e1-729.e10.

109. Price T, Hilt E, Thomas-White K, Mueller E, Wolfe A, Brubaker L. 2020. The urobiome of continent adult women: a cross-sectional study. BJOG: Int J Obstet Gy 127:193–201.

110. Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Wade WG. 1998. Design and Evaluation of Useful Bacterium-Specific PCR Primers That Amplify Genes Coding for Bacterial 16S rRNA. Applied and Environmental Microbiology 64:795–799.

111. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Research 40:e115.

112. Clifford RJ, Milillo M, Prestwood J, Quintero R, Zurawski DV, Kwak YI, Waterman PE, Lesho EP, Gann PM. 2012. Detection of Bacterial 16S rRNA and Identification of Four Clinically Important Bacteria by Real-Time PCR. PLOS ONE 7:e48558.

113. RStudio Team. RStudio: Integrated Development for R. 2015. RStudio, Inc., Boston, MA.

114. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis2nd ed. 2016. Springer International Publishing : Imprint: Springer, Cham.

115. Venables WN, Ripley BD, Venables WN. 2002. Modern applied statistics with S4th ed. Springer, New York.

116. Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723.

117. Bryer J, Speerschneider K. 2016. likert: Analysis and Visualization Likert Items (1.3.5). https://CRAN.R-project.org/package=likert. Retrieved 31 May 2022.

118. Gomez-Alvarez V, Teal TK, Schmidt TM. 2009. Systematic artifacts in metagenomes from complex microbial communities. 11. ISME J 3:1314–1317.

119. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

120. Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485.

121. Bozic C, Subramanyam M, Richman S, Plavina T, Zhang A, Ticho B. 2014. Anti-JC virus (JCV) antibody prevalence in the JCV Epidemiology in MS (JEMS) trial. Eur J Neurol 21:299–304.

VITA

Ms. Mormando was raised in South Barrington, Illinois. She attended Loyola University Chicago for her undergraduate degree, where she obtained a Bachelor of Science degree in Bioinformatics. Ms. Mormando worked as a bioinformatics research assistant in the Putonti Lab for four years focusing on bacteriophage, bacterial, and viral analyses.