Fall 2022

Exploring the Role of Lactobacillus Jensenii and Lactobacillus Mulieris in the Urogenital Tract

Adriana Daniela Ene

LOYOLA UNIVERSITY CHICAGO

EXPLORING THE ROLE OF LACTOBACILLUS JENSENII AND LACTOBACILLUS MULIERIS IN THE UROGENITAL TRACT

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN BIOINFORMATICS

BY

ADRIANA ENE

CHICAGO, IL

AUGUST 2022

ACKNOWLEDGEMENTS

I would like to thank Dr. Catherine Putonti for the guidance and mentorship she has given me throughout the years and for always supporting my research endeavors. I want to thank members of the Putonti lab for their support and assistance. I would like to also thank the members of my committee, Dr. Stefan Kanzok and Dr. Michael Burns, for lending their knowledge to my project. Finally, I would like to thank my family for their support through my academic career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Lactobacillus is a predominant species of the urogenital tract and it has been found in females with and without lower urinary tract symptoms. The Lactobacillus genus is thought to be beneficial to the health of the urogenital tract due to its capability to produce hydrogen peroxide, secondary metabolites, and lactic acid. L. jensenii is an organism frequently isolated from the urogenital tract. In March 2020, a new sister taxon of L. jensenii was described, L. mulieris. Shortly thereafter we produced a genomic analysis of all publicly available genomes (n = 43) which reclassified some L. jensenii strains as L. mulieris. This motivated the in-depth study of the difference between L. jensenii and L. mulieris presented here which included expanding our collection of representatives of these two species to 61 strains. Genome analysis was conducted including the examination of the core genome as well as secondary metabolites, prophages and virulence factors. Complementing this genetic analysis, urinary strains of both species were phenotyped for urinary tract relevant characteristics, including sugar metabolism, pH, and hydrogen peroxide production. Lastly, in an effort to ascertain the prevalence of these two species in the urinary tract, we found that the 16S rRNA gene sequence was insufficient, prompting our design of new gene markers that could specifically detect L. jensenii and L. mulieris while also being able to distinguish between the two. Using one of these gene markers, I assayed 190 urine samples and found that none of them contained either species. This led me to examine 233 urinary metagenomes for evidence of these gene markers, finding L. jensenii in

only six samples and L. mulieris in two samples. These results suggest that L. jensenii and L. mulieris are not as abundant in the urinary microbiota as previously thought.

CHAPTER 1

INTRODUCTION

Human Microbiome

The human body is inhabited by complex communities of microbes, collectively known as the human microbiota. The microbiota is estimated to exceed the human somatic and germ cells by more than a factor of 100 (Turnbaugh et al., 2007). The genetic content of the microbiota is referred to as the microbiome. High-throughput sequencing technologies enabled the identification of the species inhabiting the human body. It is widely regarded that microbes inhabit every anatomical site in the human body. Furthermore, the human body is host to a great diversity of bacterial taxa (Methé et al., 2012).

The microbiota plays an essential role in human health. It includes organisms that have evolved with humans and provide essential functions that the human body cannot perform itself. For example, *Escherichia coli* (*E. coli*) within the human gut plays the important role of synthetizing Vitamin K, which is a necessary coagulation factor in our blood (Blount, 2015). The human vagina and the microbiota that resides within are an example of mutualism. The bacterial communities that reside within the walls of the vagina play a protective role in preventing colonization of pathogenic bacteria, and in exchange, the host provides nutrients from cells and glandular secretions (Ma, Forney & Ravel, 2012).

Disruption of these microbiota (dysbiosis) has been associated with human disease. For example, a low diversity of bacterial taxa within the gut can lead to inflammatory bowel diseases

(IBS). (Turnbaugh et al., 2009; Qin et al., 2010). Also, dysbiosis of the "healthy" vaginal microbiota has been shown to be responsible for bacterial vaginosis, yeast infections, and sexually transmitted infections (STI) (Sewankambo et al., 1997; van De Wijgert et al., 2000; Wiesenfeld et al., 2003). Dysbiosis can be the result of one or many bacterial taxa increasing/decreasing in prominence. Bacterial vaginosis arises due to decreases in the abundance of Lactobacilli which enables *Gardnerella* species to quickly bloom and lead to inflammation (Macklaim et al., 2013). While exemplified here by just a few examples, there are many more. The role of the microbiota in the health of an individual is immense and the symbiosis of bacterial communities is essential.

Human Urinary Microbiome

In contrast to other anatomical sites, the urinary tract of healthy individuals was long thought to be a sterile environment. This belief dates to the mid-1800s (Thomas-White et al., 2016a). Traditionally, the presence of bacteria in urine is used to define infection within an individual (Thomas-White et al., 2016a). The inability to identify bacteria within urine from healthy individuals was due to technical difficulties in culturing these microbes (Wolfe et al., 2012; Hilt et al., 2014). However, advances in culturing methods, such as the use of the expanded quantitative urine culture (EQUC) protocol, proved that in fact the healthy urinary tract is inhibited by living microbes (Hilt et al., 2014; Price et al., 2016).

Some urinary symptoms are the direct response to pathogenic bacterial blooms. For example, *E. coli* is a frequent cause of acute UTIs in females (Klein & Hultgren, 2020). While other microbes can cause UTIs, current estimates indicate that *E. coli* is associated with ~80% of UTIs (Klein & Hultgren, 2020). The growth of *E. coli* from urine samples has been optimized in

the clinical laboratory and can easily be detected and quantified (Hooton et al., 2013). In fact, the Standard Urinary Culture (SUC) method, almost exclusively used for UTI diagnosis, was specifically developed to identify *E. coli* (Barnes et al., 2021). Nevertheless, other bacteria have been cultured from UTI patients. The bacterium *Proteus mirabilis* is frequently cultured from and the cause of UTIs of individuals with indwelling-catheters (Sabbuba, Mahenthiralingam & Stickler, 2003). In cases of recurrent UTIs (rUTIs), the uropathogen *Enterococcus faecalis* is also frequently found (Hochstedler et al., 2022). The EQUC method is efficient in culturing these and other non-*E. coli* taxa contributing to UTIs (Barnes et al., 2021).

There has been significant research dedicated to studying the differences in the urinary microbiomes of individuals with lower urinary tract symptoms, for example, overactive bladder (OAB) and individuals without any symptoms (Wu et al., 2017; Neugent et al., 2020; Li et al., 2022). Previous studies have surveyed the bacterial diversity of the urinary microbiome of females with urgency urinary incontinence (UUI) and those without lower urinary tract symptoms (LUTS) (Pearce et al., 2015; Karstens et al., 2016; Thomas-White et al., 2017; Nardos et al., 2022). A few studies have surveyed the urinary microbiome of females with and without UTIs (Moustafa et al., 2018; Chen et al., 2018; Anglim et al., 2021). Most surveys of the bacterial taxa within the urinary tract have utilized the 16S rRNA region; these studies are often limited in their resolution, able to confidently classify taxa only to the genus level.

Lactobacillus **in the Urinary Microbiome**

Lactobacillus species are common colonizers of the human microbiome, including the gastrointestinal (GI) tract, urinary tract, and vaginal microbiota (Ravel et al., 2011; Heeney, Gareau & Marco, 2018; Fok et al., 2018). The *Lactobacillus* genus is predominantly found in all

studies that have looked at the urinary microbiome (Pearce et al., 2014a; Wu et al., 2017; Thomas-White et al., 2017; Moustafa et al., 2018; Chen et al., 2018; Price et al., 2020; Anglim et al., 2021). In studies that looked at the urinary tract microbiome of females with UUI versus those without, *Lactobacillus* at the genus level was found in both cohorts (Pearce et al., 2015; Karstens et al., 2016; Thomas-White et al., 2017; Nardos et al., 2022). When the V4 hypervariable region of the 16S rRNA gene sequence was examined to the species level, *L. gasseri* was found more frequently in UUI cohort and *L. crispatus* was detected more frequently in women without UUI (Pearce et al., 2014a).

Lactobacillus species are dominant members of the healthy female urogenital microbiome, and they are thought to be the regulators of these communities. Several different Lactobacilli have been associated with mitigating growth of uropathogens. For instance, prior research has shown that *L. crispatus* provides protection against uropathogenic *E. coli* (UPEC) by mitigating some inflammatory cytokine effects (Butler, Silvestroni & Stapleton, 2016). For example, *L. jensenii* produces hydrogen peroxide (H_2O_2) (Gupta et al., 1998; Hütt et al., 2016), which provides antimicrobial defense (Singh, Hertzberger & Knaus, 2018). Lactobacilli also can create an acidic environment that restricts the growth of pathogens (Nunn & Forney, 2016). When *Lactobacilli* communities diminish, pathogens can fill the void. Furthermore, *L. gasseri*, *L. iners, L. paragasseri* as well as *L. jensenii* are associated with probiotic capabilities (Kirjavainen et al., 2009; Hütt et al., 2016; Oh et al., 2018). Thus, *Lactobacillus* species are essential to the balance and health of the environments they protect, including the urogenital tract.

Lactobacillus jensenii and Lactobacillus mulieris

Prior studies have found *L. jensenii* in the urinary microbiomes of both symptomatic and healthy individuals (Hilt et al., 2014; Thomas-White et al., 2018; Putonti et al., 2019; Price et al., 2020a,b; Komesu et al., 2020; Miller-Ensminger, Wolfe & Putonti, 2020; West-Pelak et al., 2020; Rivera et al., 2020; Ene & Putonti, 2022). In a recent study, *L. jensenii* was found to be one of the most frequently detected species within the urinary microbiome of females with SUI symptoms (Price et al., 2020b). *L. jensenii* also has been routinely reported in the vaginal microbiome (Romero et al., 2014; Fettweis et al., 2014; Komesu et al., 2020; Tortelli et al., 2020; Mehta et al., 2020; Dunlop et al., 2021; Witkin et al., 2021). Detection of *L. jensenii* in these studies often relied on 16S rRNA gene sequencing of the V4 region in urinary microbiome studies or the V1-V3 and V3-V4 regions in vaginal microbiome studies (Romero et al., 2014; Fettweis et al., 2014; Thomas-White et al., 2018; Komesu et al., 2020; Tortelli et al., 2020; Mehta et al., 2020; Dunlop et al., 2021; Witkin et al., 2021).

Recently, a new *Lactobacillus* species, *L. mulieris*, was discovered (Rocha et al., 2020). This new species is nearly indistinguishable from *L. jensenii* by the 16S rRNA gene sequence. Our prior analysis of these two species found that the 16S rRNA sequences of these two species could only be distinguished by two nucleotide differences, neither of which occurred within the commonly targeted hypervariable regions (Putonti et al., 2020). Thus, prior associations of *L. jensenii* and symptoms (or the lack thereof) in the urogenital tract are incomplete as they were conducted prior to the characterization of *L. mulieris*.

This thesis provides a comprehensive characterization of *L. jensenii* and *L. mulieris*, employing bioinformatic, microbiological, and molecular techniques. Building upon our prior

analysis of *L. jensenii* and *L. mulieris* genomes (Putonti et al., 2020), Chapter 2 examines 61 genome assemblies, including new strains sequenced here. In Chapter 3, the phenotypic differences of 40 strains from the Putonti Lab collection were analyzed. This includes looking at the difference between the two species in their lactic acid production, H_2O_2 production, secondary metabolite production, and sugar metabolism efficiency. In Chapter 4, I reassessed how prior urinary microbiome studies classified *L. jensenii*, leading to the development of new gene markers for these two species, and screened over 233 metagenomic samples. Conclusions and future directions are discussed in Chapter 5.

CHAPTER 2

GENOMIC INSIGHTS OF LACTOBACILLUS JENSENII AND LACTOBACILLUS MULIERIS

Introduction

The 16S rRNA sequence is a conserved region across all prokaryotic species and is widely used to assign taxonomy (Yarza et al., 2014). The most common method of typing bacterial species is based upon either Matrix-Assisted Laser Desorption/Ionization-Time Of Flight (MALDI-TOF) mass spectrometry or 16S rRNA gene sequencing (Toyoshima et al., 2021; Zhou et al., 2020). However, there are distinctly different species that share identical 16S rRNA gene sequences (Johnson et al., 2019), species with intragenomic variation in their 16S rRNA gene copies (Johnson et al., 2019), and even rare cases of horizontal gene transfer of 16S rRNA genes between species (Schouls, Schot & Jacobs, 2003). Therefore, 16S rRNA-based taxonomic classification systems have been eclipsed by alternative gene markers for some species as well as alternative methods.

The development of high-throughput sequencing technologies revolutionized how taxa are assigned to species. These technologies enable microbes to be quickly and inexpensively sequenced, and their complete genome to be assembled. Genome sequences are the most reliable way to distinguish between closely related taxa (Tanizawa et al., 2018; Zhou et al., 2020). 16S rRNA taxonomic classification has been replaced by Average Nucleotide Identify (ANI) (Konstantinidis & Tiedje, 2005). ANI measures the mean nucleotide sequence identity of shared

genes between two genomes; homologous genes of closely related strains of the same species often have identical sequences. ANI analyses generate a similarity index between all given pairs of genomes. Genomes with an ANI value greater than or equal to 95% are considered the same species (Konstantinidis & Tiedje, 2005).

As a single genome is not able to provide information about genetic variability of a species, multiple genomes from the same species need to be analyzed to infer the genic content and functional potential of a species. The catalog of all genes encoded by strains of the same species is referred to as the pangenome (Tettelin et al., 2005). The set of genes that are shared among all strains of a species are referred to as the core genome. The core genome includes genes and functionalities essential for microbial life as well as genes conserved amongst related taxa that enable it to thrive in its environment. Genes that are identified in one or more, but not all, genomes are considered the accessory genome. The accessory genome can include genes acquired via horizontal gene transfer and often serves as a reservoir of gene content for the species. The accessory genome within a single species can explain the phenotypic differences between strains. For example, some strains of *E. coli* can be pathogenic and lead to UTIs while other strains of *E. coli* are inhabiting a healthy urinary tract microbiome (Garretto et al., 2020) and this might be related to their strain-specific functions.

The accessory genome often contains genes that encode for a selective advantage within the bacterium's niche, including virulence factors, antibiotic resistance genes, secondary metabolites, and prophages. For example, *Helicobacter pylori* strains that encode for multiple different virulence factors have increased pathogenicity (de Brito et al., 2019). Secondary metabolites within the gut microbiota were shown to be linked with colorectal cancer (Louis,

Hold & Flint, 2014). Frequently, antibiotic resistance genes are exchanged between bacterial species and pose a mounting threat for human health (Zhu, Huang & Yang, 2022). Lastly, the accessory genome can contain prophages (bacterial viruses) that can confer protection against infection from another phage or it can drive genomic evolution (Brüssow, Canchaya & Hardt, 2004).

Genomics were integral in the redefinition of the *Lactobacillus* genus over the past few years (Wittouck, Wuyts & Lebeer, 2019; Wittouck et al., 2019; Zheng et al., 2020). *Lactobacillus* is a complex genus composed of more than 200 species, and it has a variety of food and medical applications (Wittouck et al., 2019). The *Lactobacillus* species *L. crispatus*, *L. iners*, *L. gasseri*, and *L. jensenii* are dominant members of the healthy female urogenital microbiome (Ravel et al., 2011; Heeney, Gareau & Marco, 2018; Fok et al., 2018). Genomic analyses of isolates from these four species resulted in the definition of two new sister taxa – *L. paragasseri*, a sister-taxon of *L. gasseri* (Tanizawa et al., 2018), and *L. mulieris*, a sister-taxon of *L. jensenii* (Rocha et al., 2020). Previously, we conducted ANI analysis of 43 publicly available *L. jensenii* and *L. mulieris* genomes and found that many *L. mulieris* strains were misidentified as *L. jensenii* strains (Putonti et al., 2020).

In-depth genomic analysis of *L. jensenii* and *L. mulieris* has yet to be conducted. While our own prior genomic analysis noted that *L. mulieris* genomes were predominately found in the urinary tract of women with urinary tract symptoms (Putonti et al., 2020), the sample size was limited (n=43). Here we present a comparative genomic study of 61 genomes for these two species. In addition to ANI analyses, we examined genic content in an effort to explore the divergence of these two species.

Methods

Data acquisition

All publicly available genome sequences of *L. jensenii* and *L. mulieris* were retrieved from NCBI as of September 18, 2021. Additional strains in our lab collection were sequenced as described in subsequent sections, and their genomes were deposited in NCBI and included in our analysis.

Sample acquisition

L. jensenii and *L. mulieris* strains tested were obtained through prior IRB-approved studies conducted by the Wolfe lab at Loyola University Chicago's Stritch School of Medicine (IRB approvals LUC206469, LUC207102, and LUC204195 from Loyola University Chicago and 17077AW from University of California San Diego) (Hilt et al., 2014; Pearce et al., 2014b, 2015; Price et al., 2016; Thomas-White et al., 2016b). Briefly, catheterized samples were collected and cultured using the EQUC [defined on page 3] (Hilt et al., 2014; Price et al., 2016). Strains were identified as *L. jensenii* by MALDI-TOF (as previously described (Hilt et al., 2014)) and stored at -80°C. All samples were isolated and identified via MALDI-TOF prior to the description of *L. mulieris*. Freezer stocks were first streaked on Columbia CNA agar with 5% sheep blood plates (BD 221353) and incubated at 35°C in 5% CO_2 for 48 hours in the Wolfe lab and then transferred to the Putonti lab. Upon receipt, a single colony was selected and grown in MRS+1% Tween 80 liquid media at 35°C in 5% CO_2 for 48 hours. This culture was then stored in 50/50 v/v glycerol at -80°C.

DNA extraction and sequencing

The following protocol was used to generate genomes produced as part of this work. Samples were extracted using a modified version of the Qiagen Blood and Tissue Kit Protocol. Pre-1 lysis buffer was made by combining 120 ul Triton x-100, 100 ul Tris-Cl at pH 8.0, 40 ul of EDTA, and 9.74 mL of nuclease free water. Lysis buffer was made by combining 150 ul of Pre-1 lysis buffer and 50 ul of 120 mg/mL Lysozyme. Liquid cultures of each strain, grown in MRS+1% Tween 80 liquid media at 35°C in 5% CO_2 for 48 hours, were centrifuged at 7500 x g for 10 minutes. The supernatant was removed and discarded. The pellet was resuspended in 200 ul of lysis buffer and incubated at 37°C for 30 minutes. After incubation, the protocol described by the manufacturer was followed. DNA concentrations were quantified using a Qubit fluorometer following the manufacturer's protocol. Samples were sequenced at MIGS (Pittsburgh, PA). There, sequencing libraries were prepared using the Illumina Nextera Kit and samples were sequenced using the Illumina NextSeq 550 platform (150 bp x 2, paired-end reads). We also received Illumina paired-end sequencing reads of 9 strains from the Wolfe lab. DNA extraction, library preparation, and sequencing was performed by the Wolfe lab, while assembly (as described below) was performed as part of this project.

Assembly

Raw reads were first trimmed for quality using BBDuk v. 38.92 ("BBMap") with the following parameters ftl=15 ftr=135 minlength=30 qtrim=rl maq=20 maxns=0 statscolumns=5 trimq=20. The genomes were assembled via SPAdes v. 3.15.2 using the assembly-only option (Bankevich et al., 2012). Genome assemblies were made publicly available by depositing them

in NCBI's Assembly database. When deposited, the genome assemblies were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v. 5.3 (Tatusova et al., 2016).

ANI analysis

The average nucleotide identity (ANI) was computed using pyani v0.2 (Pritchard et al., 2015). From the ANIm percentage identity values, we classified the genomes into species based on the 95% ANI threshold routinely used in the field for distinguishing between closely related bacterial taxa (Jain et al., 2018).

Secondary metabolites, antibiotic resistance, and virulence factor screening

Each genome was screened for secondary metabolites via antiSMASH using the default parameters and all extra features option of the web-based tool (Blin et al., 2021). The biosynthetic gene clusters sequences found by antiSMASH were aligned using MAFFT (v7.388) (Katoh & Standley, 2013) and a phylogenetic tree was created as described below. Genomes were screened for antibiotic resistance genes using arpcard (https://github.com/arpcard/rgi). Virulence factor genes were predicted using the Virulence Factor Database (VFDB), a curated database of pathogenic bacterial virulence factor sequences (Chen et al., 2005). Virulence factor gene sequences and metadata were retrieved from the VFDB website (http://www.mgc.ac.cn/VFs/download.htm), and a local blast database was created. Each genome was then queried locally via blastn against the database and the top hits were examined.

Prophage Identification

Each genome sequence was screened for prophage sequences using PHASTER (Arndt et al., 2016). While PHASTER predicted intact, questionable, and incomplete prophage sequences, only intact prophage sequences were examined in depth. The intact prophage nucleotide

sequences were used to create a phage network based on homologous genes shared among prophages. The prophage sequences were annotated using anvi'o (v7.1) pangenomics tool, which annotates the sequences using the COG database and then identifies homologous genes between prophage sequences (Eren et al., 2015) The network was made as follows. Prophages are represented as nodes; if two prophage sequences contain homologous genes, then they are connected by an edge. The network was visualized with Cytoscape 3.8.2 (Shannon et al., 2003).

Core and pangenome

The core and pangenome for the publicly available genomes and genomes sequenced as part of this work were determined using anvi'o v7.1 (Eren et al., 2015). Frist, contigs less than 1000bp were removed using the command anvi-script-reformat-fasta; afterwards, the command anvi-gen-contigs-database was used to generate databases for each genome. The anvi-pan-genome command was used to create the pangenome of all the genomes with an mcl of 9. The concatenated single-copy core genome was found using the command anvi-get-sequences-for-gene-clusters with the --min-num-genomes-gene-cluster-occurs number of genomes --max-num-genes-from-each-genome 1 --concatenate-gene-clusters. The output for this command is amino acid sequences. The pangenome was created using anvi-get-sequences-for-gene-clusters with default settings. The set of core genes was found using the command anvi-get-sequences-for-gene-clusters with the --min-num-genomes-gene-cluster-occurs number of genomes --max-num-genes-from-each-genome 1. The functional enrichment of the genomes was found using the anvi-compute-functional-enrichment-in-pan command.

Phylogenetic tree

Phylogenies of sequences were created as follows. Sequences were imported into Geneious Prime 2022.1.1 (Biomatters Ltd., Auckland, NZ) and aligned using the MAFFT v7.388 (Katoh & Standley, 2013) plug-in through Geneious Prime. The phylogenetic tree was derived using the FastTree 2.1.12 (Price, Dehal & Arkin, 2010) plug-in through Geneious Prime and visualized using iTOL v6 (Letunic & Bork, 2016).

16S rRNA gene sequence analysis

The 16S rRNA reference sequence for *L. jensenii* was obtained from the SILVA database (Quast et al., 2013). This sequence was used to create a local nucleotide blast database (Altschul et al., 1990). Each genome was then queried against this database using blastn, in order to find the 16S rRNA sequence. The resulting16S rRNA gene sequences were imported into Geneious Prime (Biomatters Ltd., Auckland, NZ) and aligned using the MAFFT v7.388 (Katoh & Standley, 2013) plug-in through Geneious Prime. Nucleotide differences between the sequences were identified, and their location within conserved or variable regions was determined by aligning the consensus sequence against the *E. coli* 16S rRNA gene sequence (GenBank Accession No. J01859.1) (Brosius et al., 1978). The phylogenetic tree was derived using the FastTree 2.1.12 (Price, Dehal & Arkin, 2010) plug-in through Geneious Prime and visualized using iTOL v6 (Letunic & Bork, 2016)

Results and Discussion

L. jensenii and *L. mulieris* genomes

Our genomic analysis of *L. jensenii* and *L. mulieris* included 61 genomes. Forty-five are publicly available sequences. As part of this thesis work, I extracted, sequenced, and deposited an additional 4 *L. jensenii* strains and 3 *L. mulieris* strains; these strains were isolated from urine samples (Ene & Putonti, 2022). The other 9 genomes in this analysis, also for urinary isolates, were obtained from the Wolfe Lab. Overall, the majority of the strains included in this analysis were isolated from urine (n=41). Five of the 45 publicly available genomes are from strains collected from fecal samples, 10 were isolated from vaginal samples, and 8 are from an unknown source. More information on the genomes can be found in Appendix I.

Genomic analysis

The taxonomies of these 61 strains were assigned based upon their average nucleotide identity (ANI) values. The 95% threshold for pairwise ANI values was used as the species boundary between the two sister taxa (Ramasamy et al., 2014). Pairwise ANI values ranged between 88.66% and 100%. From this ANI analysis, two species can be distinguished. In total, 36 strains were identified as *L. jensenii* and 25 strains were identified as *L. mulieris*. The average pairwise ANI between *L. jensenii* strains is 99.81%, while the average pairwise ANI of *L. mulieris* strains is 99.49%. The ANI profile for these two species is shown in Figure 1.

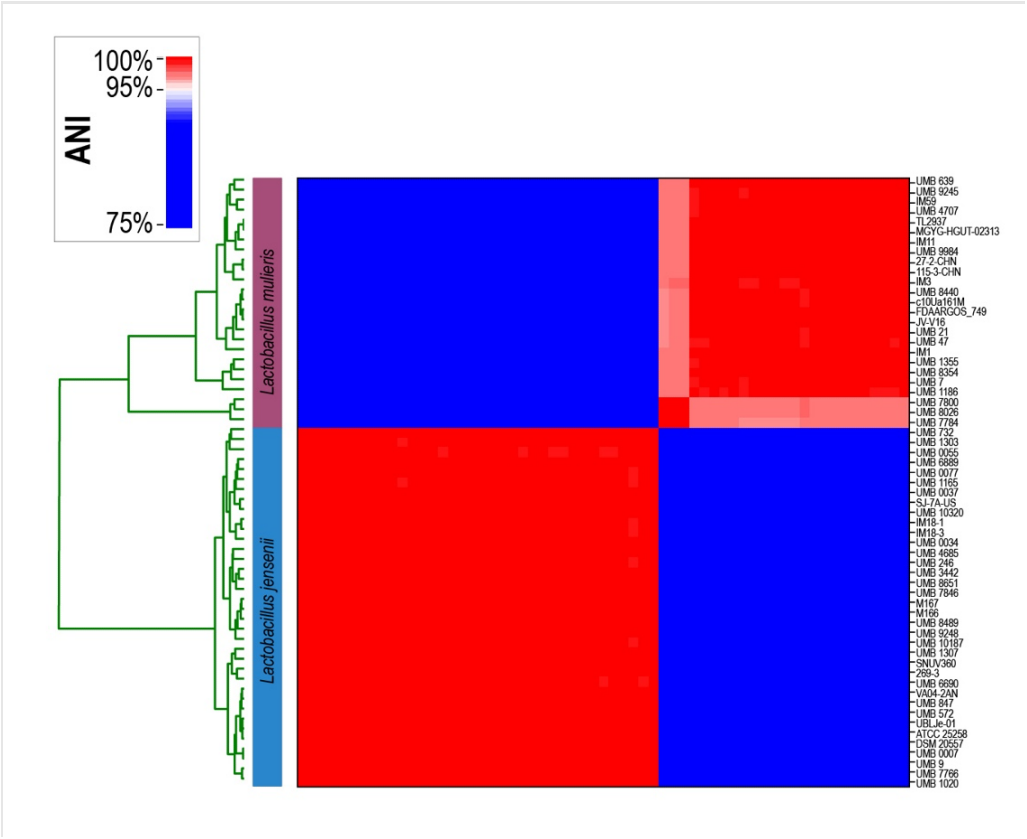

Figure 1. ANI analysis of *L. jensenii* and *L. mulieris* strains. The dark purple rectangle on the upper left side indicates the *L. mulieris* strains, while the light blue rectangle below it shows the *L. jensenii* strains.

There are three genomes within the *L. mulieris* group that form their own subgroup: UMB7784 (rUTI9), UMB7800 (rUTI12) and UMB8026 (rUTISD018). Their pairwise ANI compared with the *L. mulieris* type strain c10Ua161M (GCA_007095465), are 97.17%, 97.35% and 97.34%, respectively (Rocha et al., 2020). Thus, by the 95% threshold, they are strains of the *L. mulieris* species. These three divergent *L. mulieris* strains were isolated from three different females. However, all three females were clinically diagnosed with a rUTI infection.

Previously, we identified 17 *L. mulieris* strains (Putonti et al., 2020). Our analysis here provides 8 additional representatives of this new species. Our prior analysis also noted the ANI difference between UMB7784 and the other *L. mulieris* strains (Putonti et al., 2020). With the

subsequent sequencing of the urinary isolates UMB7800 and UMB8026, we find that other strains belong to this subclade. These three genomes might be an emerging species. Subsequent sequencing of isolates from the urogenital tract may uncover new *Lactobacillus* species or subspecies.

The pangenome of these two species, including all 61 strains, was next computed. In total, 2,636 unique genes were found. These genes can be found in one or more of the 61 genomes. The single-copy core genome, or the set of genes that are found in all genomes (in a single copy), includes 589 genes. (Henceforth, the single-copy core will simply be referred to as the "core genome".) The accessory genome is composed of 1,835 genes. This accessory genome consists of genes that are found in one or more genomes, but not the core genome.

Given that more genomes were considered here than in our prior analysis of these two species (Putonti et al., 2020), it is important to note that the core genome is larger here (589 genes) than previously found when only 43 genomes were considered, which was just 453 genes. The reason why is likely due to the different MCL threshold used. The Markov CLustering algorithm, or MCL, identifies clusters of amino acid sequence similarity, with values ranging between 2 and 10. An MCL of 2 should be used for vastly different genomes such as genomes obtained from different families of bacteria, and an MCL of 10 should be used for genomes obtained from the same species (van Dongen & Abreu-Goodger, 2012). In the previous analysis, an MCL of 10 was used, essentially considering the *L. jensenii* and *L. mulieris* genomes as members of a single species (Putonti et al., 2020). Our further investigation of these genomes, however, informed our revision of this threshold. In my analysis of the 61 genomes here, I used an MCL of 9. In contrast to an MCL of 10, this threshold permits more sequence divergence

between homologs. In doing so, more homologs (gene clusters) were identified resulting in a larger core genome for these two species.

In comparison to our prior analysis of 43 genomes (Putonti et al., 2020), the size of the accessory genome for the 61 genomes examined here grew slightly. Our previous analysis identified an accessory genome of 1,738 genes. The addition of 18 genomes only increased the accessory genome by 97 genes. This modest increase in the accessory genome size, ~5 genes per new genome considered, is indicative of a closed pangenome. Prior genomic analyses of Lactobacilli similarly posit a closed genome (Inglin, Meile & Stevens, 2018).

Next, the amino acid sequences in the core genome were aligned and a phylogenetic tree was derived. The figure shows the distinct *L. jensenii* and *L. mulieris* clades. This tree also includes the *L. mulieris* subclade of the three genomes that had more distant ANI values from the other *L. mulieris* genomes. Figure 2 also includes information about the isolation source for each strain.

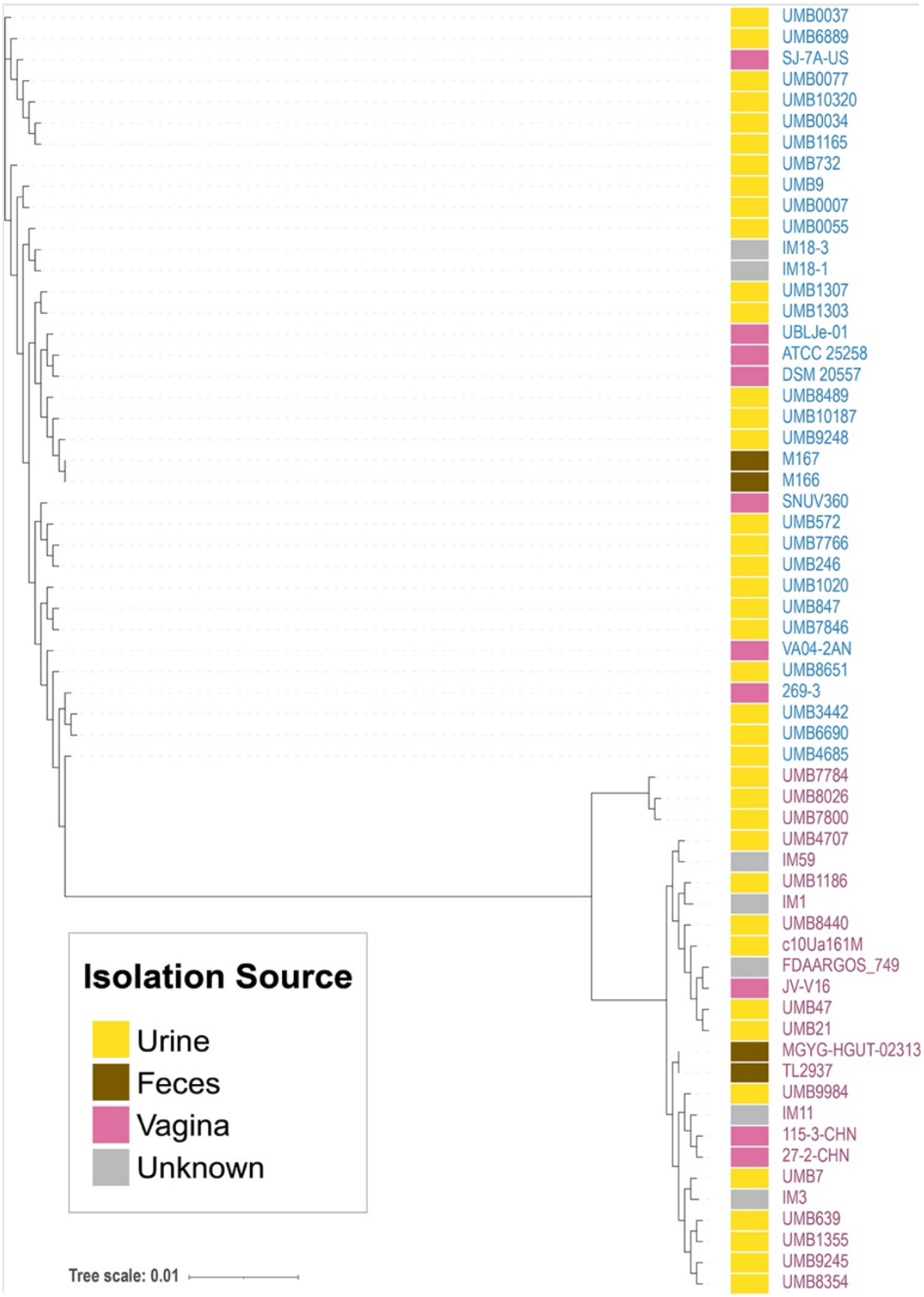

Figure 1. Core phylogenetic tree of *L. jensenii* (text in blue) and *L. mulieris* (text in pink) with the source of the strains labeled by the rectangles on the right according to the legend.

As can be seen in Figure 2, the core genome is a sufficient signal of the divergence of the two species. This tree further confirms our findings from the ANI analysis (Figure 1). The same strains in the *L. jensenii* ANI clade are in the *L. jensenii* core phylogenetic tree clade. Likewise, the same strains in the *L. mulieris* ANI clade are in the *L. mulieris* core phylogenetic tree clade. As previously mentioned, the core genome tree also shows the subgrouping of the three *L. mulieris* genomes: UMB7784, UMB7800, and UMB8026. This indicates that these genomes' core genome represents a distinct lineage within the *L. mulieris* species. Moreover, both species includes strains from the four different isolation sources: urine, fecal, and vaginal samples (Figure 2). Thus, neither is niche-specific.

To further investigate the genic content of each species, pangenome analysis was repeated, this time for genomes of each individual species. Each strain's species designation was assigned by the ANI analysis (Figure 1). For the 36 strains of *L. jensenii*, 904 genes were identified in their core genome. The 25 *L. mulieris* genomes had a larger core genome, which includes 1,190 genes. The *L. jensenii* core genome had 103 genes that were not shared with the *L. mulieris* core genome, i.e., 103 genes are conserved (in a single copy) among all *L. jensenii* genomes and are not conserved (in a single copy) among all 25 *L. mulieris* genomes. Thirty-two of these genes are specific to *L. jensenii* and are not found in any of the 25 *L. mulieris* genomes examined here. The *L. mulieris* core includes 389 genes that were not shared with the *L. jensenii* core genome, with 64 of these genes not found in any of the *L. jensenii* genomes examined here.

The differences between these two species can also be seen in the functional capacity of their encoded genes. Using the Clusters of Orthologous Groups of proteins (COGs) database, which is a phylogenetic classification of the proteins encoded in bacterial genomes, only genes

predicted to be different between the two strains were examined (Tatusov, Koonin & Lipman, 1997). Table 1 lists the predicted protein functions that are in all of the *L. mulieris* genomes examined and not in any of the *L. jensenii* genomes examined. As it can be seen in Table 1, *L. mulieris* species encode for different functionalities related to carbohydrate metabolism and transport pathway. The *Lactobacillus* genus is generally considered to be a probiotic; therefore, carbohydrate pathways are very important. The carbohydrate pathways of *Lactobacillus* are important because it confers an advantage in surviving in the various competitive environments of the body such as the gastrointestinal tract where *Lactobacillus* can utilize complex carbohydrates such as milk (Zunga, Yebra & Monedero, 2021). The carbohydrate pathways are essential in surviving competitive environments such as the bladder, where other pathogenic species such as *E. coli* can thrive (Zunga, Yebra & Monedero, 2021). For example, *L. mulieris* encodes for COG1363, Alpha-glucosidase/xylosidase which can catalyze the transfer of alpha-xylosyl residue from alpha-xyloside to xylose, glucose, maltose, nigerose, sucrose, and trehalose . This might be indicative of *L. mulieris* having different carbohydrate metabolisms (Okuyama et al., 2004). Other *L. mulieris*-specific predicted functions are part of the nucleotide transport and metabolism or cell cycle control.

Table 1. *L. mulieris*-specific COG functions

COG category	COG20_FUNCTION	COG ID
Nucleotide transport and metabolism	Cytosine/adenosine deaminase or related metal-dependent hydrolase (SsnA) (PDB:3O7U)	COG0402
Cell cycle control and mitosis	Zn-dependent membrane protease YugP (YugP)	COG2340, COG5271
Amino Acid metabolism and transport	5-carboxyvanillate decarboxylase LigW (lignin degradation), amidohydro domain (LigW) (PDB:2DVT) (PUBMED:26714575)	COG1168

Carbohydrate metabolism and transport	NADPH-dependent 2,4-dienoyl-CoA reductase, sulfur reductase, or a related oxidoreductase (FadH2) (PDB:1VRQ)	COG2159
Carbohydrate metabolism and transport	Uncharacterized conserved protein YxeA, DUF1093 family (YxeA) (PDB:2K5Q)	COG3507
Carbohydrate metabolism and transport	Beta-xylosidase (XynB2) (PDB:1Y7B)	COG1196, COG1501
Carbohydrate metabolism and transport and Amino Acid metabolism and transport	Chromosome segregation ATPase Smc (Smc) (PDB:5XG3); Alpha-glucosidase/xylosidase, GH31 family (YicI) (PDB:1WE5)	COG1363
Coenzyme metabolism	Protoheme ferro-lyase (ferrochelatase) (HemH) (PDB:1AK1)	COG0276
Coenzyme metabolism	Uncharacterized conserved protein, DUF1430 domain	COG0667
Lipid metabolism	Pyridoxal reductase PdxI or related oxidoreductase, aldo/keto reductase family (PdxI) (PDB:1LQA)	COG0446
Translation	Putative aminopeptidase FrvX (FrvX) (PDB:1VHE)	COG5271
Post-translational modification, protein turnover, chaperone functions	Ammonia channel protein AmtB (AmtB) (PDB:1U77)	COG2738
Inorganic ion transport and metabolism	Midasin, AAA ATPase with vWA domain, involved in ribosome maturation (MDN1) (PDB:6HYP)	COG0004
Function Unknown	Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities (MalY) (PDB:4DGT)	COG5294
Function Unknown	Spore germination protein YkwD and related proteins with CAP (CSP/antigen 5/PR1) domain (YkwD) (PDB:1CFE) (PUBMED:31199835)!!!Midasin, AAA ATPase with vWA domain, involved in ribosome maturation (MDN1) (PDB:6HYP)	COG4652
Function Unknown	Uncharacterized conserved protein, DUF2316 domain	COG4367

Next, the functionality of the core genes of *L. jensenii* that were not found in the core of *L. mulieris* were examined (Table 1). One functionality of *L. jensenii* that *L. mulieris* does not have is related to energy production and conversion which could explain the ability of *L. jensenii*

to metabolize carbohydrates differently than *L. mulieris*. *L. jensenii* has Glucose uptake protein GlcU (GlcU), which is a transmembrane protein that facilitates transport of glucose into the cell (Fiegler et al., 1999).

Table 2. *L. jensenii*-specific COG functions

COG category	COG20_FUNCTION	COG ID
Energy production and conversion	FAD/FMN-containing lactate dehydrogenase/glycolate oxidase (GlcD) (PDB:1AHU)	COG1154
Energy production and conversion	Cytochrome bd-type quinol oxidase, subunit 2 (AppB) (PDB:6RKO)	COG4988
Energy production and conversion	NADH dehydrogenase, FAD-containing subunit (Ndh) (PDB:5NA4)	COG0277
Energy production and conversion	Cytochrome bd-type quinol oxidase, subunit 1 (AppC) (PDB:5DOQ)	COG1294
Carbohydrate metabolism and transport	Glycogen debranching enzyme (alpha-1,6-glucosidase) (GDB1) (PDB:5D06)	COG1575
Carbohydrate metabolism and transport	Glucose uptake protein GlcU (GlcU)	COG4987
Coenzyme metabolism	1,4-dihydroxy-2-naphthoate polyprenyltransferase (MenA)	COG1252
Coenzyme metabolism and Lipid metabolism	Deoxyxylulose-5-phosphate synthase (Dxs) (PDB:2O1S)	COG3408
Post-translational modification, protein turnover, chaperone functions	ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components (CydD)	COG4975
Post-translational modification, protein turnover, chaperone functions	ABC-type transport system involved in cytochrome bd biosynthesis, fused ATPase and permease components (CydC)	COG1271
Secondary Structure	Acyl carrier protein (AcpP) (PDB:1ACP)!!!EntF, seryl-AMP synthase component of non-ribosomal peptide synthetase (EntF) (PDB:5ES5)	COG0236, COG1020

Distinguishing between *L. jensenii* and *L. mulieris* via gene markers

Traditionally, the hypervariable 16S rRNA gene regions have been used to distinguish between species (Van de Peer, Chapelle & De Wachter, 1996). Thus, I extracted the 16S rRNA gene sequences from the 61 genomes (see Methods). Seven of the genomes, however, did not include a full 16S rRNA gene sequence: UMB3442, UMB0847, 269-3, IM1, IM3, UMB9245, and UMB7, and thus were excluded from analysis. In total, 59 16S rRNA sequences were found, as some genomes contained more than one copy of a 16S rRNA sequence. A phylogenetic tree was derived (Figure 3), showing a clear distinction between the 16S rRNA gene sequences from *L. jensenii* and *L. mulieris* strains. For genomes with more than one 16S rRNA gene sequence, the intragenomic variation was less than the interspecies variation.

The 59 aligned 16S rRNA gene sequences have a length of 1,575 base pairs with an average pairwise identity of 99.9%. The two species have 1,550 (98.4%) identical residues and can be differentiated by only two nucleotide positions, neither of which is within a variable region. These differences occur at position 76, located in the conserved region C1, in which *L. mulieris* has a T while *L. jensenii* has an A. At position 399, located in the conserved region C2, *L. mulieris* has a C while *L. jensenii* has an A. Because short-read sequencing of the V1-V3 regions is a frequent target in vaginal microbiome studies (Romero et al., 2014; Fettweis et al., 2014; Thomas-White et al., 2018; Komesu et al., 2020; Tortelli et al., 2020; Mehta et al., 2020; Dunlop et al., 2021; Witkin et al., 2021), these studies may amplify positions 76 and 399 (depending upon the placement of the V1 primer in the C1 region) and thus be capable of distinguishing between the two species. In contrast, urinary microbiome 16S rRNA gene amplicon surveys have primarily targeted the V4 region (Hoffman et al.), which is upstream of both of these positions.

Therefore, the 16S rRNA gene sequence is not a robust marker for distinguishing between *L. jensenii* and *L. mulieris*.

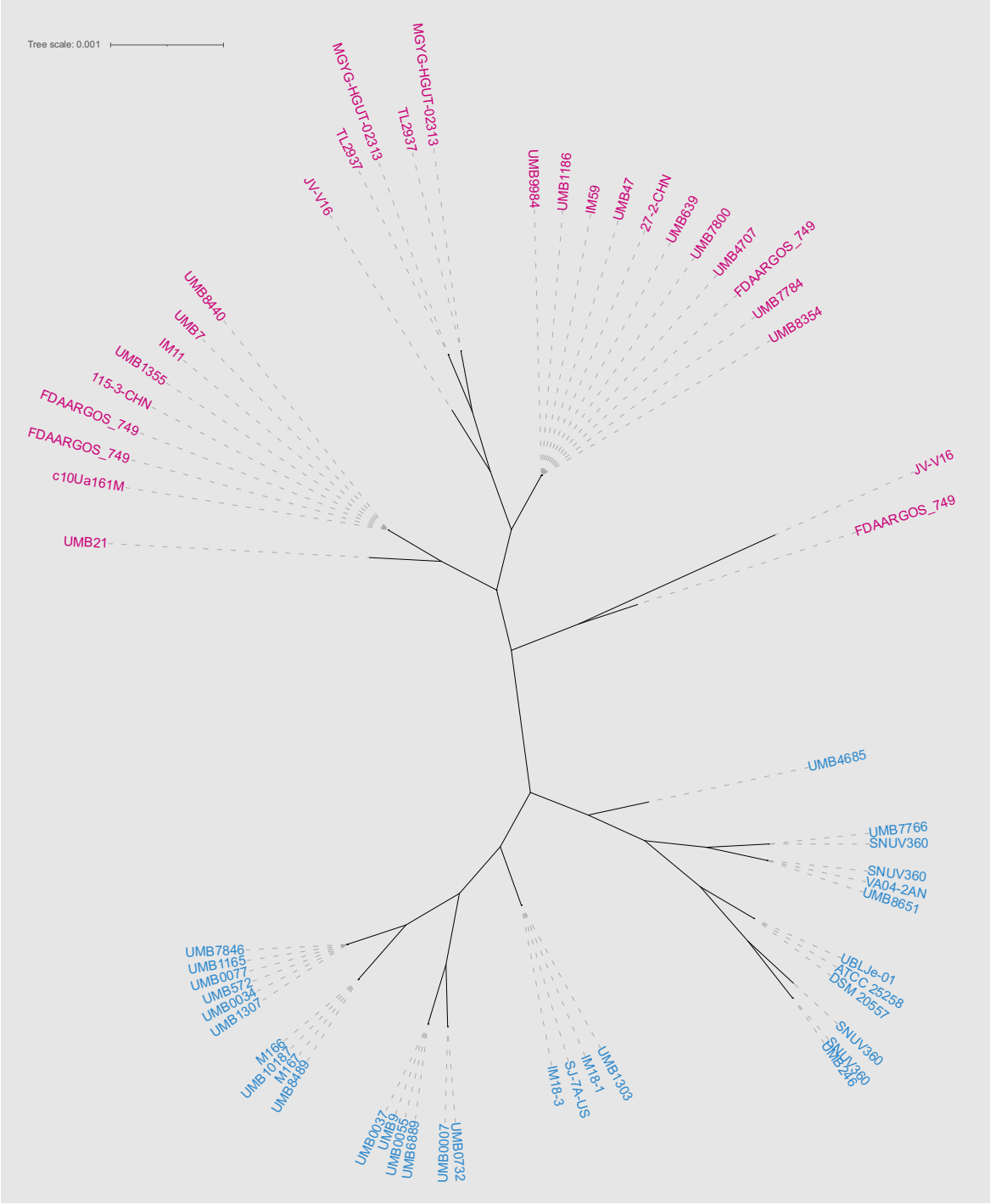

Figure 2. Phylogenetic tree of the 16S rRNA gene sequences. *L. jensenii* strain names are indicated in blue, and *L. mulieris* strain names are indicated in pink. Species designation is based upon ANI and core analysis.

Secondary metabolites

Each genome was individually screened for hypothetic gene clusters using antiSMASH (Blin et al., 2021). The results of the analysis are shown in Table 3. Three different types of biosynthetic gene clusters were identified: ribosomally synthesized and post-translationally modified peptide products (RiPP), domains with non-ribosomal peptide synthases (NRPS), and Class IV lanthipeptide clusters. While biosynthetic gene clusters with NRPS domains were only found in *L. jensenii* strains; RiPP-like and Class IV lanthipeptide clusters were only found in *L. mulieris* strains.

Table 3. Summary of secondary metabolites identified in the *L. jensenii* and *L. mulieris genomes.*

Species	RiPP-like	nrps	lanthipeptide-class-iv	No hits
L. jensenii	0	34	0	2
L. mulieris	3	0	17	5

The RiPP-like biosynthetic gene clusters were only found in three genomes: UMB7784, UMB7800, UMB8026. These genomes were shown previously to form a subclade in the *L. mulieris* group (Figures 2 and 3). To further investigate the RiPP-like clusters, their nucleotide sequences were aligned and a phylogenetic tree was derived (Figure 4). The length of the RIPP-like cluster is 5,728 bp. UMB 7800 and 8026 are 100% identical; however, their similarity with UMB7784 is 52.4%. The RiPP nucleotide sequences also were compared against the NCBI nr/nt database online using blastn. This resulted in homologous hits with a percent identity of 89.9% to *Lactobacillus sakei* IP-TX (AY206863.1), 88.13 % to *Enterococcus faecium* (AB908994.1), and 89.45 % to *Latilactobacillus curvatus* (CP031003.1). Given the lack of this gene cluster in

other *L. mulieris* strains and its presence in these other species, we hypothesize that these three *L. mulieris* strains may have acquired these genes via horizontal gene transfer.

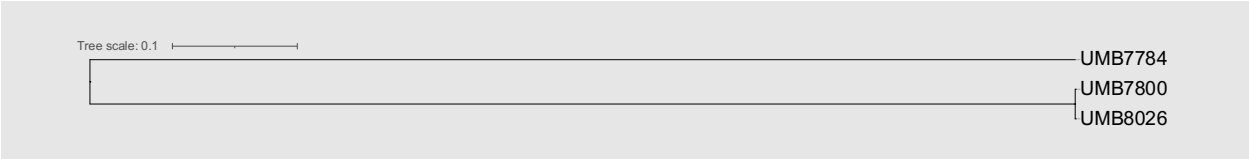

Figure 3. Phylogenetic tree of *L. mulieris* RIPP-like cluster nucleotide sequences.

Furthermore, 17 out of the 25 genomes of *L. mulieris* encode for a lanthipeptide-class-iv peptide biosynthetic gene cluster. [It is worth noting that the three strains in the *L. mulieris* subclade do not include this lanthipeptide-class-iv peptide.] Lanthipeptides are a major family of RiPPs and are classified into four different classes based on the biosynthetic enzymes that install the Lan and MeLan motifs (Arnison et al., 2012). All Class IV lanthipeptides discovered to date are structurally homologous to the first discovered IV lanthipeptide, venezuelin, from *Streptomyces venezuelae* (Goto et al., 2010). The type strain of venezuelin (HQ328852) was compared to the *L. mulieris* sequences and a phylogenetic tree was created (Figure 5). As displayed in this tree, the *L. mulieris* Class IV lanthipeptide nucleotide sequences are essentially identical among the *L. mulieris* strains and distinct from venezuelin.

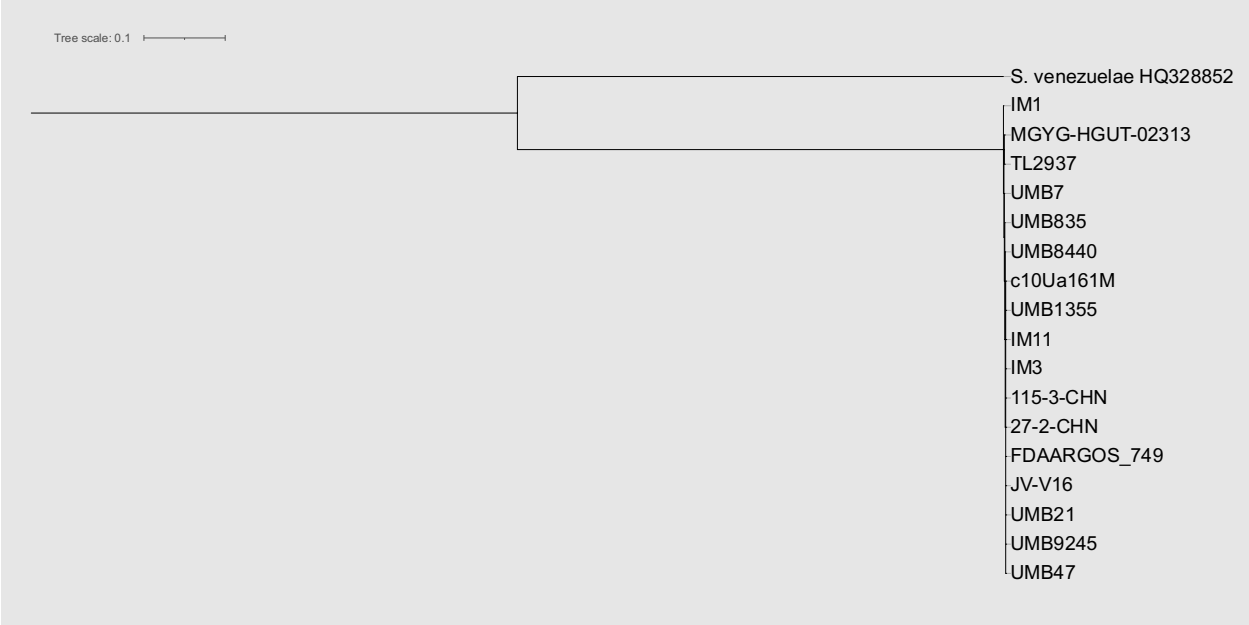

Figure 4. Lanthipeptide-class-iv phylogenetic tree.

All but 2 of the *L. jensenii* strains examined here were predicted to contain a biosynthetic gene cluster containing an NRPS domain. The two *L. jensenii* strains lacking the NRPS domain are UMB0055 and UMB8489. NRPS are widely known to have a variety of properties such as siderophores, toxin pigments, immunosuppressants, or anticancer agents (Wang et al., 2014; Martínez-Núñez & López, 2016). The *L. jensenii* NRPS sequences were aligned, and a phylogenetic tree was created (Figure 6). Upon further analysis, the nucleotide sequence of the NRPS were queried against the NCBI nr/nt database; the only records containing this sequence were from *L. jensenii* strains, suggesting that this sequence is unique to the species. However, when the NRPS sequences were queried using blastx, significant sequence similarity was found to the L. jensenii SJ-7A-US AMP-binding enzyme (EEX27994.1). These blast analyses suggest

that this NRPS biosynthetic gene cluster is specific to the Lactobacillus family.

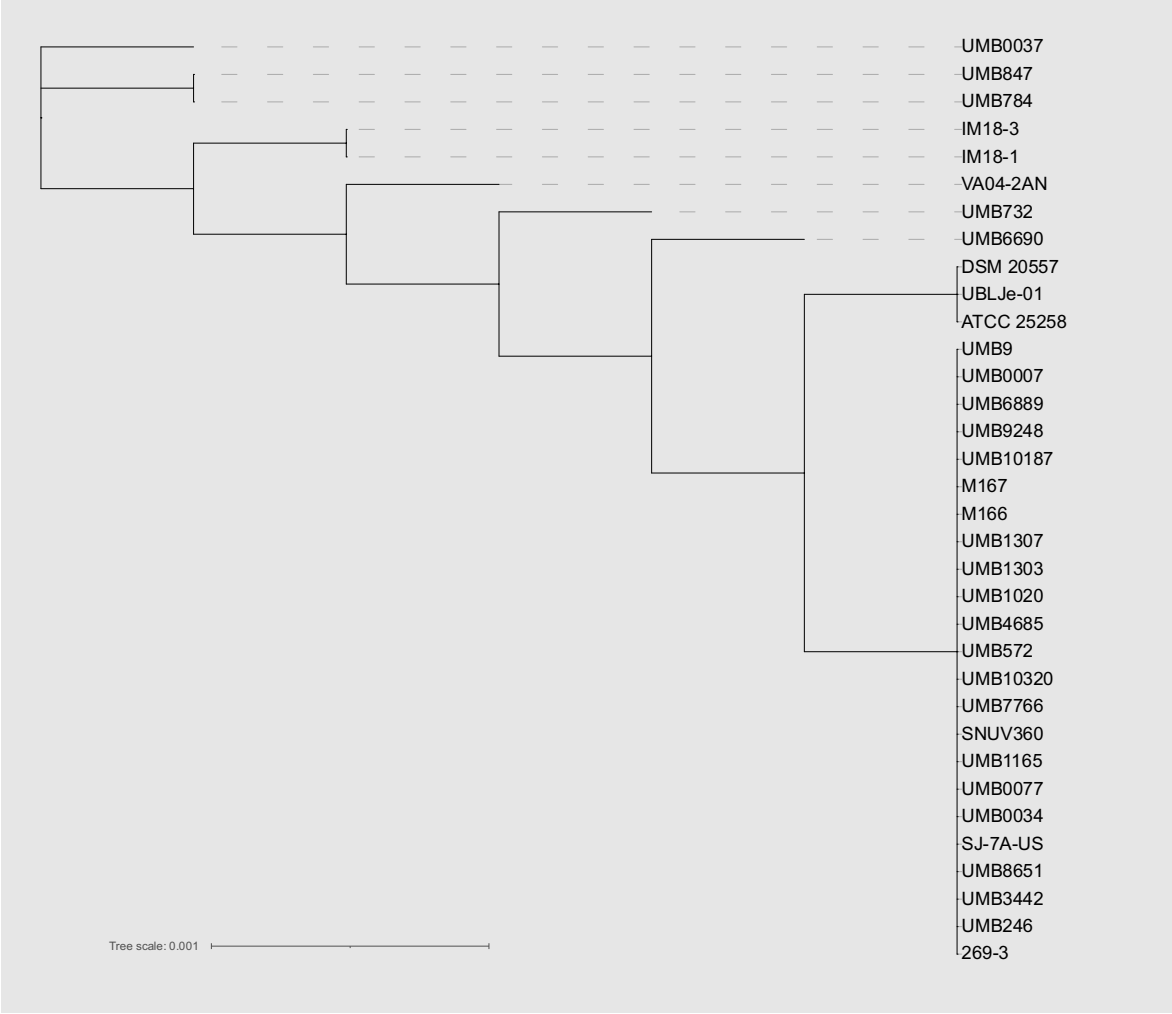

Figure 5. NRPS phylogenetic tree

Antibiotic resistance screening

The *L. jensenii* and *L. mulieris* genome sequences also were examined with regards to their antibiotic resistance gene content. All genome assemblies were screened for antibiotic resistance genes using the Comprehensive Antibiotic Resistance Database (CARD). While none of the genomes had exact (100% identity) matches to CARD antibiotic resistance reference gene sequences, four strains, *L. mulieris* UMB7800, *L. mulieris* UMB9984, *L. mulieris* UMB0047,

and *L. jensenii* UMB8651, had antibiotic resistance genes detected using CARD's Strict Algorithm (Table 4). The Strict Algorithm is designed to detect previously unknown variants of known resistance genes (Alcock et al., 2020, p. 2). All four of these strains are from urinary isolates.

Table 4. Antibiotic resistance prediction

Species	Strain (symptom status)	Drug Class	Resistance Mechanism	AMR Gene Family
L. mulieris	UMB0047 (no LUTS)	fluoroquinolone antibiotic	antibiotic target protection	quinolone resistance protein (qnr)
L. mulieris	UMB7800 (rUTI)	phenicol antibiotic	antibiotic inactivation	chloramphenicol acetyltransferase (CAT)
L. mulieris	UMB9984 (UTI)	tetracycline antibiotic	antibiotic target protection	tetracycline-resistant ribosomal protection protein
L. jensenii	UMB8651 (OAB)	carbapenem; cephalosporin; cephamycin; penam; penem	antibiotic inactivation	IMP beta-lactamase

Virulence factors

The 61 genomes were next queried for virulence factors, revealing 426 virulence genes. In total three unique virulence factors were found in the *L. jensenii* and *L. mulieris* strains. These virulence factors were part of Immune modulation, Adherence, and Exoenzyme virulence factor categories. One of these virulence genes, the Beta-C protein is documented to bind the Fc portion of serum IgA and the complement inhibitor factor H; it is believed to play a role in preventing opsonophagocytosis (Jerlström et al., 1996). Another identified virulence gene, the C5a peptidase, cleaves C5a, the major neutrophil chemoattractant produced by activation of the complement cascade, thus causing impaired recruitment to sites of infection (Cheng et al., 2001).

This virulence gene may function as an invasion and Fibronectin binding protein. And lastly, the hyaluronidase virulence gene helps facilitate the spread of bacteria by breaking down the hyalurone polymers present in the extracellular matrices of the host; the GBS hyaluronate lyase also has limited specificity for achondroitin sulphate and cleaves the chain at unsulphated sites. This action may facilitate deep tissue penetration during infection (Baker & Pritchard, 2000).

Prophage prediction

Finally, the *L. mulieris* and *L. jensenii* genomes were screened for prophage sequences. In total 6 questionable (low confidence) prophage sequences, 50 incomplete (partial) prophage sequences, and 34 intact prophage sequences were found. We further focused on the intact prophage sequences as these are most likely to be viable phages. The prophage sequences were queried against the nr/nt database via blastn. The 34 prophage sequences had sequence similarity to seven different phages: five metagenome-assembled genomes (MAGs) from human metagenomes (phage ct8SJ1, n=4; phage ctX581, n=6; phage ctXC91, n=2; phage ctXdY1, n=1; and phage ct8NQ1, n=1) (Tisza & Buck, 2021), *Lactobacillus* phage Lv-1 (n=11) (Martín et al., 2009; Martín, Escobedo & Suárez, 2010), and *Lactobacillus* phage Lu-1 (n=9) (Miller-Ensminger et al., 2020). All seven of these phages are predicted to be tailed siphoviruses, although only *Lactobacillus* phage Lv-1 has been characterized in the lab (Martín et al., 2009; Martín, Escobedo & Suárez, 2010). *Lactobacillus* phage Lu-1 was discovered by our group from the genomes of previously sequenced urinary *L. jensenii* strains, including a few in this study (Miller-Ensminger et al., 2020).

To compare these prophage sequences, they were annotated, and homologous genes were identified (see Methods). The 34 predicted intact prophages are highly similar with regards to

their gene content. To visualize this similarity, a network of prophages was created in which each prophage was represented by a node, and if two prophages share one or more homologous genes, they are connected by an edge Figure 7). The fact that all 34 prophages are connected indicates that the phages all have shared gene content. This suggests that these phages evolved from a common ancestor and/or exchange gene content. Furthermore, this network analysis suggests that these phages are capable of infecting both *L. jensenii* and *L. mulieris*. Prophages similar to MAG ct8SJ1, MAG ctX581, *Lactobacillus* Lv-1, and *Lactobacillus* Lu-1 phages are found in both *L. jensenii* and *L. mulieris* strains.

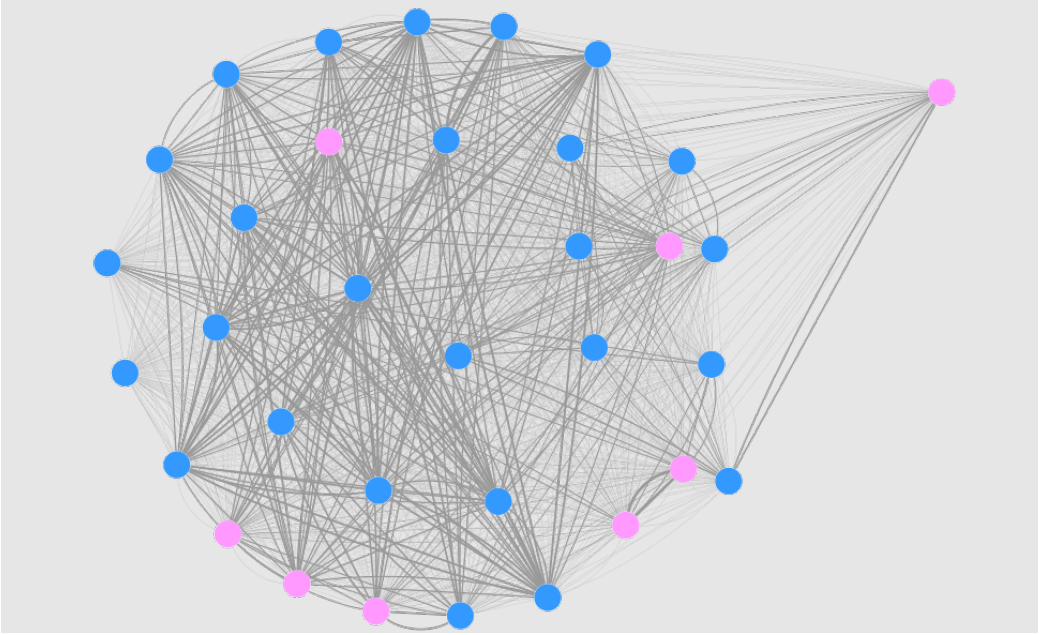

Figure 6. Phage network with prophages from *L. mulieris* strains represented by pink nodes and prophages from *L. jensenii* strains represented by blue nodes. Two nodes (prophages) are connected by an edge if they share homologous genes. The more genes shared between the 2 nodes, the darker the color of the edge is.

Conclusion

Even though the two species are very closely related, this genomic analysis highlights the distinctions between them. There are many genes unique to each species, as well as functions unique to each species. Most notable is the type of biosynthetic clusters found, which may play a role in the deviation and subsequent evolution of these two species

CHAPTER 3

PHENOTYPIC DIFFERENCES BETWEEN THE TWO SPECIES AND DO THEY RELATE TO URINARY SYMPTOMS

Introduction

The human urinary tract varies in temperature, nutrient composition, and pH. The urinary environment temperature ranges from 36.9-39.9°C (Neugent *et al.* 2020) and nutrient composition varies over time due to the voiding of urine. Nevertheless, urine most commonly contains amino acids, electrolytes, and carbohydrates (Forsyth et al., 2018). While the pH of urine collected from healthy individuals is between 5 and 9 (Simerville, Maxted & Pahira, 2005), certain health conditions can alter this pH (Maalouf et al., 2007). *Lactobacillus* species are known to change the pH due to its production of lactic acid, and thus play a role in regulating the pH of the urinary tract and vagina (Gupta et al., 1998; Boskey et al., 1999). In the vaginal tract, the lactic acid can lower the pH to 3 or 4 (Boskey et al., 1999; Ravel et al., 2011).

As previously discussed, *L. jensenii* has been frequently isolated from the urinary tract and the vaginal tract (Mendes-Soares et al., 2014; Thomas-White et al., 2018). *L. jensenii* has been shown to be a probiotic due to its ability to produce lactic acid, thus lowering the pH of its surrounding environment (Gupta et al., 1998; Boskey et al., 1999). Furthermore, prior studies have shown that *L. jensenii* produces H_2O_2 (Antonio, Hawes & Hillier, 1999). While variation has been observed between *L. jensenii* strains (Antonio, Hawes & Hillier, 1999; Hütt et al., 2016), prior studies of Lactobacilli found that the H_2O_2 produced can reduce the growth of

common uropathogens, including *E. coli* and species of *Pseudomonas*, *Proteus*, and *Bacillus* (Price & Lee, 1970; Martín & Suárez, 2010; Hertzberger et al., 2014). Furthermore, prior research has shown that *L. jensenii* can help inhibit sexually transmitted diseases, as well as pathogenic bacteria, and may offer protection against UTIs (Kalyoussef et al., 2012; Yamamoto, Xu & Fichorova, 2013).

In March 2020 a new sister taxon of *L. jensenii* was identified, named *L. mulieris* (Rocha et al., 2020). Rocha et al., showed that there are distinct phenotypic characteristics distinguishing the two species. For example, it was found that *L. mulieris* cannot metabolize ribose and trehalose while *L. jensenii* can (Rocha et al., 2020). A prior meta-analysis found that human urine is dominated by amino acids and carbohydrates (Bouatra et al., 2013). D-Ribose was found in all urines screened and maltose was found in almost all urine samples (Bouatra et al., 2013). *L. crispatus* and *L. iners*, both other colonizers of the female urogenital tract, have been shown to ferment glucose, trehalose, maltose, and mannose (France, Mendes-Soares & Forney, 2016). The ability to metabolize carbohydrates available in the urine is essential for bacterial growth and persistence within this niche.

Prior phenotypic characterization of the *L. mulieris* species included a single representative of the new species, *L. mulieris* c10Ua161M^T, and a single representative of *L. jensenii*, strain DSM 20557^T. Furthermore, this characterization was primarily focused on metabolic processes. We thus focused our phenotypic characterization of these two species on urogenital-relevant aspects, namely pH, H_2O_2 production, and carbohydrate metabolism. Rather than selecting a single representative of each species, multiple isolates from both *L. jensenii* and *L. mulieris* were assayed.

Methods

Carbohydrate metabolic assays

Lactobacillus isolates from our collection (described in Chapter 2 Methods) were streaked onto MRS + 1% Tween 80 agar plates from freezer stocks and incubated overnight at 35°C with 5% CO_2. Colonies were picked from these plates to inoculate 10 mL of MRS + 1% Tween 80 liquid media and grown for 48 hours at 35°C and 5% CO_2. This was performed in triplicate for each strain. After 48 hours, each bacterial culture was pelleted by centrifugation. The spent media was removed, and the pellet was washed with 10 ml of PBS. The pellet was washed again before resuspension in 15 ml SDM basal medium without any carbohydrates (Dextrose: 20 g, Tween 80: 1 g, Ammonium Citrate: 2 g, Sodium acetate buffer solution, pH 5.2 $\pm$ 1 at 25 º C 20.316 ml, $MgSO_4 \bullet 7\ H_2O$: 0.1 g, $MnSO_4$: 0.05 g, K_2HPO_4: 2 g, Yeast nitrogen base (Diftco^{TM}): 5 g, and Bacto Casiton: 10 g in 1L H_2O). This recipe was adapted from (Kimmel & Roberts, 1998). The triplicate samples for each strain were then combined into 50 ml conical tubes and homogenized by vortexing for 30 seconds. 1 ml of the vortexed culture was added to 19 ml of SDM media supplemented with 20 g of one of the four sugars: (1) Trehalose dihydrate (VWR), (2) D-(-)-Ribose, 98% (BTC), (3) Maltose monohydrate (VWR), or (4) D-Glucose (Dextrose) (TEKnova). SDM media with no sugar was used to serve as a control. Each strain had three biological replicates.

For each of these cultures, bacterial growth was measured using a spectrophotometer (wavelength=660 nm) after 0 hours, 24 hours, 48 hours, and 72 hours. The spectrophotometer was calibrated for each of the medias using the SDM media + sugar (or no sugar in the case of

the control) without bacteria. At each time point, 1 mL was removed from the culture for measurement. All measurements were recorded for statistical analyses.

Sugar statistical analysis

For each time point, *L. jensenii* and *L. mulieris* strain measurements, which were conducted in triplicate (biological replicates), were considered independent replicates of their respective species. While this was mainly done for convenience, we also verified statistically that there was no significant difference between the different strains of a species. Exploratory data analysis (EDA) in R was used to visualize our data. We then performed a repeated measures analysis of variance (ANOVA) (Montgomery, 2013) where the factors were Species (two levels: *L. mulieris*, *L. jensenii*) and Sugar (five levels: each sugar and the no sugar control), over the four time points (0 h, 24 h, 48 h and 72 h). This was followed by two-way ANOVA in R to identify further significant differences.

***Lactobacillus* Hydrogen Peroxide Measurements**

Twenty-two *L. jensenii* and 14 *L. mulieris* strains were grown from freezer stocks in 10 mL of MRS + 1% Tween 80 for 48 h at 35°C in an incubator with 5% CO_2. The cultures were centrifuged for 10 minutes. 2 mL of the spent media was removed and placed in a microcentrifuge tube and filtrated using a sterile 0.22 um Cellulose Acetate syringe filter (Restek). The pH of each sample was determined using pH paper strips and adjusted to range between pH 7 and 8 via 1M NaOH and 3M HCl. This step is required by the Amplex Red Hydrogen Peroxide/ Peroxidase Assay Kit (Thermo Fisher Scientific) which was used to measure the H_2O_2 of each culture. The Amplex Red Hydrogen Peroxide/ Peroxidase Assay Kit manufacturer's instructions were used to measure H_2O_2 in each sample. For each sample, three

biological replicates were assayed, and for each biological replicate, a technical replicate was measured. The absorbance was measured using a plate reader (560 nm).

Lactobacillus pH measurements

Twenty-three *L. jensenii* and 13 *L. mulieris* strains were grown from freezer stocks streaked on MRS + 1% Tween 80 agar plates overnight at 35°C and 5% CO_2. Afterwards, colonies were picked to be grown in 10 mL of MRS + 1% Tween 80 liquid media for 48 h at 35°C and 5% CO_2. Each strain was grown in triplicate. A pH meter was calibrated using buffers at pH 4, 7 and 10 prior to measurement.

Secondary metabolite screening

Lactobacillus strains were grown from freezer stocks streaked on MRS + 1% Tween 80 agar plates overnight at 37°C and 5% CO_2. Afterwards, colonies were picked to be grown in 10 mL of MRS + 1% Tween 80 liquid media for 48 h at 37°C and 5% CO_2. The culture was filtrated using a sterile 0.22 um Cellulose Acetate syringe filter (Restek). 10 ul of the filtrate was then spotted onto bacterial lawns. The bacterial lawns were made using the following bacteria: *P. aeruginosa* ATCC 15692, *E. coli* UMB8140, *E. coli* UMB8289, and *E. coli* C. The two UMB strains were obtained from prior IRB-approved studies (Price et al., 2020b). *E. coli* C and *P. aeruginosa* ATCC 15692 were obtained from the ATCC collection. All four of these bacterial strains were grown from our freezer collection, inoculating Lysogeny Broth (LB) with a loop of bacteria which was then grown overnight at 37 ºC. 1 ml of culture was added to 3 ml of LB soft agar (0.7% agar), poured onto an LB agar plate (1.7% agar), and left to dry before the 10 ul spot was applied. These plates were incubated overnight and were assessed the next day to see if a clearing in the plate had formed where the spot had been applied.

pH and H_2O_2 statistical analysis

L. jensenii and *L. mulieris* strain measurements, all of which included biological triplicates, were treated as independent replicates of their respective species. While this was mainly done for convenience, we also verified statistically that there was no significant difference between the different strains of a species. The H_2O_2 measurements also included two technical replicates of each sample. We performed one-way ANOVA for both datasets (pH and hydrogen peroxide) in order to identify significant difference between the two species followed by Fisher's Least Significant difference (LSD) test to quantify the significant difference (if any) between species. These analysis were done in R .

Results and Discussion

pH measurements

Thirty-six strains of *L. jensenii* and *L. mulieris* from the urinary tract were grown and their effect on the pH in their surrounding media was assayed. In total, 23 *L. jensenii* strains and 13 *L. mulieris* strains were examined and three biological replicates were performed for each. Overall pH values varied between 3.16 and 5.81. In order to evaluate putative differences between pH effects of the two species, we next considered the difference in pH measurements. An ANOVA revealed no significant difference between the two species (p-value = 0.2812). The 95% confidence interval between *L. jensenii* and *L. mulieris* ranges from -0.0843 to 0.2874, clearly including zero and therefore proving no significant difference at alpha = 0.05. The mean pH for each species can be seen in Figure 8. Based upon the samples investigated here, there is no discernable difference in the pH effect of these two species.

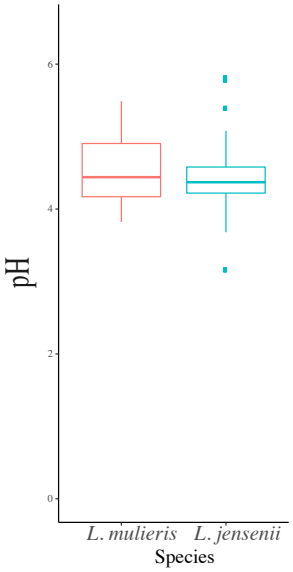

Figure 7. pH measurements for *L. jensenii* and *L. mulieris* urinary strains. The species designation is on the x-axis and the pH is on the y-axis.

Previously, it was shown that the pH of urine from healthy individuals is between 5 and 9 (Simerville, Maxted & Pahira, 2005). *Lactobacillus* is known to produce lactic acid and it was indicative that it lowers the pH of the urine (Gupta et al., 1998; Boskey et al., 1999). Its production of lactic acid is believed to prevent colonization of pathogens in the urogenital tract (Boskey et al., 2001). Our analysis showed that the average pH of 4.48 was lower than prior measurements for voided urine. It is important to note that the strains were grown in MRS + 1% Tween 80 media which is a very different environment with different nutrients than the urinary tract. Even though there is no significant difference between the two species, there is a detectable change in the pH of its surrounding environment. The pH of MRS media is around 6.2 ± 0.2, according to the manufacturer (see Methods). The observed increased acidity of the environment suggests that both *L. jensenii* and *L. mulieris* have the potential to inhibit pathogenic species.

H_2O_2 production

The H_2O_2 production of each species also was analyzed. Here 36 strains were examined: 23 *L. jensenii* strains and 13 *L. mulieris* strains. Each strain was assayed with three biological replicates, and two technical replicates were performed for each biological replicate when measuring H_2O_2 levels. An ANOVA analysis revealed a statistically significant difference between the H_2O_2 production of these two species (p-value = 0.0153). The 95% confidence interval between *L. jensenii* and *L. mulieris* ranges from 0.0304 to 0.2830 which indicates that *L. mulieris* produces more H_2O_2 compared to *L. jensenii*, but the difference between the two is small (Figure 9). One of the biological replicates of *L. jensenii* strain UMB8651 was an outlier to this trend, with measured H_2O_2 production levels > 2.

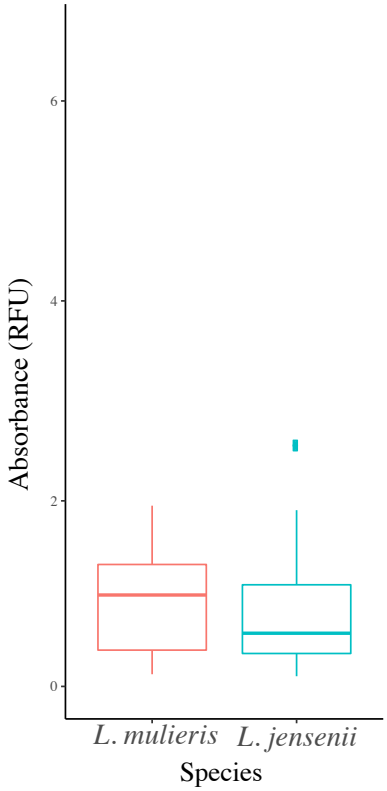

Figure 8. H_2O_2 measurements for *L. jensenii* and *L. mulieris* urinary strains. The species are labeled on the x-axis while the absorbance units are on the y-axis.

Production of H_2O_2 has previously been observed for *L. jensenii* strains (Antonio, Hawes & Hillier, 1999). Our results confirm this prior study and show that that *L. mulieris* strains are also H_2O_2 producers. Furthermore, we find that the difference in H_2O_2 between the two species is statistically significant. The mean measured H_2O_2 RFU for *L. jensenii* strains is 0.667, while the mean for the *L. mulieris* strains is 0.823. Prior studies (conducted before the characterization of *L. mulieris*) noted the variation between H_2O_2 produced by *L. jensenii* strains (Wilks et al., 2004). This observation may be the result of assessing both *L. jensenii* and *L. mulieris* strains. In a recent study, a urinary *L. jensenii* strain was shown to produce up to 10 times more H_2O_2 than another urinary bacteria, *Streptococcus mitis,* which is known to inhibit pathogen growth in the

urogenital tract (Mores et al., 2021). This further suggests that both *L. jensenii* and *L. mulieris* have the potential through H_2O_2 production to inhibit pathogenic growth including UPEC strains.

Secondary metabolites

Our genomic screening of the *L. jensenii* and *L. mulieris* strains identified RiPPs, NRPS, and Class IV lanthipeptides (Chapter 2). This led me to investigate their putative effect on uropathogens. To do this 33 strains of *L. jensenii* and *L. mulieris* were cultured and filtered, and their filtrate was tested for its inhibitory potential against the uropathogenic species *P. aeruginosa* (n=1) and *E. coli* (n=3). Strains containing each of the prdicted biosyntehtic clusters, as well as strains with no predicted biosynthetic clusters, were tested (Table 5), but no effect on any of the uropathogenic species was observed. Previous work has showed that the cell-free supernatants of *L. jensenii* strains were bacteriostatic towards other UPEC strains (Mores et al., 2021). While this could be due to their lactic acid production, hydrogen peroxide production, or secondary metabolite production, the 34 different strains tested here did not show signs of inhibiting *P. aeruginosa* or *E. coli* growth.

Table 5. Number of strains tested for each class of secondary metabolites.

	NRPS	lanthipeptide-class-iv	No Hits	RiPP-like
L. mulieris	0	7	2	2
L. jensenii	20	0	2	0

Sugar metabolism analysis

Forty strains from our lab collection of *L. jensenii* (n=24) and *L. mulieris* (n=13) were grown and their ability to metabolize four different sugars, ribose, maltose, glucose, and trehalose, was tested alongside media with no sugar (control). Each strain was tested with three

biological replicates and growth was measured at four different time points (0, 24, 48, and 72 hours). Therefore, the measurements were stopped at 72h.

Figure 10 presents the results of these measurements. Each time point is represented for each of the four sugars tested and the control. Here we have divided the results for the *L. jensenii* strains (right panel) and the *L. mulieris* strains (left panel). First, we anticipated and observed no substantial growth for the "no sugar" group, which served as a control. Next, as the exponential growth phase for the species concludes around 72 hours of growth, we anticipated growth to plateau or decline at this measurement. We observed a mean decrease in growth for *L. jensenii* strains cultured in the presence of trehalose and glucose. A reduced growth rate is detected for *L. mulieris* strains cultured in the presence of maltose and glucose and *L. jensenii* strains cultured in the presence of maltose.

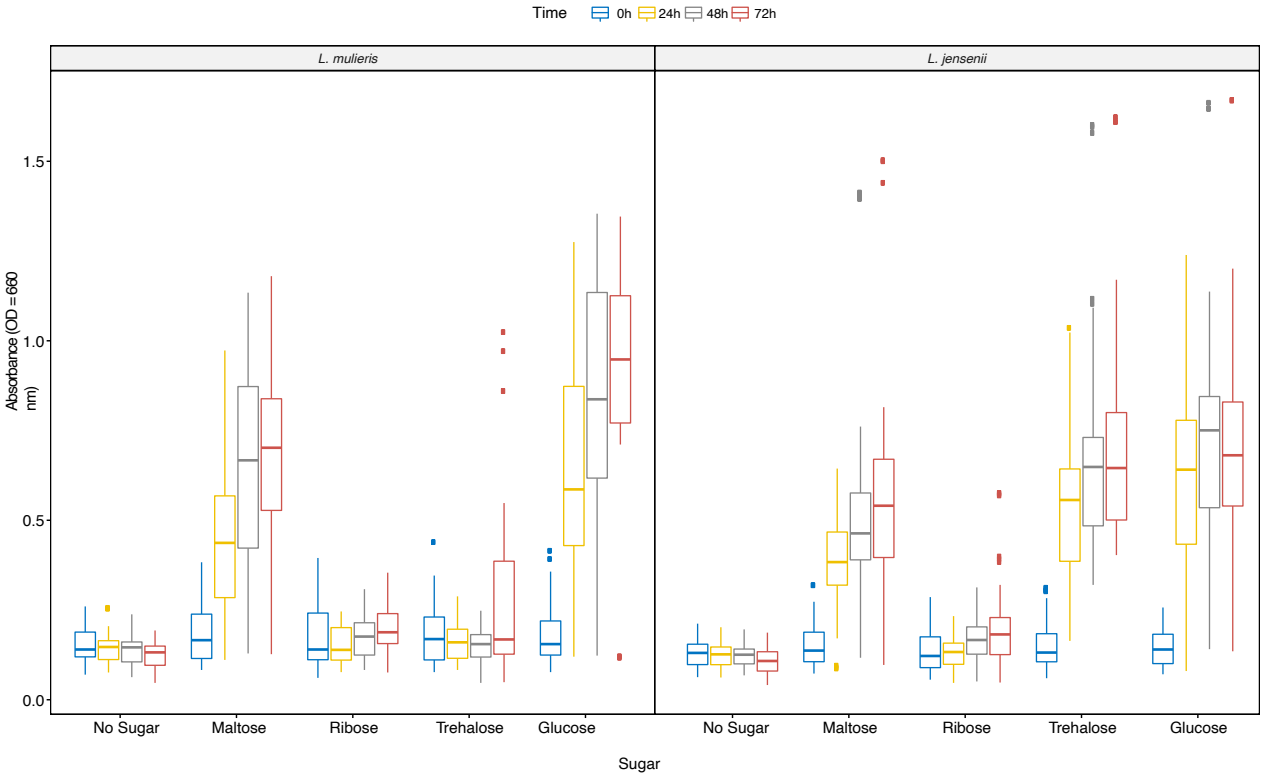

Figure 9. Carbohydrates analysis. On the left side *L. mulieris* is shown while on the right *L. jensenii*. On the x-axis each sugar is shown, and the different box plot represents each sugar growth at that time point. Each box plot is a different time point for all the species in the analysis.

Our assays suggest that L. jensenii strains can efficiently metabolize maltose, trehalose, and glucose, but not ribose. L. jensenii was previously shown to metabolize ribose (Rocha et al., 2020); however, the study assessed just a single strain of L. jensenii from the vagina: L. jensenii DSM 20557. There was also no information provided on how L. jensenii DSM 20557 was grown and what time points were used in the analysis. Our strains are clinical isolates from the urinary tract of females with or without lower urinary tract symptoms. We conducted a pairwise comparison of the measurements for L. jensenii strains from the control treatment (no sugar) and ribose treatment. At time 0h and 24h, there is no significant difference between the growth of the L. jensenii strains. However, when analyzing L. jensenii at time point 48h there is a statistically significant difference

between (p-value=4.38E-06) as well as time point 72h (p-value=1.10E-08).his indicates that L. jensenii is able to metabolize ribose, but not efficiently. It is worth noting that this is direct contrast to pairwise comparisons of the other sugar treatments to the control treatment; above timepoint 48h all strains were efficient in their metabolism of carbohydrates for L. *jensenii* strains. In comparison to maltose, trehalose, and glucose, the *L. jensenii* strains cannot metabolize ribose nearly as well.

Similar pairwise comparisons were conducted for L. mulieris strains grown with no sugars (control) and trehalose and ribose treatments. Again, there was no statistically significant difference between the control and treatment groups for 0h and 24h. However, at time point 48h and 72h there was a statistically significant difference between the no sugar and ribose treatment (p-value= 0.000182 (48h) and p-value = 1.01E-08 (72h)), as well as no sugar and trehalose treatment (p-value= 0.026 (48h) and p-value = 1.01E-08 (72h)). These findings suggest that the *L. mulieris* strains cannot efficiently metabolize ribose or trehalose. In the phenotypic characterization of *L. mulieris*, the species was characterized by its inability to metabolize ribose and trehalose (Rocha et al., 2020). We find this to be true for < 48h measurements. Certainly, in comparison to other sugars tested (Figure 10), the 13 strains cannot metabolize trehalose or ribose as efficiently as maltose or glucose. Figure 10 also includes outliers to these general trends. For *L. mulieris*, there are outliers for trehalose at 0h and 72h. At timepoint 0h, the outlier is one of the biological replicates for UMB0047, which comes from a no LUTS (control) patient. Measurements for this line at subsequent timepoints, however, do not deviate from the measurements from other strains. At timepoint 72h, the three outlier points are the 3 biological replicates of UMB4707, a clinical

isolated from an OAB+ patient. The increased growth of this strain at 72h suggests that this strain can now metabolize trehalose.

Outliers are observed for the *L. jensenii* strains for growth supplemented with maltose, trehalose, and glucose; these include the highest growth rates observed (Figure 10). The outliers of *L. jensenii* for maltose timepoint 0h are the biological replicates of UMB0055, a strain from a female with no LUTS. At timepoint 48h and 72h, the outliers are the three replicates of UMB1303, a clinical isolated from a patient with UTI. In a prior study, carbohydrate levels in urine were linked to diet and sex of an individual (Mack et al., 2018). For example, maltose is present in low concentrations in urine. However, the vaginal microbiota can degrade α-amylase activity and breakdown glycogen to maltose. This degradation is due to the presence of *Lactobacillus* species (Spear, 2014; Spear et al., 2014; Nunn & Forney, 2016).

Other outliers of *L. jensenii* can be found in trehalose at timepoint 48h and 72h. These are again UMB1303 at both timepoints. Upon further inspection of UMB1303, it grew best in all conditions of carbohydrate including ribose and glucose at all time points. Future investigations of this strain can include metabolic pathway analysis of the genome as well as testing this strain under various concentrations of these sugars and/or additional timepoint measurements.

Conclusion

While the pH effects of *L. jensenii* and *L. mulieris* strains are not statistically significant. H_2O_2 production, which can inhibit uropathogen growth, is statistically significantly different between the two species. This may have implications for urinary health. *L. mulieris* is a promising species to study for future studies of maintaining a "healthy" urinary microbiome or directly modulating the urinary microbiome. Our phenotypic characterization of these species found discord with prior work, namely the ability (or inability) for *L. jensenii* to efficiently metabolize ribose (Rocha et al., 2020). Further investigation into the metabolic pathways of strains of these two species will be illuminating with regards to gaining insight into how Lactobacilli persist in the urinary tract.

CHAPTER 4

PREVALENCE OF L. JENSENII AND L. MULIERIS IN THE URINARY MICROBIOTA OF FEMALES WITH AND WITHOUT SYMPTOMS.

Introduction

Profiling metagenomic communities using the 16S rRNA gene marker has been an essential tool for exploring and profiling microbial communities (Woo et al., 2008). It was instrumental in early surveys of the human microbiome prokaryotic diversity as part of the first stages of the Human Microbiome Project (NIH HMP Working Group et al., 2009). The 16S rRNA gene marker is a gene that is shared across all prokaryotes and is conserved within species. Hypervariable regions within this gene provide resolution such that taxonomic identification can be made for bacterial genera and in many cases species-level designations. As such, individual hypervariable regions are targeted by 16S rRNA amplicon surveys as they are informative with regards to taxonomic classification and have lengths that can be sequenced by high-throughput short-read sequencing technologies. Studies that characterize whole communities based on the 16S rRNA hypervariable gene regions provide a cost-effective means of identifying organisms within complex communities including those that are difficult or impossible to grow in laboratory settings (The Human Microbiome Project Consortium, 2012).

The most common way of profiling a microbial community using the 16S rRNA gene is by identifying operational taxonomic units (OTUs) with a de facto threshold for clustering of 97% (Chiarello et al., 2022). However, it is well-known that using the 97% clustering threshold for OTUs cannot provide a comprehensive or accurate image of a microbiome at the species level (Edgar, 2018). Despite this limitation, genus-level taxonomic classifications are often possible. OTUs are widely used in the characterization of the urinary microbiome (Siddiqui et al., 2011, 2012; Fouts et al., 2012; Lewis et al., 2013; Pearce et al., 2014a,b; Karstens et al., 2016; Wu et al., 2017; Thomas-White et al., 2017; Gottschick et al., 2017; Chen et al., 2018; Zeng et al., 2020; Price et al., 2020a; Komesu et al., 2020; Xu et al., 2021; Anglim et al., 2021; Richter et al., 2022; Li et al., 2022).

An alternative approach for profiling microbial communities is by amplicon sequence variants (ASVs). ASVs follow denoising algorithms which produce clusters with 100% sequence similarity; however, research has shown that these clusters are not always 100% sequence similarity (Callahan, McMurdie & Holmes, 2017). Nevertheless, ASVs have a higher sensitivity in identifying different strains, but they are not perfect and should not be used to the species level (Chiarello et al., 2022). They are accurate in giving a genus level approach of a metagenome.

16s rRNA surveys have widely been used given the cost effectiveness possible with current high-throughput sequencing technologies. Because these short-read sequencing technologies can only sequence 500-600 bp (using the paired-end approach), the entire 16S rRNA gene cannot be sequenced (Pace, 1997) . Thus, variable regions of the 16S rRNA gene sequence are targeted. The V3-V4 region of the 16S rRNA gene is the most sequenced region for metagenomic studies of the urinary and vaginal microbiomes (Hugerth et al., 2020; Hoffman et

al., 2021a). Critical to these studies is the reference database used to make taxonomic calls based upon the variable region sequences; for the urinary microbiota, the NCBI database and SILVA database have been found to perform best (Hoffman et al., 2021b).

Several studies have suggested that *L. jensenii* is abundant in microbial communities of the urinary tract and vagina (Komesu et al., 2020; Tortelli et al., 2020; Mehta et al., 2020; Witkin et al., 2021). Most of the studies are based on 16S rRNA sequencing targeting the V4, V1-V3, V1-V5, or V1-V4 regions and OTU clustering (Komesu et al., 2020; Tortelli et al., 2020; Mehta et al., 2020; Witkin et al., 2021). Very few studies have looked at the species level of the vaginal tract from the V3-V4 16S rRNA variable region and found *L. jensenii* (Dunlop et al., 2021). It is important to note that these studies were conducted prior to the characterization of *L. mulieris*. As my prior investigation of these two species revealed (Chapter 2), the full-length *L. jensenii* and *L. mulieris* 16S rRNA sequence differs only by two nucleotide positions. Neither of these nucleotide positions are in variable regions. This prompted my investigation into the viability of using the V4 region, the preferred region to target for urinary microbiome studies (Hoffman et al., 2021b), for classification of *L. jensenii* and *L. mulieris*. The results of this inquiry incited subsequent work in identifying species-specific markers for *L. jensenii* and *L. mulieris* identification in metagenomic sequencing.

Methods

Lactobacillus 16S rRNA gene marker phylogenetic tree

On February 1st, 2022, all *Lactobacillus* 16S rRNA gene sequences were retrieved from the SILVA database version 138.1 (Quast et al., 2013). Partial sequences were removed from the data set. This included the removal of several partial *L. jensenii* sequences leaving 11 full length *L. jensenii* sequences. Using Geneious Prime (Biomatters Ltd., Auckland, NZ), the retrieved sequences were trimmed to only contain the V4 region, the region targeted by urinary and vaginal microbiome studies (Hugerth et al., 2020; Hoffman et al., 2021a). The sequences were clustered using usearch v11 (Edgar, 2010) with a 97% threshold. Each cluster was manually inspected. The 16S rRNA sequences from SILVA and the 16S rRNA gene sequences from the genomes of the type strains *L. jensenii* ATCC 25258 (GCA_018094625.1) and *L. mulieris* c10Ua161M (GCA_007095465.1) were aligned using the MAFFT v7.388 (Katoh & Standley, 2013) plug-in through Geneious Prime.

Divergent genes

The core genes of the 61 *L. jensenii* and *L. mulieris* genomes analyzed in Chapter 2 were further investigated here. A script was created to identify genes in which single nucleotide polymorphisms (SNPs) can accurately distinguish between strains of the two species. The Levenstein distance package (https://github.com/ztane/python-Levenshtein/) was used in python to compute the similarity/distance between two gene sequences. For each core gene, a matrix was created of the Levenstein ratios for each pairwise genome comparison. Each gene was then clustered with k clusters = 2 using the Biopython kcluster package. If the 2 clusters consisted of

strains that corresponded with the species designation, then that gene was considered a core divergent gene.

Phylogenetic tree

Sequences were aligned using the MAFFT v7.388 (Katoh & Standley, 2013) plug-in through Geneious Prime. The phylogenetic tree was derived using the FastTree 2.1.12 (Price, Dehal & Arkin, 2010) plug-in through Geneious Prime and visualized using iTOL v6 (Letunic & Bork, 2016).

Primer Design

Primers were designed for select divergent core genes using Primer3 (https://bioinfo.ut.ee/primer3-0.4.0/) (Untergasser et al., 2012) and Primer-BLAST (Ye et al., 2012). Self-dimerization was checked using the ThermoFisher Multiple Primer Analyzer tool (https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html). Primers were synthesized by Eurofins MWG.

Screening for *L. mulieris* and *L. jensenii*

Serving as a control for the targeted PCR, DNA extracted from *L. jensenii* and *L. mulieris* strains in our collection (described in Chapter 3 Methods) was amplified for each of the 4 primer pairs tested (Table 6) using the following conditions: 25 uL of Go-Taq Master Mix (Promega, Madison, WI), 23 uL of nuclease-free H_2O, 1uL of each primer (10mM), and 5 uL of extracted bacterial DNA. Nuclease-free water was used as a negative control. PCR amplification was verified using a 1.2% agarose gel. Amplicons were cleaned using the E.Z.N.A. Cyle Pure Kit (Omega Bio-Tek) and sequenced via Sanger sequencing (Genewiz, South Plainfield, NJ).

Separate sequencing reactions were prepared for the forward and reverse primers. Sequence results in FASTA format were assembled using Geneious Prime (Biomatters Ltd., Auckland, NZ). The sequences were aligned via MAFTT and the differences in their alignments were identified.

To assess the specificity of these primers relative to other Lactobacilli of the urinary tract and the vagina, the primers were tested against strains of other *Lactobacillus* species. The other tested strains are listed in Table 9, and they were collected as part of prior IRB-approved studies (Loyola University Chicago, 206469, 204133, and 207102) (Hilt et al., 2014; Pearce et al., 2014b,b; Thomas-White et al., 2016b; Price et al., 2020b).

Table 6. Primers for *L. jensenii* and *L. mulieris* detection

Primer Pair Name	Right sequence	Left Sequence
341_p1	TATTCKTGGRTAACCAACTTTTTCAC	ACGACTGTAGGTTGGGCAAC
341_p2	CCTGCTTCRAATTGTCTAACAAAAC	TGGGACAYCATTTATCRAATTCTC
137	CATCSGTATCAACATATACTTTGATATCCATT	GGGATTGCTGGTGGTTCTG
105	ACACCACACTTGTCATCCCC	ACGTGAAGTTGATTTCCAAATG

Urine collection and extraction

Urine was collected as part of prior IRB-approved studies (Loyola University Chicago, 207102, 204195, 206449, 207152, 209545, and 204133). For all urine specimens examined here, 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 was then stored at -80°C.

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 500 uL of urine (rather than 1.75 mL) and adjusted the volume of the Binding Solution used accordingly. DNA concentration was quantified using a Qubit fluorometer.

Illumina Primer sequencing

One set of primers from the isolate screening was selected for metagenomic sequencing. The partial Illumina adapter sequences were added to the 5' end. The forward sequencing read adapter sequence was 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'. The reverse sequencing read adapter sequence was 5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'. These read adapters were added to our primer sequence 341_2 from Table 20. The primers were synthesized by Genewiz and Eurofins. Using these primers, the DNA extracted from the urine samples were screened for the presence of *L. jensenii* and *L. mulieris*. The following PCR reaction was used: 1 uL forward primer (10 mM), 1 uL reverse primer (10 mM), Go-Taq Master Mix (Promega, Madison, WI), 18 uL nuclease-free H_2O. *L. jensenii* and *L. mulieris* isolate DNA was used as a positive control; nuclease-free water was used as a negative control.

Metagenomic screening of samples

Metagenomic sequences from 233 publicly available urinary microbiome datasets were retrieved from NCBI. The SRA database accession numbers can be found in Appendix II. Raw sequencing reads for each dataset were mapped to the select divergent core genes using bowtie2 (version 2.3.2) via the Geneious Prime (Biomatters Ltd., Auckland, NZ) plug-in (Langmead & Salzberg, 2012).

Results and Discussion

16S rRNA gene sequence analysis of *Lactobacilli*

3,721 full length 16S rRNA sequences from the *Lactobacillus* genus were retrieved from the SILVA database (Quast et al., 2013). This set included 11 sequences from *L. jensenii* strains. The 3,721 sequences were trimmed to contain only the V4 region, and their sequences were clustered with a 97% identity threshold. After clustering, 273 clusters were identified, and the 11 *L. jensenii* sequences were divided into two different clusters. One cluster had only three *L. jensenii* sequences, which are 100% identical to each other. The other cluster had eight *L. jensenii* sequences and 164 other sequences (Figure 11). These 164 other sequences included strains of *L. delbrueckii* and *L. delbruckii* subspecies (subspecies: *lactis, sunkii, bulgaricus, jakobsenii, delbrueckii, Indicus, Bulgaricus*) as well as strains from the species *L. porci, L. equicursoris, L. psittaci, L. acidophilus, L. fornicalis, L. gigeriorum,* and *L. kalixensis*. Additionally, the cluster included strains for which no species designation has been made, including *L. sp.* DJF_CR11*, L. sp.* RA2113*, L. sp.* AB5231*, L. sp.* AB5262*, L. sp.* MW503*, L. sp.* AB51159. In total, 21 other species were found in this cluster with *L. jensenii*. The V4 region sequences within this cluster were aligned producing a 292 bp alignment with 247 identical sites (80.7%). The average pairwise sequence identity is 99.3%. Further examination revealed that there are only 10 unique sequences within this cluster (Table 7), with 142 of the sequences being identical to one another. These 142 sequences, which forms the biggest clade in Figure 11, are all *L. delbrueckii* strains, including several different subspecies.

Table 7. Number of sequences identical

Size of Clade	*Dominant Species Designation*

142	*L. delbrueckii*
17	*Lactobacillus delbrueckii*
5	*Lactobacillus delbrueckii*
4	*Lactobacillus delbrueckii*
3	*Lactobacillus delbrueckii*
2	Lactobacillus kalixensis
1	Lactobacillus gigeriorum
4	*Lactobacillus delbrueckii*
10	Lactobacillus jensenii
1	*Lactobacillus acidophilus*

Even though the studies that investigated the prevalence of L. jensenii in the urinary microbiome characterized it based on 16S rRNA OTU clustering (Komesu et al., 2020; Tortelli et al., 2020; Mehta et al., 2020; Witkin et al., 2021), I decided to also look at the ASV clustering of the *Lactobacillus* 16S rRNA V4 region. A phylogenetic tree was created from the cluster that contained the 8 *L. jensenii* strains and other *Lactobacilli*. Each leaf of the phylogenetic tree can be considered an ASV classification. As it can be seen in Figure 11, there is a clade that has all 8 *L. jensenii* sequences and an *L. fornicalis* strain are identical. Therefore, this shows that there is one ASV with *L. jensenii* and *L. fornicalis* clustered together.

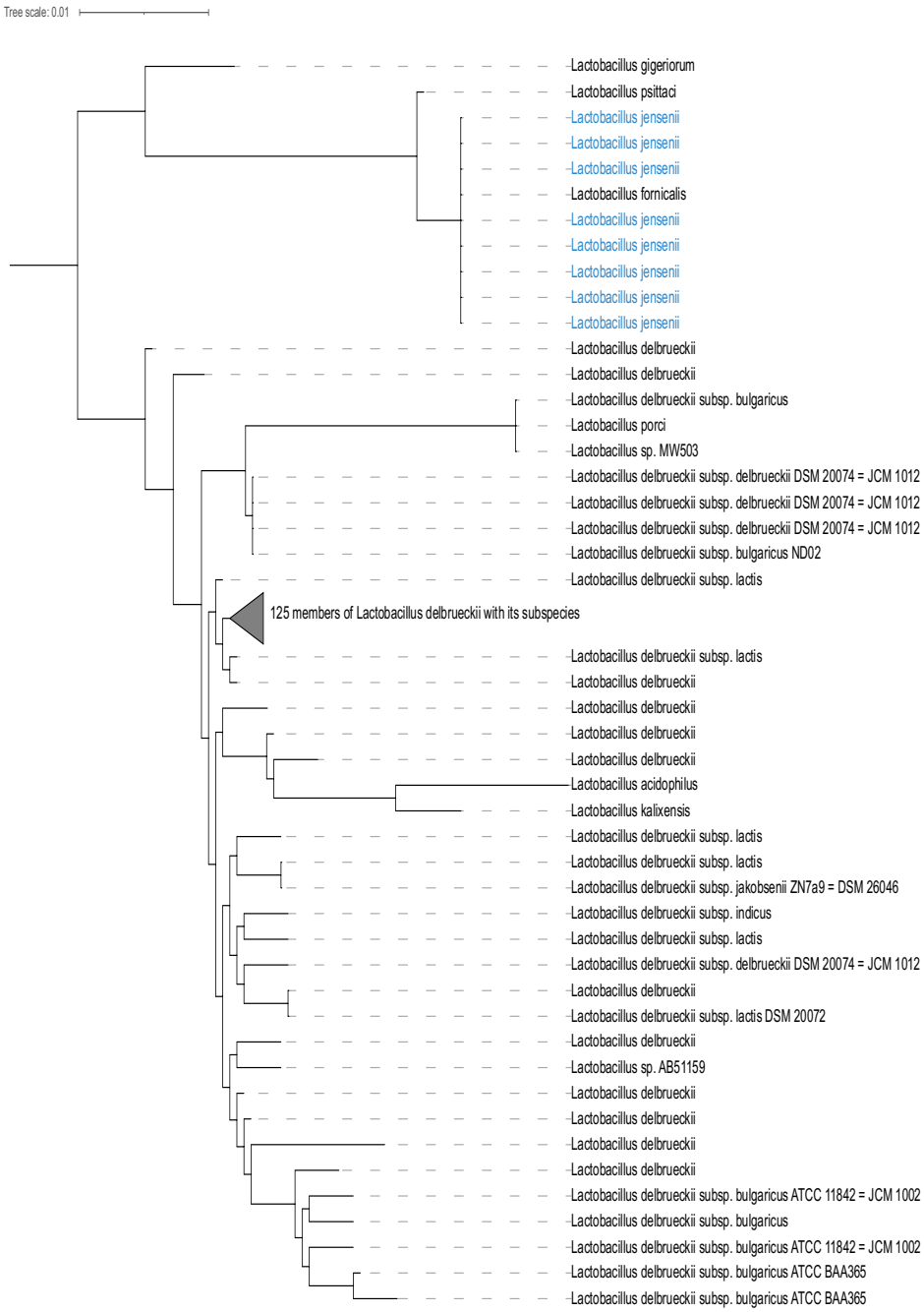

Figure 10. Phylogenetic tree represented ASVs

While the SILVA database does not contain a representative of *L. mulieris*, our prior analysis of the full length 16S rRNA gene sequences of *L. jensenii* and *L. mulieris* found that the two species cannot be distinguished by the V4 region (Chapter 2). The V4 region of the *L. jensenii* and *L. mulieris* strains examined in Chapter 2 are identical to the OTU cluster that contained *L. jensenii* and other Lactobacilli. Thus, our analysis of the V4 region using both the OTU and ASV methods finds that *L. jensenii* cannot be distinguished from other Lactobacilli. Furthermore, the *L. jensenii* and *L. mulieris* strain V4 sequences are identical to the *L. fornicalis* strain the SILVA database. The type strain of the species *L. fornicalis* was isolated from the vagina and thus could be present in the urinary tract as well (Dicks et al., 2000). This is concerning as the SILVA database is routinely used for human microbiome studies, including those of the urinary tract (Brubaker et al., 2021).

Next, I aligned the 11 full length 16S rRNA gene sequences of *L. jensenii* from SILVA and the 16S rRNA gene sequences of the type strains of *L. jensenii* ATCC 25258 and *L. mulieris* c10Ua161M. Figure 12 shows the resulting phylogenetic tree. The full length 16S rRNA gene sequences of the *L. jensenii* and *L. mulieris* type strains are most similar to the large OTU cluster that contained *L. jensenii* and other *Lactobacillus* species, shown in black font. Thus, the full length *L. jensenii* and *L. mulieris* 16S rRNA gene sequence would not be able to be distinguishable from other Lactobacilli in the SILVA database using the 97% threshold that is routinely used for OTU-based analyses.

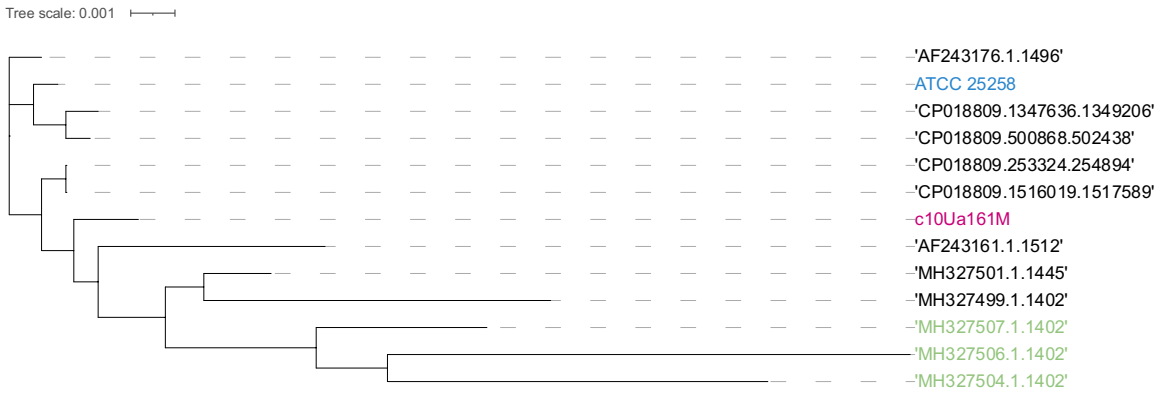

Figure 11. Phylogenetic tree of the 11 *L. jensenii* sequences from SILVA and the type strains of *L. jensenii* (in blue) and *L. mulieris* (in pink). The 3 sequences in green font were clustered as a separate OTU from the remaining SILVA *L. jensenii* sequences (shown in black font).

Divergent Core Genes

Given the limitations of the 16S rRNA gene sequence to accurately and unequivocally detect and discern between *L. jensenii* and *L. mulieris* and other Lactobacilli, I next developed novel gene markers for these two species. From the core genes of *L. jensenii* and *L. mulieris*, I identified 659 genes belonging to the "divergent core gene set". These genes are conserved in all 61 *L. jensenii* and *L. mulieris* strains (Chapter 2), and the sequences of these genes differ between *L. jensenii* and *L. mulieris* strains such that SNPs can be used to distinguish between the two species. Gene cluster GC_00000130, which is the 3-5 helicase subunit RecB of the DNA repair enzyme RecBCD (exonuclease V) (RecB) (PDB:3U44) (PUBMED:29901759;30601118) based on its COG annotation, is the most divergent gene sequence with 1,675 nucleotide differences between the two species. The second most divergent gene cluster is GC_00000118, annotated as an Isoleucyl-tRNA synthetase (IleS) (PDB:1FFY).

From the 659 divergent core genes, I chose three to test as gene markers. I looked at gene clusters that are relatively divergent (around 600 nucleotide differences) and those that are not found in other Lactobacillus species, assessed by querying their sequence against the nt/nr NCBI database. GC_00000105 is a Di- or tripeptidase (PepD2) (PDB:2QYV) with a length of 1239 bp. When queried against the NCBI nr/nt database, the only hit found is to an L. jensenii sequence (CP018809.1) with a 98.8% identity. To further verify that this gene can be a good marker gene to distinguish between the two species, the gene sequences from the 61 genomes were aligned and a phylogenetic tree was created (Figure 13). The tree shows two distinct clusters, one containing L. jensenii strains and the second containing L. mulieris strains. It is worth noting that the three strains in the L. mulieris subgroup that were previously identified in the ANI and core genome analysis (Chapter 2) form a subclade here as well.

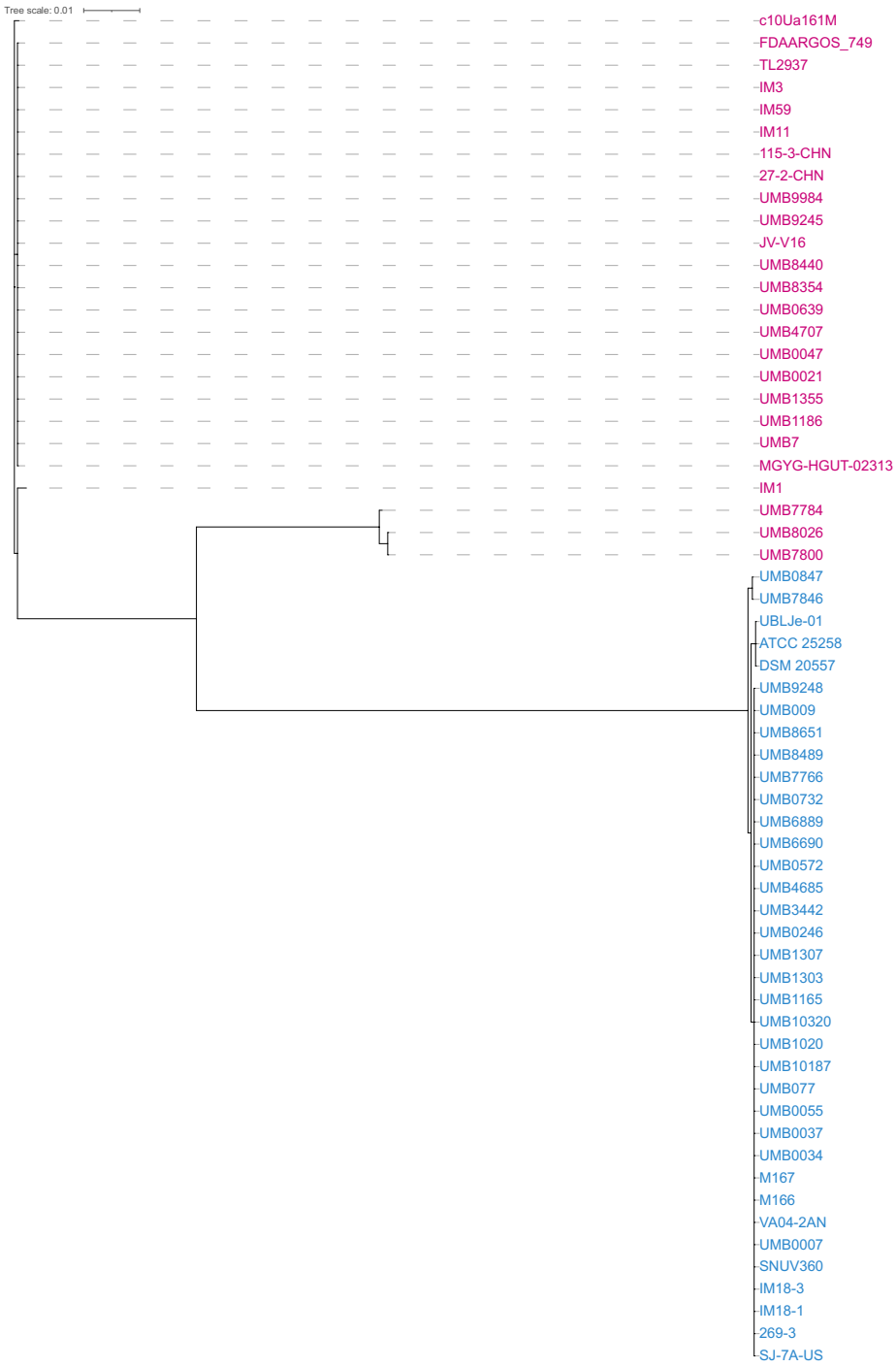

Figure 12. GC_00000105 phylogenetic tree. L. mulieris species are in pink while L. jensenii species are in blue.

The next core divergent gene that I chose to look at is an GC_00000341, namely Uncharacterized membrane protein YjjP, DUF1212 family with a length of 1,340 bp. This gene had, at most, 982 nucleotide differences between strains of the two species. Upon querying the gene sequence to the NCBI nr/nt database, only sequence similarity to an *L. jensenii* record was identified. The gene cluster's sequences were aligned, and a phylogenetic tree was created (Fig 14). Again the two species clade independently and the *L. mulieris*' subgroup can be found independently.

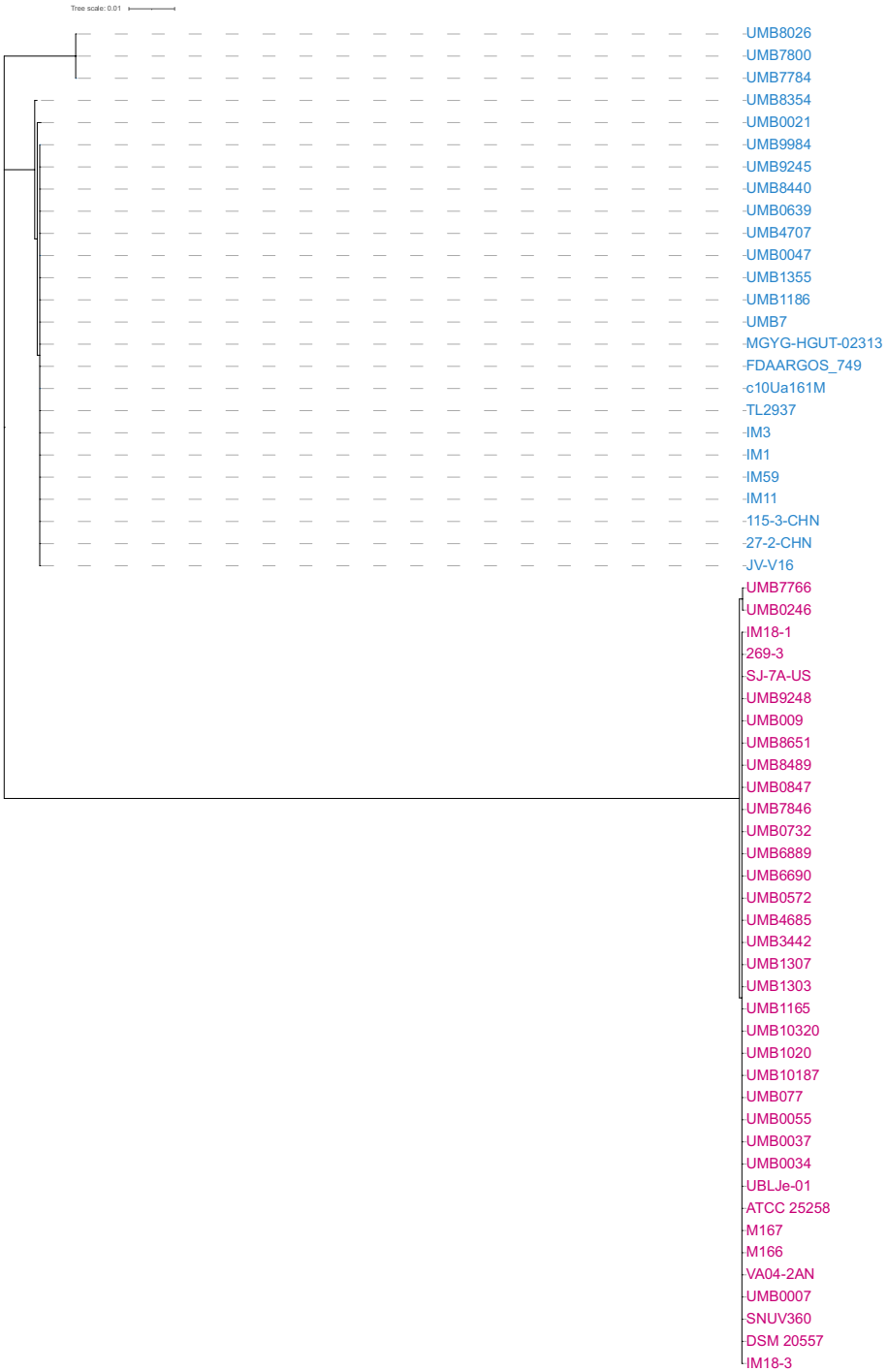

Figure 13. GC_00000341 phylogenetic tree. *L. mulieris* species are in pink while *L. jensenii* species are in blue.

I also chose to look at a core gene that the clustering method did not detect as able todistinguish between the two species. Based on clustering, it was almost perfect but had one L. mulieris strain erroneously cluster in the L. jensenii cluster. The gene cluster was GC_00000137 and its COG annotation was Uridine kinase (Udk) (PDB:5B3F) with homologous hits only to L. jensenii strains. The gene cluster sequences were retrieved, aligned and a phylogenetic tree was created. As shown in Figure 15, the phylogentic analysis of these homologous sequences indicates that it in fact can distinguish between the two species. This indicates that because the clustering algorithm used a heuristic algorithm, some clusters that could have been indicative of the L. jensenii or L. mulieris species failed our initial test.

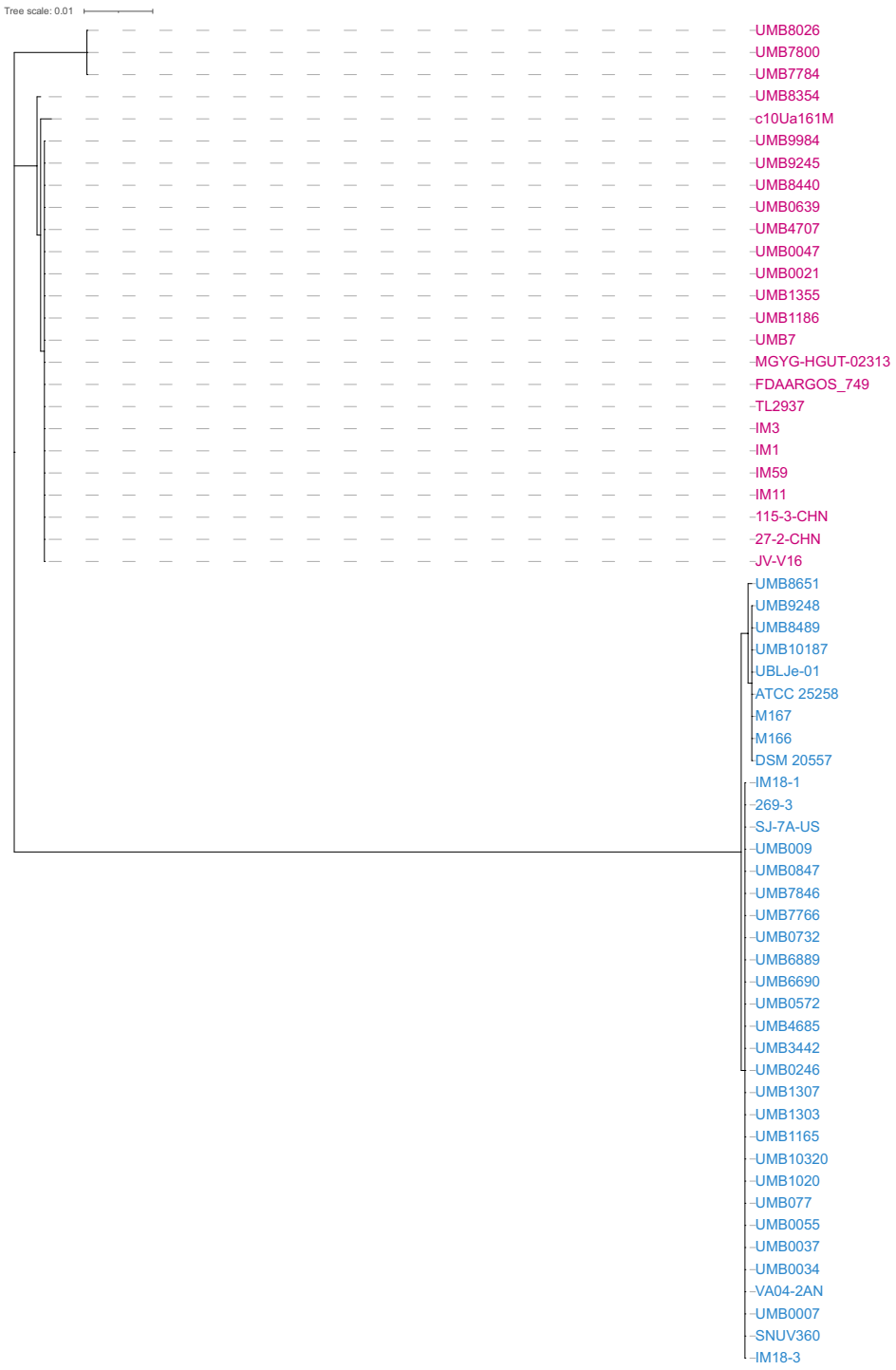

Figure 14. GC_00000137 phylogenetic tree. *L. mulieris* species are in pink while *L. jensenii* species are in blue.

I next assessed the ability for these three genes to be used as gene markers for detecting and distinguishing between the two species. PCR primers were designed to amplify variable regions of the gene sequences. Two primer pairs were designed for GC_000000341. The primer sequences are listed in Table 6. The primers were first tested against isolates of *L. jensenii* and *L. mulieris* as well as several other *Lactobacillus* species commonly found in the urinary tract and vaginal communities (Table 8).

Table 8. Results of PCR screening using the marker gene primer sets. "+" indicates that an amplicon was detected; "-" indicates that an amplicon was not detected.

Species	*Strain*	341_p1	341_p2L	137	105
L. gasseri	UMB1065	-	-	-	-
L. gasseri	UMB1303	-	-	-	-
L. gasseri	UMB607	-	-	-	-
L. iners	UMB436	-	-	-	-
L. crispatus	UMB452	-	-	-	-
L. crispatus	UMB393	-	-	-	-
L. jensenii	UMB6690	+	+	+	+
L. jensenii	UMB3442	+	+	+	+
L. jensenii	UMB732	+	+	+	+
L. jensenii	UMB246	+	+	+	+
L. jensenii	UMB4685	+	+	+	+
L. mulieris	UMB9245	+	+	+	+
L. mulieris	UMB4707	+	+	+	+
L. mulieris	UMB639	+	+	+	+
L. mulieris	UMB9984	+	+	+	+
L. mulieris	UMB8440	+	+	+	+

As Table 8 shows, the primers only amplified *L. jensenii* and *L. mulieris* strains. From the strains tested, the amplicons for five of the strains (*L. jensenii* UMB732, UMB3342, and UMB0034 and *L. mulieris* UMB639 and UMB855) were sequenced to verify the amplified region was the gene of interest. For all four primer pairs, the resulting sequences were (1) amplicons of the intended gene target and (2) capable of distinguishing between the two species.

For instance, the primer pair 137 sequence analysis shows SNPs distinguishing the sequences of the two species (Figure 16).

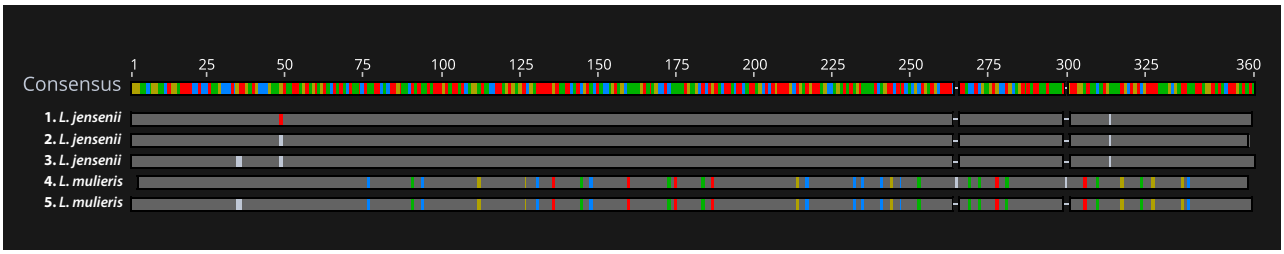

Figure 15. Alignment of amplicon GC_0000137 As you can see the 2 bottom rows are *L. mulieris* strains while the top 3 rows are L. jensenii strains. When we look at the alignment between *L. jensenii* and *L. mulieris*, we can see that *L. mulieris* has distinct mutations from *L. jensenii*, which can let us distinguish between the two species. This is the results from primer pair 137 made for GC_0000137.

Screening urinary samples for *L. jensenii* and *L. mulieris*

The literature of metagenomic studies on urine suggests that *L. jensenii* is prevalent in the urinary tract; however, my results indicate the contrary. 190 urine samples were screened and there was no *L. jensenii* or *L. mulieris* found. The symptom status of the urine from the individual tested can be found in Table 9. There was no amplification by the primers in any of the urine tested. *L. jensenii* and *L. mulieris* DNA was used as a positive control for these PCRs, reliably producing the expected amplicon. This result directly contradicts the findings of prior 16S rRNA surveys of the urinary microbiome in which *L. jensenii* was detected in abundance or prior EQUC isolations of the species. (Hilt et al., 2014; Thomas-White et al., 2018; Putonti et al., 2019; Price et al., 2020a,b; Komesu et al., 2020; Miller-Ensminger, Wolfe & Putonti, 2020; West-Pelak et al., 2020; Rivera et al., 2020; Ene & Putonti, 2022).

Table 9. Symptom status of urine tested

Symptom Status	noOAB/noLUTS	OAB+
Number of Urine screened	62	128
Positive results	0	0

Metagenomic screening

I screened 233 metagenomic datasets from publicly available databases and our own collection. These datasets were screened for the gene GC_00001137 and GC_00001341. I only found six samples that include *L. jensenii* and two samples that include *L. mulieris;* these samples come from 2 different published studies (Moustafa et al., 2018; Adu-Oppong et al., 2022) and an additional study conducted by our group (Mormando et al., in prep). The symptoms of the metagenomic data screened can be found in Table 11. To determine which species was detected, the gene sequences from the metagenomic data was compared to the gene sequences of representatives of *L. jensenii* and *L. mulieris* genomes. Both species were detected as shown in Figure 17 and Figure 18. Given the fact that very few samples contain these species gene markers, we believe that *L. jensenii* and *L. mulieris* are not abundant in the urinary microbiota.

Table 10. Number of metagenomes screened and their symptom status.

Symptom Status	#Metagenomic datasets	# Samples with evidence of genes of interest	Species
Asymptomatic UTI	30	0	n/a
OAB+	17	2	*L. mulieris*
Contaminated Urine Sample	6	1	*L. jensenii*
No OAB/No Luts	2	0	n/a
UTI	111	4	*L. jensenii*
UTI Negative	67	1	n/a

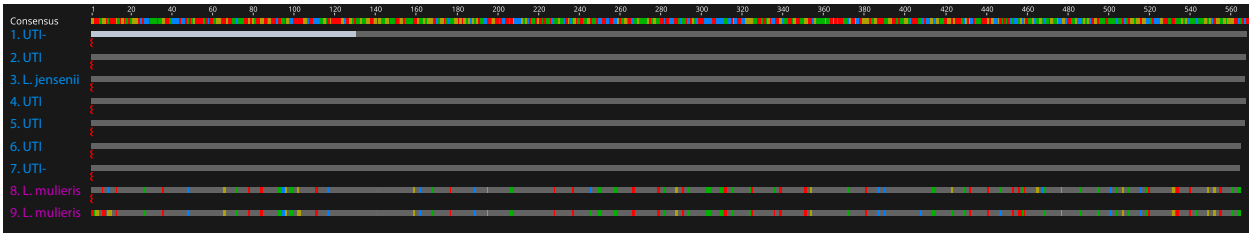

Figure 16. GC_00000341 alignment figure. L. jensenii strains are in blue and the symptom status are showing while *L. mulieris* genes are in pink.

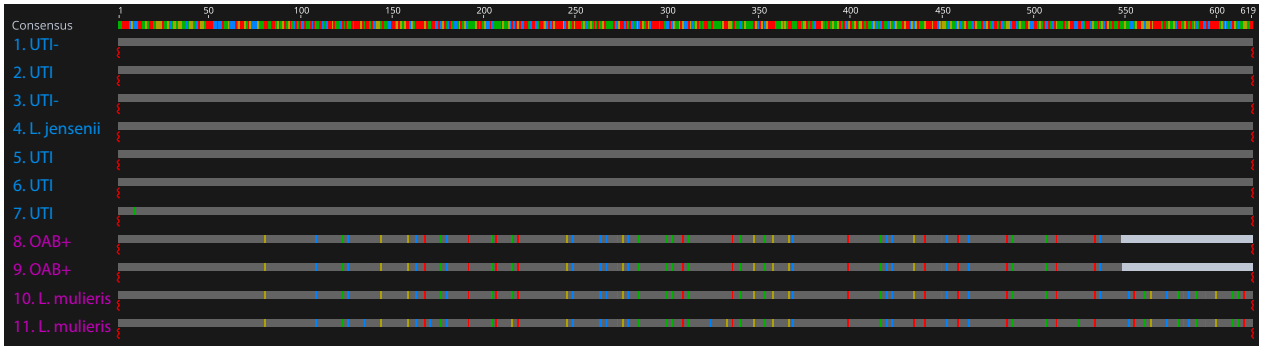

Figure 17. GC_00000137 alignment figure. *L. jensenii* strains are in blue and the symptom status are showing while *L. mulieris* is in pink

CHAPTER 5

CONCLUSIONS

The *Lactobacillus* genus has been isolated from different environments. They inhabit the urogenital tract, the gut and skin. This project has examined two of these species - *L. jensenii* and *L. mulieris.* This included a genomic assessment, as well as a phenotypic characterization. In March 2020, *L. mulieris* was declared a new species (Rocha et al., 2020). Previously, our lab analyzed a small subset (n = 43) of publicly available *L. jensenii* genome and found that some were misclassified and instead were supposed to be *L. mulieris (Putonti et al., 2020).* As it was shown in Chapter 3, the two species have 589 core genes, but they also encode for genes that are unique to and conserved among the strains of each species. Further distinguishing these two species is the secondary metabolites that they encode. Further research is needed to ascertain how or if these secondary metabolites drove the divergence and evolution of these two species.

Of critical importance, given their potential to control the growth of uropathogens in the urinary tract, bladder-relevant attributes were assessed. While both species produce H_2O_2, an known inhibitor of uropathogens in the urinary tract, *L. mulieris* produces more hydrogen peroxide. Thus, *L. mulieris* might be a better probiotic. The different efficiencies for metabolism of specific carbohydrates was particularly of interest. Previously it was shown that ribose can be metabolized by *L. jensenii*; however, this assessment was done using a single strain from the

vaginal tract (Rocha et al., 2020). When characterizing a bacterial species, it is important to test phenotypes of more than one representative per species.

Furthermore, *L. jensenii* was thought to be a prevalent *Lactobacillus* found in the urogenital tract. These analyses were based on 16S rRNA studies of the urogenital metagenome; however, based upon my assessment, *L. jensenii* cannot be distinguished from *L. psittaci and L. fornicalis.* Our lab has done substantial testing of DNA extraction protocols and selected a method that was successful for the extraction of *Lactobacillus* species when processing the 190 urine samples. However, the urine analyzed contained no *L. jensenii* or *L. mulieris.* I looked at the prevalence of *L. jensenii* using whole metagenomes by creating species specific divergent genes. *L. jensenii and L. mulieris* were only found in 3.43% of the metagenomes examined. In the future I would like to analyze more than 2 genes and see if other genes are detected in these metagenomes. However, the two genes I picked are *L. jensenii/L. mulieris* specific and are found in all strains sequenced to date.

While other *Lactobacilli* of the urogenital microbiota have been associated with symptoms (or the lack of symptoms), *L. jensenii/L. mulieris* have been isolated from both females with and without lower urinary tract symptoms. Further work isolating these two species is needed to determine its potential role in urogenital health.

APPENDIX A

STRAINS STUDIED AND THEIR ASSOCIATED INFORMATION

Strain name	Species	Source	Symptoms
JV-V16	*L. mulieris*	Vagina	n/a
27-2-CHN	*L. mulieris*	Vagina	n/a
115-3-CHN	*L. mulieris*	Vagina	n/a
IM11	*L. mulieris*	Unknown	n/a
IM59	*L. mulieris*	Unknown	n/a
IM1	*L. mulieris*	Unknown	n/a
IM3	*L. mulieris*	Unknown	n/a
TL2937	*L. mulieris*	Feces	healthy
c10Ua161M	*L. mulieris*	Urine	n/a
FDAARGOS_749	*L. mulieris*	Unknown	n/a
MGYG-HGUT-02313	*L. mulieris*	Feces	n/a
UMB1186	*L. mulieris*	Urine	UTI
UMB7	*L. mulieris*	Urine	no LUTS
UMB1355	*L. mulieris*	Urine	UTI
UMB21	*L. mulieris*	Urine	no LUTS
UMB4707	*L. mulieris*	Urine	OAB
UMB47	*L. mulieris*	Urine	no LUTS
UMB639	*L. mulieris*	Urine	OAB
UMB7784	*L. mulieris*	Urine	rUTI
UMB7800	*L. mulieris*	Urine	rUTI
UMB8026	*L. mulieris*	Urine	rUTI
UMB8354	*L. mulieris*	Urine	OAB
UMB8440	*L. mulieris*	Urine	OAB
UMB9245	*L. mulieris*	Urine	rUTI
UMB9984	*L. mulieris*	Urine	UTI
SJ-7A-US	*L. jensenii*	Vagina	n/a
269-3	*L. jensenii*	Vagina	n/a
IM18-1	*L. jensenii*	Unknown	n/a
IM18-3	*L. jensenii*	Unknown	n/a
DSM 20557	*L. jensenii*	Vagina	n/a
SNUV360	*L. jensenii*	Vagina	n/a
UMB0007	*L. jensenii*	Urine	no LUTS
VA04-2AN	*L. jensenii*	Vagina	n/a
M166	*L. jensenii*	feces	n/a

M167	*L. jensenii*	feces	n/a
ATCC 25258	*L. jensenii*	Vagina	n/a
UBLJe-01	*L. jensenii*	Vagina	healthy
UMB0034	*L. jensenii*	Urine	no LUTS
UMB0037	*L. jensenii*	Urine	no LUTS
UMB0055	*L. jensenii*	Urine	no LUTS
UMB0077	*L. jensenii*	Urine	OAB
UMB10187	*L. jensenii*	Urine	no LUTS
UMB1020	*L. jensenii*	Urine	UTI
UMB10320	*L. jensenii*	Urine	no LUTS
UMB1165	*L. jensenii*	Urine	UTI
UMB1303	*L. jensenii*	Urine	UTI
UMB1307	*L. jensenii*	Urine	UTI
UMB246	*L. jensenii*	Urine	OAB
UMB3442	*L. jensenii*	Urine	OAB
UMB4685	*L. jensenii*	Urine	OAB
UMB572	*L. jensenii*	Urine	UUI
UMB6690	*L. jensenii*	Urine	rUTI
UMB6889	*L. jensenii*	Urine	no LUTS
UMB732	*L. jensenii*	Urine	no LUTS
UMB7766	*L. jensenii*	Urine	rUTI
UMB7846	*L. jensenii*	Urine	OAB
UMB847	*L. jensenii*	Urine	no LUTS (pregnant)
UMB8489	*L. jensenii*	Urine	OAB
UMB8651	*L. jensenii*	Urine	OAB
UMB9	*L. jensenii*	Urine	no LUTS
UMB9248	*L. jensenii*	Urine	rUTI

APPENDIX B

PUBLICLY AVAILABLE METAGENOMES SCREENED

SRR13740714, SRR5535763, SRR13740673, SRR13740715, SRR13740788, SRR5535724, SRR5535764, SRR13740674, SRR13740716, SRR13740789, SRR5535725, SRR5535765, SRR13740675, SRR13740717, SRR13740790, SRR5535726, SRR5535766, SRR13740676, SRR13740718, SRR13740791, SRR5535727, SRR5535767, SRR13740677, SRR13740719, SRR13740792, SRR5535728, SRR5535768, SRR13740678, SRR13740720, SRR13740793, SRR5535729, SRR5535769, SRR13740679, SRR13740721, SRR13740794, SRR5535730, SRR5535770, SRR13740680, SRR13740722, SRR19149285, SRR5535731, SRR5535771, SRR13740681, SRR13740723, SRR19149284, SRR5535732, SRR5535772, SRR13740682, SRR13740724, SRR19149273, SRR5535733, SRR13740643, SRR13740683, SRR13740725, SRR19149262, SRR5535734, SRR13740644, SRR13740684, SRR13740726, SRR19149258, SRR5535735, SRR13740645, SRR13740685, SRR13740727, SRR19149257, SRR5535736, SRR13740646, SRR13740686, SRR13740728, SRR19149256, SRR5535737, SRR13740647, SRR13740687, SRR13740729, SRR19149255, SRR5535738, SRR13740648, SRR13740688, SRR13740730, SRR19149254, SRR5535739, SRR13740649, SRR13740689, SRR13740731, SRR19149253, SRR5535740, SRR13740650, SRR13740690, SRR13740732, SRR19149283, SRR5535741, SRR13740651, SRR13740691, SRR13740733, SRR19149282, SRR5535742, SRR13740652, SRR13740693, SRR13740734, SRR19149281, SRR5535743, SRR13740653, SRR13740694, SRR13740735, SRR19149280, SRR5535744, SRR13740654, SRR13740695, SRR13740736, SRR19149279, SRR5535745, SRR13740655, SRR13740696, SRR13740737, SRR19149278, SRR5535746, SRR13740656, SRR13740697, SRR13740738, SRR19149277, SRR5535747, SRR13740657, SRR13740698, SRR13740739, SRR19149276, SRR5535748, SRR13740658, SRR13740699, SRR13740740, SRR19149275, SRR5535749, SRR13740659, SRR13740700, SRR13740741, SRR19149274, SRR5535750, SRR13740660, SRR13740701, SRR13740742, SRR19149272, SRR5535751, SRR13740661, SRR13740702, SRR13740743, SRR19149271, SRR5535752, SRR13740662, SRR13740703, SRR13740744, SRR19149270, SRR5535753, SRR13740663, SRR13740704, SRR13740745, SRR19149269, SRR5535754, SRR13740664, SRR13740705, SRR13740747, SRR19149268, SRR5535755, SRR13740665, SRR13740706, SRR13740748, SRR19149267, SRR5535756, SRR13740666, SRR13740707, SRR13740749, SRR19149266, SRR5535757, SRR13740667, SRR13740708, SRR13740750, SRR19149265, SRR5535758, SRR13740668, SRR13740709, SRR13740751, SRR19149264, SRR5535759, SRR13740669, SRR13740710, SRR13740752, SRR19149263, SRR5535760, SRR13740670, SRR13740711, SRR13740753, SRR19149261, SRR5535761, SRR13740671, SRR13740712, SRR13740754, SRR19149260, SRR5535762, SRR13740672, SRR13740713, SRR13740755, SRR19149259, SRR13740775, SRR13740766, SRR13740784, SRR13740760, SRR13740756, SRR13740776, SRR13740767, SRR13740785, SRR13740761, SRR13740757, SRR13740777, SRR13740768, SRR13740786, SRR13740762, SRR13740758, SRR13740778, SRR13740769, SRR13740787, SRR13740763, SRR13740759, SRR13740779, SRR13740770, SRR13740772, SRR13740764, SRR13740774, SRR13740780, SRR13740771, SRR13740773, SRR13740765, SRR13740783, SRR13740781, SRR13740782

BIBLIOGRAPHY

Adu-Oppong B, Thänert R, Wallace MA, Burnham C-AD, Dantas G. 2022. Substantial overlap between symptomatic and asymptomatic genitourinary microbiota states. *Microbiome* 10:6. DOI: 10.1186/s40168-021-01204-9.

Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research* 48:D517–D525. DOI: 10.1093/nar/gkz935.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. DOI: 10.1016/S0022-2836(05)80360-2.

Anglim B, Phillips C, Shynlova O, Alarab M. 2021. The effect of local estrogen therapy on the urinary microbiome composition of postmenopausal women with and without recurrent urinary tract infections. *International Urogynecology Journal*. DOI: 10.1007/s00192-021-04832-9.

Antonio MA, Hawes SE, Hillier SL. 1999. The identification of vaginal Lactobacillus species and the demographic and microbiologic characteristics of women colonized by these species. *The Journal of Infectious Diseases* 180:1950–1956. DOI: 10.1086/315109.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* 44:W16-21. DOI: 10.1093/nar/gkw387.

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, Donk WA van der. 2012. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports* 30:108–160. DOI: 10.1039/C2NP20085F.

Baker JR, Pritchard DG. 2000. Action pattern and substrate specificity of the hyaluronan lyase from group B streptococci. *The Biochemical Journal* 348 Pt 2:465–471.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19:455–477. DOI: 10.1089/cmb.2012.0021.

Barnes HC, Wolff B, Abdul-Rahim O, Harrington A, Hilt EE, Price TK, Halverson T, Hochstedler BR, Pham T, Acevedo-Alvarez M, Joyce C, Fitzgerald CM, Schreckenberger PC, Brubaker L, Wolfe AJ, Mueller ER. 2021. A Randomized Clinical Trial of Standard versus Expanded Cultures to Diagnose Urinary Tract Infections in Women. *The Journal of Urology* 206:1212–1221. DOI: 10.1097/JU.0000000000001949.

BBMap. *Available at https://sourceforge.net/projects/bbmap/* (accessed May 17, 2022).

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49:W29–W35. DOI: 10.1093/nar/gkab335.

Blount ZD. 2015. The unexhausted potential of E. coli. *eLife* 4. DOI: 10.7554/eLife.05826.

Boskey ER, Cone RA, Whaley KJ, Moench TR. 2001. Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. *Human Reproduction (Oxford, England)* 16:1809–1813. DOI: 10.1093/humrep/16.9.1809.

Boskey ER, Telsch KM, Whaley KJ, Moench TR, Cone RA. 1999. Acid Production by Vaginal Flora In Vitro Is Consistent with the Rate and Extent of Vaginal Acidification. *Infection and Immunity* 67:5170–5175.

Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT, Poelzer J, Huynh J, Yallou FS, Psychogios N, Dong E, Bogumil R, Roehring C, Wishart DS. 2013. The human urine metabolome. *PloS One* 8:e73076. DOI: 10.1371/journal.pone.0073076.

de Brito BB, da Silva FAF, Soares AS, Pereira VA, Santos MLC, Sampaio MM, Neves PHM, de Melo FF. 2019. Pathogenesis and clinical management of Helicobacter pylori gastric infection. *World Journal of Gastroenterology* 25:5578–5589. DOI: 10.3748/wjg.v25.i37.5578.

Brosius J, Palmer ML, Kennedy PJ, Noller HF. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 75:4801–4805.

Brubaker L, Gourdine J-PF, Siddiqui NY, Holland A, Halverson T, Limeria R, Pride D, Ackerman L, Forster CS, Jacobs KM, Thomas-White KJ, Putonti C, Dong Q, Weinstein M, Lukacz ES, Karstens L, Wolfe AJ. 2021. Forming Consensus To Advance Urobiome Research. *mSystems* 6:e0137120. DOI: 10.1128/mSystems.01371-20.

Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and molecular biology reviews: MMBR* 68:560–602, table of contents. DOI: 10.1128/MMBR.68.3.560-602.2004.

Butler DSC, Silvestroni A, Stapleton AE. 2016. Cytoprotective Effect of Lactobacillus crispatus CTV-05 against Uropathogenic E. coli. *Pathogens* 5. DOI: 10.3390/pathogens5010027.

Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal* 11:2639–2643. DOI: 10.1038/ismej.2017.119.

Chen Z, Phan M-D, Bates LJ, Peters KM, Mukerjee C, Moore KH, Schembri MA. 2018. The urinary microbiome in patients with refractory urge incontinence and recurrent urinary tract infection. *International Urogynecology Journal* 29:1775–1782. DOI: 10.1007/s00192-018-3679-2.

Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* 33:D325-328. DOI: 10.1093/nar/gki008.

Cheng Q, Carlson B, Pillai S, Eby R, Edwards L, Olmsted SB, Cleary P. 2001. Antibody against surface-bound C5a peptidase is opsonic and initiates macrophage killing of group B streptococci. *Infection and Immunity* 69:2302–2308. DOI: 10.1128/IAI.69.4.2302-2308.2001.

Chiarello M, McCauley M, Villéger S, Jackson CR. 2022. Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. *PLOS ONE* 17:e0264443. DOI: 10.1371/journal.pone.0264443.

van De Wijgert JH, Mason PR, Gwanzura L, Mbizvo MT, Chirenje ZM, Iliff V, Shiboski S, Padian NS. 2000. Intravaginal practices, vaginal flora disturbances, and acquisition of sexually transmitted diseases in Zimbabwean women. *The Journal of Infectious Diseases* 181:587–594. DOI: 10.1086/315227.

Dicks LM, Silvester M, Lawson PA, Collins MD. 2000. Lactobacillus fornicalis sp. nov., isolated from the posterior fornix of the human vagina. *International Journal of Systematic and Evolutionary Microbiology* 50 Pt 3:1253–1258. DOI: 10.1099/00207713-50-3-1253.

van Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. *Methods in Molecular Biology (Clifton, N.J.)* 804:281–295. DOI: 10.1007/978-1-61779-361-5_15.

Dunlop AL, Satten GA, Hu Y-J, Knight AK, Hill CC, Wright ML, Smith AK, Read TD, Pearce BD, Corwin EJ. 2021. Vaginal Microbiome Composition in Early Pregnancy and Risk of Spontaneous Preterm and Early Term Birth Among African American Women. *Frontiers in Cellular and Infection Microbiology* 11:641005. DOI: 10.3389/fcimb.2021.641005.

Edgar RC. 2010.Search and clustering orders of magnitude faster than BLAST | Bioinformatics | Oxford Academic. *Available at https://academic.oup.com/bioinformatics/article/26/19/2460/230188?login=true* (accessed May 12, 2022).

Edgar RC. 2018.Taxonomy annotation and guide tree errors in 16S rRNA databases [PeerJ]. *Available at https://peerj.com/articles/5030/* (accessed January 14, 2022).

Ene A, Putonti C. 2022. Draft Genome Assemblies of 4 Lactobacillus jensenii and 3 Lactobacillus mulieris Strains from the Urinary Tract. *Microbiology Resource Announcements*:e0003222. DOI: 10.1128/mra.00032-22.

Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. DOI: 10.7717/peerj.1319.

Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, Strauss JF, The Vaginal Microbiome Consortium null, Jefferson KK, Buck GA. 2014. Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology (Reading, England)* 160:2272–2282. DOI: 10.1099/mic.0.081034-0.

Fiegler H, Bassias J, Jankovic I, Brückner R. 1999. Identification of a Gene in Staphylococcus xylosus Encoding a Novel Glucose Uptake Protein. *Journal of Bacteriology* 181:4929–4936. DOI: 10.1128/JB.181.16.4929-4936.1999.

Fok CS, Gao X, Lin H, Thomas-White KJ, Mueller ER, Wolfe AJ, Dong Q, Brubaker L. 2018. Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. *International Urogynecology Journal* 29:1765–1771. DOI: 10.1007/s00192-018-3732-1.

Forsyth VS, Armbruster CE, Smith SN, Pirani A, Springman AC, Walters MS, Nielubowicz GR, Himpsl SD, Snitkin ES, Mobley HLT. 2018. Rapid Growth of Uropathogenic Escherichia coli during Human Urinary Tract Infection. *mBio* 9:e00186-18. DOI: 10.1128/mBio.00186-18.

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. DOI: 10.1186/1479-5876-10-174.

France MT, Mendes-Soares H, Forney LJ. 2016. Genomic Comparisons of Lactobacillus crispatus and Lactobacillus iners Reveal Potential Ecological Drivers of Community Composition in the Vagina. *Applied and Environmental Microbiology* 82:7063–7073. DOI: 10.1128/AEM.02385-16.

Garretto A, Miller-Ensminger T, Ene A, Merchant Z, Shah A, Gerodias A, Biancofiori A, Canchola S, Canchola S, Castillo E, Chowdhury T, Gandhi N, Hamilton S, Hatton K, Hyder S, Krull K, Lagios D, Lam T, Mitchell K, Mortensen C, Murphy A, Richburg J, Rokas M, Ryclik S, Sulit P, Szwajnos T, Widuch M, Willis J, Woloszyn M, Brassil B, Johnson G, Mormando R, Maskeri L, Batrich M, Stark N, Shapiro JW, Montelongo Hernandez C, Banerjee S, Wolfe AJ, Putonti C. 2020. Genomic Survey of E. coli From the Bladders of Women With and Without Lower Urinary Tract Symptoms. *Frontiers in Microbiology* 11:2094. DOI: 10.3389/fmicb.2020.02094.

Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA. 2010. Discovery of Unique Lanthionine Synthetases Reveals New Mechanistic and Evolutionary Insights. *PLoS Biology* 8:e1000339. DOI: 10.1371/journal.pbio.1000339.

Gottschick C, Deng Z-L, Vital M, Masur C, Abels C, Pieper DH, Wagner-Döbler I. 2017. The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. *Microbiome* 5:99. DOI: 10.1186/s40168-017-0305-3.

Gupta K, Stapleton AE, Hooton TM, Roberts PL, Fennell CL, Stamm WE. 1998. Inverse association of H2O2-producing lactobacilli and vaginal Escherichia coli colonization in women with recurrent urinary tract infections. *The Journal of Infectious Diseases* 178:446–450. DOI: 10.1086/515635.

Heeney DD, Gareau MG, Marco ML. 2018. Intestinal Lactobacillus in health and disease, a driver or just along for the ride? *Current Opinion in Biotechnology* 49:140–147. DOI: 10.1016/j.copbio.2017.08.004.

Hertzberger R, Arents J, Dekker HL, Pridmore RD, Gysler C, Kleerebezem M, de Mattos MJT. 2014. H(2)O(2) production in species of the Lactobacillus acidophilus group: a central role for a novel NADH-dependent flavin reductase. *Applied and Environmental Microbiology* 80:2229–2239. DOI: 10.1128/AEM.04272-13.

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. DOI: 10.1128/JCM.02876-13.

Hochstedler BR, Burnett L, Price TK, Jung C, Wolfe AJ, Brubaker L. 2022. Urinary microbiota of women with recurrent urinary tract infection: collection and culture methods. *International Urogynecology Journal* 33:563–570. DOI: 10.1007/s00192-021-04780-4.

Hoffman C, Siddiqui NY, Fields I, Gregory WT, Simon HM, Mooney MA, Wolfe AJ, Karstens L. 2021a. Species-Level Resolution of Female Bladder Microbiota from 16S rRNA Amplicon Sequencing. *mSystems* 6:e0051821. DOI: 10.1128/mSystems.00518-21.

Hoffman C, Siddiqui NY, Fields I, Gregory WT, Simon HM, Mooney MA, Wolfe AJ, Karstens L. 2021b. Species-Level Resolution of Female Bladder Microbiota from 16S rRNA Amplicon Sequencing. *mSystems* 6:e0051821. DOI: 10.1128/mSystems.00518-21.

Hoffman C, Siddiqui NY, Fields I, Gregory WT, Simon HM, Mooney MA, Wolfe AJ, Karstens L. Species-Level Resolution of Female Bladder Microbiota from 16S rRNA Amplicon Sequencing. *mSystems* 6:e00518-21. DOI: 10.1128/mSystems.00518-21.

Hooton TM, Roberts PL, Cox ME, Stapleton AE. 2013. Voided midstream urine culture and acute cystitis in premenopausal women. *The New England Journal of Medicine* 369:1883–1891. DOI: 10.1056/NEJMoa1302186.

Hugerth LW, Pereira M, Zha Y, Seifert M, Kaldhusdal V, Boulund F, Krog MC, Bashir Z, Hamsten M, Fransson E, Nielsen HS, Schuppe-Koistinen I, Engstrand L. 2020. Assessment of *In Vitro* and *In Silico* Protocols for Sequence-Based Characterization of the Human Vaginal Microbiome. *mSphere* 5. DOI: 10.1128/mSphere.00448-20.

Hütt P, Lapp E, Štšepetova J, Smidt I, Taelma H, Borovkova N, Oopkaup H, Ahelik A, Rööp T, Hoidmets D, Samuel K, Salumets A, Mändar R. 2016. Characterisation of probiotic properties in human vaginal lactobacilli strains. *Microbial Ecology in Health and Disease* 27. DOI: 10.3402/mehd.v27.30484.

Inglin RC, Meile L, Stevens MJA. 2018. Clustering of Pan- and Core-genome of Lactobacillus provides Novel Evolutionary Insights for Differentiation. *BMC genomics* 19:284. DOI: 10.1186/s12864-018-4601-5.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9:5114. DOI: 10.1038/s41467-018-07641-9.

Jerlström PG, Talay SR, Valentin-Weigand P, Timmis KN, Chhatwal GS. 1996. Identification of an immunoglobulin A binding motif located in the beta-antigen of the c protein complex of group B streptococci. *Infection and Immunity* 64:2787–2793. DOI: 10.1128/iai.64.7.2787-2793.1996.

Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM. 2019. Evaluation of

16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications* 10:5029. DOI: 10.1038/s41467-019-13036-1.

Kalyoussef S, Nieves E, Dinerman E, Carpenter C, Shankar V, Oh J, Burd B, Angeletti RH, Buckheit KW, Fredricks DN, Madan RP, Keller MJ, Herold BC. 2012. Lactobacillus proteins are associated with the bactericidal activity against E. coli of female genital tract secretions. *PloS One* 7:e49506. DOI: 10.1371/journal.pone.0049506.

Karstens L, Asquith M, Davin S, Stauffer P, Fair D, Gregory WT, Rosenbaum JT, McWeeney SK, Nardos R. 2016. Does the Urinary Microbiome Play a Role in Urgency Urinary Incontinence and Its Severity? *Frontiers in Cellular and Infection Microbiology* 6:78. DOI: 10.3389/fcimb.2016.00078.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780. DOI: 10.1093/molbev/mst010.

Kimmel SA, Roberts RF. 1998. Development of a growth medium suitable for exopolysaccharide production by Lactobacillus delbrueckii ssp. bulgaricus RR. *International Journal of Food Microbiology* 40:87–92. DOI: 10.1016/s0168-1605(98)00023-3.

Kirjavainen PV, Pautler S, Baroja ML, Anukam K, Crowley K, Carter K, Reid G. 2009. Abnormal Immunological Profile and Vaginal Microbiota in Women Prone to Urinary Tract Infections. *Clinical and Vaccine Immunology* 16:29–36. DOI: 10.1128/CVI.00323-08.

Klein RD, Hultgren SJ. 2020. Urinary tract infections: microbial pathogenesis, host-pathogen interactions and new treatment strategies. *Nature Reviews. Microbiology* 18:211–226. DOI: 10.1038/s41579-020-0324-0.

Komesu YM, Dinwiddie DL, Richter HE, Lukacz ES, Sung VW, Siddiqui NY, Zyczynski HM, Ridgeway B, Rogers RG, Arya LA, Mazloomdoost D, Levy J, Carper B, Gantz MG, Eunice Kennedy Shriver National Institute of Child Health and Human Development Pelvic Floor Disorders Network. 2020. Defining the relationship between vaginal and urinary microbiomes. *American Journal of Obstetrics and Gynecology* 222:154.e1-154.e10. DOI: 10.1016/j.ajog.2019.08.011.

Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 102:2567–2572. DOI: 10.1073/pnas.0409727102.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9:357–359. DOI: 10.1038/nmeth.1923.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44:W242–W245. DOI: 10.1093/nar/gkw290.

Lewis DA, Brown R, Williams J, White P, Jacobson SK, Marchesi JR, Drake MJ. 2013. The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. *Frontiers in Cellular and Infection Microbiology* 3:41. DOI: 10.3389/fcimb.2013.00041.

Li K, Chen C, Zeng J, Wen Y, Chen W, Zhao J, Wu P. 2022. Interplay between bladder microbiota and overactive bladder symptom severity: a cross-sectional study. *BMC urology* 22:39. DOI: 10.1186/s12894-022-00990-0.

Louis P, Hold GL, Flint HJ. 2014. The gut microbiota, bacterial metabolites and colorectal cancer. *Nature Reviews. Microbiology* 12:661–672. DOI: 10.1038/nrmicro3344.

Ma B, Forney LJ, Ravel J. 2012. The vaginal microbiome: rethinking health and diseases. *Annual review of microbiology* 66:371–389. DOI: 10.1146/annurev-micro-092611-150157.

Maalouf NM, Cameron MA, Moe OW, Adams-Huet B, Sakhaee K. 2007. Low urine pH: a novel feature of the metabolic syndrome. *Clinical journal of the American Society of Nephrology: CJASN* 2:883–888. DOI: 10.2215/CJN.00670207.

Mack CI, Weinert CH, Egert B, Ferrario PG, Bub A, Hoffmann I, Watzl B, Daniel H, Kulling SE. 2018. The complex human urinary sugar profile: determinants revealed in the cross-sectional KarMeN study. *The American Journal of Clinical Nutrition* 108:502–516. DOI: 10.1093/ajcn/nqy131.

Macklaim JM, Fernandes AD, Di Bella JM, Hammond J-A, Reid G, Gloor GB. 2013. Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis. *Microbiome* 1:12. DOI: 10.1186/2049-2618-1-12.

Martín R, Escobedo S, Suárez JE. 2010. Induction, structural characterization, and genome sequence of Lv1, a prophage from a human vaginal Lactobacillus jensenii strain. *International Microbiology: The Official Journal of the Spanish Society for Microbiology* 13:113–121. DOI: 10.2436/20.1501.01.116.

Martín R, Soberón N, Escobedo S, Suárez JE. 2009. Bacteriophage induction versus vaginal homeostasis: role of H(2)O(2) in the selection of Lactobacillus defective prophages. *International Microbiology: The Official Journal of the Spanish Society for Microbiology* 12:131–136.

Martín R, Suárez JE. 2010. Biosynthesis and degradation of H2O2 by vaginal lactobacilli. *Applied and Environmental Microbiology* 76:400–405. DOI: 10.1128/AEM.01631-09.

Martínez-Núñez MA, López VEL y. 2016. Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes* 4:13. DOI: 10.1186/s40508-016-0057-6.

Mehta O, Ghosh TS, Kothidar A, Gowtham MR, Mitra R, Kshetrapal P, Wadhwa N, Thiruvengadam R, GARBH-Ini study group, Nair GB, Bhatnagar S, Das B. 2020. Vaginal Microbiome of Pregnant Indian Women: Insights into the Genome of Dominant Lactobacillus Species. *Microbial Ecology* 80:487–499. DOI: 10.1007/s00248-020-01501-0.

Mendes-Soares H, Suzuki H, Hickey RJ, Forney LJ. 2014. Comparative functional genomics of Lactobacillus spp. reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. *Journal of Bacteriology* 196:1458–1470. DOI: 10.1128/JB.01439-13.

Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, Badger JH, Chinwalla AT, Earl AM, FitzGerald MG, Fulton RS, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi VR, Brooks P, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PS, Chen I-MA, Chen L, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Ding Y, Dooling DJ, Dugan SP, Michael Dunne W, Scott Durkin A, Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney L, Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A, Griggs A, Gujja S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA, Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, Kinder-Haake S, King NB, Knight R, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides NC, La Rosa PS, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu Y, Lo C-C, Lozupone CA, Dwayne Lunsford R, Madden T, Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavrommatis K, McCorrison JM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ, Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Priest ME, Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodriguez-Mueller B, Rogers Y-H, Ross MC, Russ C, Sanka RK, Sankar P, Fah Sathirapongsasuti J, Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM,

Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward DV, Warren W, Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Qing Wu Y, Wylie KM, Wylie T, Yandava C, Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Weinstock GM, Wilson RK, White O, The Human Microbiome Project Consortium. 2012. A framework for human microbiome research. *Nature* 486:215–221. DOI: 10.1038/nature11209.

Miller-Ensminger T, Mormando R, Maskeri L, Shapiro JW, Wolfe AJ, Putonti C. 2020. Introducing Lu-1, a Novel Lactobacillus jensenii Phage Abundant in the Urogenital Tract. *PLOS ONE* 15:e0234159. DOI: 10.1371/journal.pone.0234159.

Miller-Ensminger T, Wolfe AJ, Putonti C. 2020. Draft Genome Sequence of Lactobacillus jensenii Strain UMB7766, Isolated from the Female Bladder. *Microbiology Resource Announcements* 9:e00392-20. DOI: 10.1128/MRA.00392-20.

Montgomery DC. 2013. *Design and analysis of experiments*. Hoboken, NJ: John Wiley & Sons, Inc.

Mores CR, Price TK, Wolff B, Halverson T, Limeira R, Brubaker L, Mueller ER, Putonti C, Wolfe AJ. 2021. Genomic relatedness and clinical significance of Streptococcus mitis strains isolated from the urogenital tract of sexual partners. *Microbial Genomics* 7:mgen000535. DOI: 10.1099/mgen.0.000535.

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. *Scientific Reports* 8:4333. DOI: 10.1038/s41598-018-22660-8.

Nardos R, Leung ET, Dahl EM, Davin S, Asquith M, Gregory WT, Karstens L. 2022. Network-Based Differences in the Vaginal and Bladder Microbial Communities Between Women With and Without Urgency Urinary Incontinence. *Frontiers in Cellular and Infection Microbiology* 12:759156. DOI: 10.3389/fcimb.2022.759156.

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. DOI: 10.1128/mBio.00218-20.

NIH HMP Working Group, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M. 2009. The NIH Human Microbiome Project. *Genome Research* 19:2317–2323. DOI: 10.1101/gr.096651.109.

Nunn KL, Forney LJ. 2016. Unraveling the Dynamics of the Human Vaginal Microbiome. *The Yale Journal of Biology and Medicine* 89:331–337.

Oh NS, Joung JY, Lee JY, Kim Y. 2018. Probiotic and anti-inflammatory potential of Lactobacillus rhamnosus 4B15 and Lactobacillus gasseri 4M13 isolated from infant feces. *PLoS ONE* 13. DOI: 10.1371/journal.pone.0192021.

Okuyama M, Mori H, Chiba S, Kimura A. 2004. Overexpression and characterization of two unknown proteins, YicI and YihQ, originated from Escherichia coli. *Protein Expression and Purification* 37:170–179. DOI: 10.1016/j.pep.2004.05.008.

Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science (New York, N.Y.)* 276:734–740. DOI: 10.1126/science.276.5313.734.

Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014a. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *mBio* 5:e01283-01214. DOI: 10.1128/mBio.01283-14.

Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014b. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *mBio* 5:e01283-01214. DOI: 10.1128/mBio.01283-14.

Pearce MM, Zilliox MJ, Rosenfeld AB, Thomas-White KJ, Richter HE, Nager CW, Visco AG, Nygaard IE, Barber MD, Schaffer J, Moalli P, Sung VW, Smith AL, Rogers R, Nolen TL, Wallace D, Meikle SF, Gai X, Wolfe AJ, Brubaker L, Pelvic Floor Disorders Network. 2015. The female urinary microbiome in urgency urinary incontinence. *American Journal of Obstetrics and Gynecology* 213:347.e1–11. DOI: 10.1016/j.ajog.2015.07.009.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* 5:e9490. DOI: 10.1371/journal.pone.0009490.

Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, Brubaker L, Wolfe AJ, Mueller ER, Schreckenberger PC. 2016. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. *Journal of Clinical Microbiology* 54:1216–1222. DOI: 10.1128/JCM.00044-16.

Price TK, Hilt EE, Thomas-White K, Mueller ER, Wolfe AJ, Brubaker L. 2020a. The urobiome of continent adult women: a cross-sectional study. *BJOG: an international journal of obstetrics and gynaecology* 127:193–201. DOI: 10.1111/1471-0528.15920.

Price RJ, Lee JS. 1970. INHIBITION OF PSEUDOMONAS SPECIES BY HYDROGEN PEROXIDE PRODUCING LACTOBACILLI1. *Journal of Milk and Food Technology* 33:13–18. DOI: 10.4315/0022-2747-33.1.13.

Price TK, Lin H, Gao X, Thomas-White KJ, Hilt EE, Mueller ER, Wolfe AJ, Dong Q, Brubaker L. 2020b. Bladder bacterial diversity differs in continent and incontinent women: a cross-sectional study. *American Journal of Obstetrics and Gynecology* 223:729.e1-729.e10. DOI: 10.1016/j.ajog.2020.04.033.

Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2015. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* 8:12–24. DOI: 10.1039/C5AY02550H.

Putonti C, Ahmad A, Baddoo G, Diaz J, Do M, Gallian N, Lorentzen C, Mohammed H, Murphy J, Olu-Ajeigbe A, Yang T, Miller-Ensminger T, Stark N, Maskeri L, Van Dusen J, Wolfe AJ. 2019. Draft Genome Sequences of 11 Lactobacillus jensenii Strains Isolated from the Female Bladder. *Microbiology Resource Announcements* 8:e00970-19. DOI: 10.1128/MRA.00970-19.

Putonti C, Shapiro JW, Ene A, Tsibere O, Wolfe AJ. 2020. Comparative Genomic Study of Lactobacillus jensenii and the Newly Defined Lactobacillus mulieris Species Identifies Species-Specific Functionality. *mSphere* 5:e00560-20. DOI: 10.1128/mSphere.00560-20.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J-M, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD, Wang J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65. DOI: 10.1038/nature08821.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:D590–D596. DOI: 10.1093/nar/gks1219.

Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M, Sentausa E, Raoult D, Fournier P-E. 2014. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *International Journal of Systematic and Evolutionary Microbiology* 64:384–391. DOI: 10.1099/ijs.0.057091-0.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America* 108:4680–4687. DOI: 10.1073/pnas.1002611107.

Richter HE, Carnes MU, Komesu YM, Lukacz ES, Arya L, Bradley M, Rogers RG, Sung VW, Siddiqui NY, Carper B, Mazloomdoost D, Dinwiddie D, Gantz MG, Eunice Kennedy

Shriver National Institute of Child Health and Human Development Pelvic Floor Disorders Network. 2022. Association between the urogenital microbiome and surgical treatment response in women undergoing midurethral sling operation for mixed urinary incontinence. *American Journal of Obstetrics and Gynecology* 226:93.e1-93.e15. DOI: 10.1016/j.ajog.2021.07.008.

Rivera I, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. 2020. Draft Genome Sequence of Lactobacillus jensenii UMB0836, Isolated from the Female Bladder. *Microbiology Resource Announcements* 9:e00410-20. DOI: 10.1128/MRA.00410-20.

Rocha J, Botelho J, Ksiezarek M, Perovic SU, Machado M, Carriço JA, Pimentel LL, Salsinha S, Rodríguez-Alcalá LM, Pintado M, Ribeiro TG, Peixe L. 2020. Lactobacillus mulieris sp. nov., a new species of Lactobacillus delbrueckii group. *International Journal of Systematic and Evolutionary Microbiology* 70:1522–1527. DOI: 10.1099/ijsem.0.003901.

Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Nikita L, Galuppi M, Lamont RF, Chaemsaithong P, Miranda J, Chaiworapongsa T, Ravel J. 2014. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* 2:4. DOI: 10.1186/2049-2618-2-4.

Sabbuba NA, Mahenthiralingam E, Stickler DJ. 2003. Molecular epidemiology of Proteus mirabilis infections of the catheterized urinary tract. *Journal of Clinical Microbiology* 41:4961–4965. DOI: 10.1128/JCM.41.11.4961-4965.2003.

Schouls LM, Schot CS, Jacobs JA. 2003. Horizontal transfer of segments of the 16S rRNA genes between species of the Streptococcus anginosus group. *Journal of Bacteriology* 185:7241–7246. DOI: 10.1128/JB.185.24.7241-7246.2003.

Sewankambo N, Gray RH, Wawer MJ, Paxton L, McNaim D, Wabwire-Mangen F, Serwadda D, Li C, Kiwanuka N, Hillier SL, Rabe L, Gaydos CA, Quinn TC, Konde-Lule J. 1997. HIV-1 infection associated with abnormal vaginal flora morphology and bacterial vaginosis. *Lancet (London, England)* 350:546–550. DOI: 10.1016/s0140-6736(97)01063-5.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498–2504. DOI: 10.1101/gr.1239303.

Siddiqui H, Lagesen K, Nederbragt AJ, Jeansson SL, Jakobsen KS. 2012. Alterations of microbiota in urine from women with interstitial cystitis. *BMC microbiology* 12:205. DOI: 10.1186/1471-2180-12-205.

Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS. 2011. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. *BMC microbiology* 11:244. DOI: 10.1186/1471-2180-11-244.

Simerville JA, Maxted WC, Pahira JJ. 2005. Urinalysis: a comprehensive review. *American Family Physician* 71:1153–1162.

Singh AK, Hertzberger RY, Knaus UG. 2018. Hydrogen peroxide production by lactobacilli promotes epithelial restitution during colitis. *Redox Biology* 16:11–20. DOI: 10.1016/j.redox.2018.02.003.

Spear GT. 2014.Free glycogen in vaginal fluids is associated with Lactobacillus colonization and low vaginal pH - PubMed. *Available at https://pubmed.ncbi.nlm.nih.gov/25033265/* (accessed May 18, 2022).

Spear GT, French AL, Gilbert D, Zariffard MR, Mirmonsef P, Sullivan TH, Spear WW, Landay A, Micci S, Lee B-H, Hamaker BR. 2014. Human α-amylase present in lower-genital-tract mucosal fluid processes glycogen to support vaginal colonization by Lactobacillus. *The Journal of Infectious Diseases* 210:1019–1028. DOI: 10.1093/infdis/jiu231.

Tanizawa Y, Tada I, Kobayashi H, Endo A, Maeno S, Toyoda A, Arita M, Nakamura Y, Sakamoto M, Ohkuma M, Tohno M. 2018. Lactobacillus paragasseri sp. nov., a sister taxon of Lactobacillus gasseri, based on whole-genome sequence analyses. *International Journal of Systematic and Evolutionary Microbiology* 68:3512–3517. DOI: 10.1099/ijsem.0.003020.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science (New York, N.Y.)* 278:631–637. DOI: 10.1126/science.278.5338.631.

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* 44:6614–6624. DOI: 10.1093/nar/gkw569.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." *Proceedings of the National Academy of Sciences* 102:13950–13955. DOI: 10.1073/pnas.0506758102.

The Human Microbiome Project Consortium. 2012. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* 486:207–214. DOI: 10.1038/nature11234.

Thomas-White K, Brady M, Wolfe AJ, Mueller ER. 2016a. The bladder is not sterile: History and current discoveries on the urinary microbiome. *Current Bladder Dysfunction Reports* 11:18–24. DOI: 10.1007/s11884-016-0345-8.

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. *Nature Communications* 9:1557. DOI: 10.1038/s41467-018-03968-5.

Thomas-White KJ, Hilt EE, Fok C, Pearce MM, Mueller ER, Kliethermes S, Jacobs K, Zilliox MJ, Brincat C, Price TK, Kuffel G, Schreckenberger P, Gai X, Brubaker L, Wolfe AJ. 2016b. Incontinence medication response relates to the female urinary microbiota. *International Urogynecology Journal* 27:723–733. DOI: 10.1007/s00192-015-2847-x.

Thomas-White KJ, Kliethermes S, Rickey L, Lukacz ES, Richter HE, Moalli P, Zimmern P, Norton P, Kusek JW, Wolfe AJ, Brubaker L, National Institute of Diabetes and Digestive and Kidney Diseases Urinary Incontinence Treatment Network. 2017. Evaluation of the urinary microbiota of women with uncomplicated stress urinary incontinence. *American Journal of Obstetrics and Gynecology* 216:55.e1-55.e16. DOI: 10.1016/j.ajog.2016.07.049.

Tisza MJ, Buck CB. 2021. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proceedings of the National Academy of Sciences of the United States of America* 118:e2023202118. DOI: 10.1073/pnas.2023202118.

Tortelli BA, Lewis WG, Allsworth JE, Member-Meneh N, Foster LR, Reno HE, Peipert JF, Fay JC, Lewis AL. 2020. Associations between the vaginal microbiome and Candida colonization in women of reproductive age. *American Journal of Obstetrics and Gynecology* 222:471.e1-471.e9. DOI: 10.1016/j.ajog.2019.10.008.

Toyoshima H, Shibahara T, Tanigawa M, Masuda N, Ishiguro C, Tanaka H, Nakanishi Y, Sakabe S. 2021. Lactobacillus paragasseri as a novel causative pathogen of cavernosal abscess. *IDCases* 26:e01320. DOI: 10.1016/j.idcr.2021.e01320.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. 2009. A core gut microbiome in obese and lean twins. *Nature* 457:480–484. DOI: 10.1038/nature07540.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The Human Microbiome Project. *Nature* 449:804–810. DOI: 10.1038/nature06244.

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. DOI: 10.1093/nar/gks596.

Van de Peer Y, Chapelle S, De Wachter R. 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research* 24:3381–3391. DOI: 10.1093/nar/24.17.3381.

Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences* 111:9259–9264. DOI: 10.1073/pnas.1401734111.

West-Pelak E, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. 2020. Draft Genome Sequence of Lactobacillus jensenii UMB0847, Isolated from the Female Urinary Tract. *Microbiology Resource Announcements* 9:e00395-20. DOI: 10.1128/MRA.00395-20.

Wiesenfeld HC, Hillier SL, Krohn MA, Landers DV, Sweet RL. 2003. Bacterial vaginosis is a strong predictor of Neisseria gonorrhoeae and Chlamydia trachomatis infection. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 36:663–668. DOI: 10.1086/367658.

Wilks M, Wiggins R, Whiley A, Hennessy E, Warwick S, Porter H, Corfield A, Millar M. 2004. Identification and H2O2 Production of Vaginal Lactobacilli from Pregnant Women at High Risk of Preterm Birth and Relation with Outcome. *Journal of Clinical Microbiology* 42:713–717. DOI: 10.1128/JCM.42.2.713-717.2004.

Witkin SS, Moron AF, Linhares IM, Forney LJ. 2021. Influence of Lactobacillus crispatus, Lactobacillus iners and Gardnerella vaginalis on bacterial vaginal composition in pregnant women. *Archives of Gynecology and Obstetrics* 304:395–400. DOI: 10.1007/s00404-021-05978-z.

Wittouck S, Wuyts S, Lebeer S. 2019. Towards a Genome-Based Reclassification of the Genus Lactobacillus. *Applied and Environmental Microbiology* 85:e02155-18. DOI: 10.1128/AEM.02155-18.

Wittouck S, Wuyts S, Meehan CJ, van Noort V, Lebeer S. 2019. A Genome-Based Species Taxonomy of the *Lactobacillus* Genus Complex. *mSystems* 4:e00264-19. DOI: 10.1128/mSystems.00264-19.

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. *Journal of Clinical Microbiology* 50:1376–1383. DOI: 10.1128/JCM.05852-11.

Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen K-Y. 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiology and Infection* 14:908–934. DOI: 10.1111/j.1469-0691.2008.02070.x.

Wu P, Chen Y, Zhao J, Zhang G, Chen J, Wang J, Zhang H. 2017. Urinary Microbiome and Psychological Factors in Women with Overactive Bladder. *Frontiers in Cellular and Infection Microbiology* 7:488. DOI: 10.3389/fcimb.2017.00488.

Xu H, Tamrat NE, Gao J, Xu J, Zhou Y, Zhang S, Chen Z, Shao Y, Ding L, Shen B, Wei Z. 2021. Combined Signature of the Urinary Microbiome and Metabolome in Patients With Interstitial Cystitis. *Frontiers in Cellular and Infection Microbiology* 11:711746. DOI: 10.3389/fcimb.2021.711746.

Yamamoto HS, Xu Q, Fichorova RN. 2013. Homeostatic properties of Lactobacillus jensenii engineered as a live vaginal anti-HIV microbicide. *BMC microbiology* 13:4. DOI: 10.1186/1471-2180-13-4.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews. Microbiology* 12:635–645. DOI: 10.1038/nrmicro3330.

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. DOI: 10.1186/1471-2105-13-134.

Zeng J, Zhang G, Chen C, Li K, Wen Y, Zhao J, Wu P. 2020. Alterations in Urobiome in Patients With Bladder Cancer and Implications for Clinical Outcome: A Single-Institution Study. *Frontiers in Cellular and Infection Microbiology* 10:555508. DOI: 10.3389/fcimb.2020.555508.

Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *International Journal of Systematic and Evolutionary Microbiology* 70:2782–2858. DOI: 10.1099/ijsem.0.004107.

Zhou X, Yang B, Stanton C, Ross RP, Zhao J, Zhang H, Chen W. 2020. Comparative analysis of Lactobacillus gasseri from Chinese subjects reveals a new species-level taxa. *BMC genomics* 21:119. DOI: 10.1186/s12864-020-6527-y.

Zhu Y, Huang WE, Yang Q. 2022. Clinical Perspective of Antimicrobial Resistance in Bacteria. *Infection and Drug Resistance* 15:735–746. DOI: 10.2147/IDR.S345574.

Zunga M, Yebra MJ, Monedero V. 2021. Complex Oligosaccharide Utilization Pathways in Lactobacillus. *Current Issues in Molecular Biology* 40:49–80. DOI: 10.21775/cimb.040.049.

VITA

Adriana Ene was born in Satu Mare, Romania. She was raised there until she was 16, when she moved to the United States. From a young age, she was fascinated by science and technology. She attended Loyola University Chicago where she received a Bachelor's of Science. She double majored in Biology and Bioinformatics and graduated with Interdisciplinary Honors. At Loyola University Chicago, she joined the Putonti Lab where her research involving bacteria and bacteriophages started. During this time, she has been a co-author on 8 publications. She is planning to apply to medical school and attend medical school in Fall 2023.