7-2-2024

Disambiguation of Aerococcus Urinae Species Diversity

Brian Ian Choi
Loyola University of Chicago Graduate School

LOYOLA UNIVERSITY CHICAGO

DISAMBIGUATION OF *AEROCOCCUS URINAE* SPECIES DIVERSITY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN MICROBIOLOGY AND IMMUNOLOGY

BY

BRIAN IAN CHOI

CHICAGO, ILLINOIS

AUGUST 2024

ACKNOWLEDGEMENTS

I am endlessly thankful to have been granted such an invigorating and enthralling academic journey while working on this dissertation. To have been guided by such kind and earnest mentors has been an honor. To have been surrounded by people who have supported and celebrated my accomplishments at every stage has been a delightful blessing. I am forever grateful for this opportunity to pursue my passion of microbiology and to share its mysteries with everyone.

To my mentor, Alan J. Wolfe, I am the most thankful. His keen abilities in acknowledging my potential, honing my talents, and leading me down new paths had me engaged at every moment. There was a sense of urgency in his excitement at every new project or collaboration, as if the next discovery could not come soon enough. This excitement on his face and in his voice would never get old, and I would feel his passion carry through my works. I could not ask for a better scientist and friend in a mentor than Alan J. Wolfe.

My education and scholarship was cultivated by such an esteemed group of mentors and colleagues, to whom I express my thankfulness. The microbiology experience of Karen Visick and the bioinformatic guidance of Catherine Putonti was critical in my navigation of this dissertation. I am also thankful for the insightful advice from my other committee members of John Allen, David Hecht, and Nicole Gilbert. To my overseas collaborators, Jens J. Christensen, Erik Senneby, Rimtas Dargis, and Christian H. Schouw, I am grateful for your patience and

persistence in seeing our combined efforts come to fruition. The collection, organization, and interpretation of data in this project was supported in large part by Melline Fontes Noronha, Wilson Geaman, Adriana Ene, Jyoti Lamichhane, and Jake Kaindl whose combined work likely saved me several years of labor. I am also thankful for the support of the rest of my lab team, LUEREC colleagues, and department members in fostering such a warm and inviting community.

I particularly want to express my gratitude towards those who granted me the chance to arrive at Loyola University and find my passions. To the late Andrew Dingwall, I will always cherish the long moments we shared chatting about anything and everything. His tireless efforts in bringing the MD/PhD community together as family created a place of warmth and shelter. I also extend this gratitude to my other directors of Charles Hemenway, Katherine Knight, Mitchell Denning, James Mendez, Gregory Gruener, and Sam Marzo for upholding and protecting our institution through the tumultuous COVID-19 pandemic. Lastly, of course, I am thankful for my friends and family for providing all the safety and support I could ever wish for. I will carry this gratitude forward in all my future efforts and accomplishments as I continue on my life journey of medicine and academia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANI	Average Nucleotide Identity
API	Analytical Profile Index
ARG	Antibiotic Resistance Gene
ASP	Aerococcus Surface Protein
ATCC	American Type Culture Collection
AUC	Aerococcus urinae Complex
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CAMHB	Cation-Adjusted Muller-Hinton Broth
CARD	Comprehensive Antibiotic Resistance Database
CCUG	Culture Collection University of Gothenburg
CDC	Centers for Disease Control and Prevention
CDM	Chemically Defined Medium
CFU	Colony Forming Unit
CLSI	Clinical and Laboratory Standards Institute
CPS	Capsule Biosynthesis Operon
CST	Community State Type
DDH	DNA-DNA Hybridization

DSM	German Collection of Microorganisms and Cell Cultures GmbH
EQUC	Expanded Quantitative Urine Culture
IC/PBS	Interstitial Cystitis / Painful Bladder Syndrome
ICE	Integrative and Conjugative Element
ICSP	International Committee on Systematics of Prokaryotes
IJSEM	International Journal of Systematic and Evolutionary Microbiology
LHB	Lysed Horse Blood
LUTS	Lower Urinary Tract Symptoms
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization-Time of Flight
MHL	Main Hypervariable Locus
MIC	Minimum Inhibitory Concentration
MLST	Multi-Locus Sequence Typing
MS	Mass Spectrometry
MSCRAMM	Microbial Surface Components Recognizing Adhesive Matrix Molecules
NCBI	National Center for Biotechnology Information
NCS	National Center for Streptococcus
NYCIII	New York City III
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PsrP	Pneumococcal serine-rich repeat Protein
PTS	Phosphotransferase System
rRNA	Ribosomal RNA
SHP	Short Hydrophobic Peptide
sRNA	small RNA
SRR	Serine-Rich Repeat
SUC	Standard Urinary Culture

SUI	Stress Urinary Incontinence
UTI	Urinary Tract Infection
UUI	Urgency Urinary Incontinence

ABSTRACT

Aerococcus urinae is an uncommon inhabitant of the human urogenital tract. The Gram-positive bacterium has been implicated in a wide range of urinary tract diseases that, if left untreated or undiagnosed, can progress to severe disease such as infective endocarditis. Comparative analysis of 115 *A. urinae* isolates from infection episodes of urinary tract infection, bacteremia, infective endocarditis, urgency urinary incontinence, stress urinary incontinence, interstitial cystitis / painful bladder syndrome, and asymptomatic individuals reveal substantial diversity with respect to predicted virulence factors, metabolism, and observed biofilm behavior. Through whole genome sequencing analysis, these strains can be reorganized into six new taxonomic groups, representing a complex of species.

The newly named species, *A. tenax, A. mictus*, and *A. loyolae*, and the newly defined *A. urinae* each display their own unique phenotypes including metabolism and flocking biofilm ability that allow each to be differentiated from one another. It is hypothesized that these species-specific behaviors confer pathogenic potential through tissue adherence, host-tissue tropisms, and growth dynamics. This study analyzes these differences through comparative genomics, identifying distinguishing features that can explain speciation. Among these features, horizontal gene transfer was found to play a significant role in genomic variation. Vectors such as integrative and conjugative elements (ICEs) were identified as key in introducing large genomic content such as antibiotic resistance genes.

To facilitate rapid and unambiguous identification of these new species, a multiplex PCR test was created. This test was evaluated on 189 clinical isolates previously identified as *A. urinae* and successfully identified 88% of isolates. From this dataset, the species distribution in a tertiary hospital system was determined, finding that certain species are far more commonly isolated than others. It is anticipated that the findings of this study will lead to the development of better identification, detection, and diagnostic practices that will ultimately contribute towards treatment strategies preventing any *Aerococcus* infection from exacerbating to fatality. The strategies developed in this study can also be used as a template to investigate other poorly understood bacteria, particularly uncharacterized emerging uropathogens, broadening our understanding of the human urinary microbiome.

CHAPTER ONE

INTRODUCTION AND LITERATURE REVIEW

When Koch's Postulates Fail

How do you go about studying a living animal that no one has ever seen before? It is like finding a mythical creature, deep in the unexplored forest. While this creature may be effectively invisible to human eyes, its presence is still felt through its impact on the local ecosystem; no lifeform exists in a vacuum. Microorganisms remained ostensibly invisible until their initial discovery beneath a microscope, yet their presence was still felt in the diseases they spread and food they fermented. To this day, an inscrutable diversity of microbes remains unseen. Unobserved and unnamed, their lives continue to remain a mystery until the moment a human figures out how to capture one.

Like ascending a treacherous mountain, many the microbiologist has attempted to reach new heights in exploration with veterans leaving behind well-worn paths and crafted techniques. However, time and time again these legacies would be up-ended by rule-breaking specimens, tearing dogma apart. From bacteria to viruses to prions, the forms that microscopic entities take to infect and destroy us are outlandish, always pressing our own definitions of what qualifies as "alive." Indeed, we only need look to Dr. Robert Koch, one of the founders of bacteriology, whose famous postulates delivered his discovery of Mycobacterium tuberculosis, the causative agent of tuberculosis. Yet these postulates quickly folded as he failed to isolate tuberculosis's cousin, *Mycobacterium leprae*, the causative agent of leprosy that, today,

continues to remain uncultivable (1).The task of establishing causality between parasite and disease is still a daunting challenge even with today's technology. Similar to dark matter in space, microbiologists are aware of the existence of thousands of unnamed microbial species, but we have not yet arrived at the techniques to isolate and observe these organisms in the laboratory environment (2). This microbial dark matter, although invisible to direct observation, can still be characterized through indirect methods. As stated before, no lifeform exists in a vacuum. As such, scientists can study an unknown creature by what it leaves behind – DNA. The advent of high-throughput sequencing technologies has enabled the reconstruction of genomes belonging to microbes that have even yet to be named. As more and more of these genomes are annotated, tests can be created to identify these microbes in their native environment, shedding light on their hidden lives.

The information contained in this introduction briefly describes the literature behind several of the techniques used to identify these hidden microbes, how the technology provided the foundation for the study of the urinary microbiome and, finally, how a particular urinary bacterial species, *Aerococcus urinae*, earned its classification as an "emerging uropathogen."

Growth and Isolation of Human Microorganisms

One of the first major hints of microbial dark matter arose in a conundrum known as the "great plate-count anomaly" (3). The puzzle is that of a numerical discrepancy; the number of distinct colony morphologies arising from individual bacterial cells are far less than the count of cellular morphologies enumerated when visualized beneath a microscope. The problem lies in the medium on which the bacteria are fostered. Under laboratory conditions that are often room temperature, ambient atmosphere, and biochemically discrete, bacteria are challenged to

find survival in an environment removed from their natural habitat. Thus, it is impossible to expect to replicate the infinite diversity of microbial habitats within a single growth plate.

It is estimated that less than 1% of this microbial diversity is capable of cultivation with today's known techniques (4, 5). Of course, this estimation is really a conjecture of convenience. If we truly desired to cultivate all of the microbes in a rainforest, then all that would need be done is to carve out a piece of the rainforest and bring it home to the laboratory – trees, dirt, atmosphere, sun exposure, and all. Of course, that is easier said than done. On another note, to perfectly cultivate all of the microorganisms within a sick patient, then scientists would need to find a willing participant that they could subject to similar pathogenic conditions, an ethical catastrophe that Robert Koch's postulates could also not overcome. Thus, to provide a cheap, replicable, and safe culture environment to isolate and study all microorganisms is a rather insurmountable task.

Nevertheless, the field of medical microbiology attempts to meet that challenge for the sake of diagnosis and therapy. In general, there are two forms of media that are utilized: liquid broth and solid agar. A liquid broth medium was first developed by Dr. Louis Pasteur in 1860 and featured several basic components that met the basic requirements of nitrogen, carbon, and vitamins (6). The second form was introduced in 1881 by Dr. Robert Koch in a solid medium form that is agar-based. This medium utilized a meat extract to provide nutrients for the bacteria and would eventually be held in a glass plate known as the Petri dish developed by Dr. Julius Petri in 1887. From this point on, variations in media compositions would revolve around determining the nutritional and environmental requirements of target bacteria for their isolation and purification. Media recipes shuffled different carbon source combinations such as

sugars and starches, nitrogen sources such tryptone or peptone, and various growth factors such as vitamins and minerals. Livestock replacements of human components became standardized, such as with sheep's blood, bovine serum, and rumen fluid (7). Lastly, environmental factors would be manipulated, including pH, salinity, gas pressures, temperature, and agitation (shaking). Naturally, bacterial growth requirements spanned all of these variables and more, leading to hundreds of different media types.

Overturning the Paradigm of Urine Sterility

When bacteria were first characterized beneath the microscope, their life cycles were poorly understood. In the 1880's, there existed two competing theories as to the origins of these microorganisms. The roots of the incumbent theory of spontaneous generation dates back to the Greek philosopher Aristotle positing that life could regularly arise spontaneously from organic matter without the need of any preexisting ancestor (8). Supporters of this theory, like Dr. Henry Bastion, produced works demonstrating how boiled urine in just two days could become "fertile" with bacteria (9). But these claims would be famously overturned by the experiments of Dr. Louis Pasteur and Dr. John Tyndall demonstrating that it was the air itself that carried bacteria into the urine media, introducing germ theory into wider acceptance (10, 11)

The common medium utilized in these competing experiments was urine, an easily accessible organic broth within which bacteria could grow (12). Standardized methodologies for obtaining this urine were described by Dr. Pasteur and Dr. William Roberts, the latter who characterized fresh and healthy urine as microorganism-free (13, 14). This assumption that healthy urine was devoid of microbes was well compatible with germ theory. Theorizing that

disease was caused by the presence of bacteria would easily lead one to conclude that urine sterility means to be healthy.

Although not as common today, many prominent deaths in history have been attributed to urinary tract infections from holy roman emperor Henry II to Charlie's Angels actress Tanya Roberts. With infections accounting for some of the greatest morbidity and mortality of mankind, the miracle of antibiotics was propagandized to both physician and patient as a cure-all silver bullet (15). Thus, the physician did not consider the identity of the offending microbe found within a patient's urine to be significant, but rather that there was significant bacteriuria at all that would merit antimicrobial treatment. The threshold for therapy most widely adopted by modern physicians was developed by Dr. Edward Kass in the 1960's where he identified bacterial counts at colony forming units per milliliter urine (CFUs/mL) greater than 10^5 usually predicting pathogenic risk over urine specimen contamination (16, 17). What Dr. Kass was describing, however, was asymptomatic bacteriuria or lower urinary tract bacterial colonization without accompanying symptoms. Today, asymptomatic bacteriuria is increasingly recognized as a benign phenomenon that does not immediately require antimicrobial therapy (18). However, this view was not widely accepted historically and continues to be denied in several medical circles.

The true rate of asymptomatic bacteriuria among populations is unknown because, as mentioned before, it is not possible to identify all the microbes from within the human body even with today's technology. Thus, the standard method for urinary tract infection (UTI) diagnosis by culturing uropathogens on growth media from urine samples inherently carries the risk of misdiagnosis by means of false-negative results. This was a concern outlined in the work

of Dr. Rosalind Maskell in 1979 wherein she documented patients expressing UTI symptoms while demonstrating negative urine culture results (19). Dr. Maskell concluded that the standard media conditions were not sufficient to accurately identify bacteria of several urinary disorders, and she expressed these concerns to the greater scientific community, only to be ignored (20, 14). Unfortunately, her work was devalued by critical reviews, ultimately suppressed by prevailing dogma (21, 22). Later, Maskell's findings would be supported by in-depth studies utilizing more advanced culturing techniques and DNA-based technologies that would rapidly expand the field of urinary microbiome research (14).

Phylogenetics and Genome Sequencing

As mentioned, the greatest challenge in identifying bacteria by culture is finding the proper conditions that allow the bacteria to grow in the laboratory, removed from their native environment. If the required conditions cannot be supported, then those bacteria simply remain invisible to culturing. However, these bacteria are not completely lost to microbial dark matter. Indirect identification methods through detection of what these bacteria leave behind (DNA) exist through sequencing technologies. These culture-independent techniques have greatly expanded microbial taxonomies and have allowed a greater appreciation of the diversity within the urinary microbiome.

Identification of bacteria through DNA sequencing began in the 1980's when it was first established that ribosomal DNA behaved as a molecular chronometer or an evolutionary timepiece of sorts (23). From this observation, a specific portion of ribosomal DNA, the 16S rRNA gene, could be related directly to a particular bacterium's taxonomic lineage (24). Sequencing small regions of variability within this gene allowed for differentiation between

organisms at the species level, revolutionizing the field of bacterial phylogenomics. This practice of bacterial identification is still used today in a method called 16S rRNA amplicon sequencing that combines polymerase chain reaction (PCR) DNA amplification with next generation sequencing technologies (25). This biased method is effective in evaluating the diversity of microbial samples, even in human urine samples (26). While the use of next generation sequencing in 16S rRNA amplicon sequencing allows for greater depth or coverage of DNA, it also limits the sequencing breadth, restricting bacterial identification to the genus or family level. An alternative method, referred to as shotgun sequencing, takes an unbiased approach of sequencing microbial DNA to assess microbial diversity while also sampling the microbial genetic contents (27). These two methods together make up much of the genomic data that is analyzed through metagenomics research, leading the study of human microbiomes.

A major limitation in these culture-independent identification methods, however, is the requirement that sequenced genes be compared against a reference database for bacterial identification. In other words, if a DNA portion is sequenced and does not match any known bacterial reference sequence, then it falls into the category of bacterial dark matter. To address this gap, more advanced culture-dependent methods (metaculturomics) have been developed to identify purified bacteria isolates that give rise to this bacterial dark matter (28, 29).

In this manner, new microbial species discoveries have followed the widespread use of DNA sequencing with extensive reorganization of microbial taxonomies. Whereas two species may be completely identical in physical appearance, they may be differentiated by DNA content (25). This new concept of establishing species identity by DNA follows a complex history of microbial classification woes and disagreements whether they be discrimination by

morphology, biochemical behavior, serotyping, or antibiotic inhibition patterns (30). **Table 1** outlines several sequencing-based microbial differentiation methods with current standards of species cutoffs. These standards are not mutually exclusive, vary widely by microbial taxon, and are constantly undergoing reevaluation and controversy (31, 32, 33, 34).

Table 1. Sequencing-based Identification Methods and Cutoffs

Identification Method	Species Similarity Threshold
DNA-DNA Hybridization (DDH)	70%
16S rRNA	97-98%
Average Nucleotide Identity (ANI)	94-96%
Multi-locus Sequence Typing (MLST)	species dependent

DNA-DNA Hybridization evaluates genetic similarity by experimentally hybridizing DNA single strands from a query and reference organism to each other and comparing the temperature at which the DNA-DNA hybrid melts with the temperature of the reference organism. This method can be simulated digitally in a method referred to as digital DNA-DNA Hybridization. 16S rRNA analysis compares variable regions of a query 16S rRNA gene sequence with reference sequences. Average Nucleotide Identity (ANI) identification involves a computational analysis and similarity scoring of lineage-conserved genes between a query and reference genome; ANI analysis outputs can vary depending on the bioinformatic algorithms used. Finally, Multi-locus Sequence Typing (MLST) involves sequencing comparison of various species-defined housekeeping genes.

At some point or another, each of the above methods have claimed themselves to be the “gold standard” of bacterial identification methods when, in practice, they each have their own flaws and shortcomings (35, 36, 37, 38). DDH is often described as laborious in

methodology, and results vary on whether one genome is used as the query or reference (39). 16S rRNA analysis depends heavily on the choice of variable region used for identification with variation being unequal in effectiveness between regions for different organisms (40). ANI suffers from a similar problem with variation within different microbial genera varying unequally (41). Finally, MLST is species-formulated leaving unannotated species in the dark. Ultimately, the choice of identification method lies in practicality and purpose. Some taxonomies go as deep as strain or sub-species-level differentiation often in the interest of medical diagnosis. As new bacteria are identified, their contribution towards health and disease greatly determines the strategies used in their identification.

The Human Urinary Microbiome

It is no surprise that among the bacteria on Earth, those that pose existential threats to human survival are the most studied. Human biology is inherently inseparable from microbial influence both in sickness and in health. These bacteria are collectively referred to as the human microbiome and are at the forefront of many medical research disciplines (42). Understanding the roles these bacteria hold within the human body is key to both diagnosis and therapy.

With the human urinary tract now understood to be home to its own microbiome, studies have been extensively conducted assessing its role in various urinary diseases. While some diseases such as pyelonephritis and UTIs have more obvious microbial infectious etiologies, diseases such as urgency urinary incontinence (UUI), interstitial cystitis, and even cancer have unclear connections with the urinary microbiome. With the advent of

metagenomics and bioinformatics, interest in elucidating these connections has been expanding as more and more studies contribute to the growing urinary microbiome field.

The earliest studies utilizing the new metagenomic techniques primarily focused on females and the ability to identify bacteria in asymptomatic individuals (28, 43, 44, 45, 46, 47). The observation that the standard urinary culture (SUC) method used by clinical microbiology laboratories around the world was missing whole categories of microbes in both symptomatic and asymptomatic females led to the development of a more comprehensive culturing protocol, called the expanded quantitative urine culture (EQUC) method (48). This method provided for more comprehensive microbial growth conditions including greater culturing volumes, longer incubation times, various growth media, and various atmospheric conditions. After culturing, colonies are isolated and identified through a mass spectrometry technique called matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS). Studies utilizing this new method found greater representation of fastidious and anaerobic bacteria that previously could only be detected via DNA sequencing (49, 50). Thus, investigations have been reworking several hypotheses from the ground up as previous diagnostic standards utilizing SUC are overturned. For example, it has long been held that *Escherichia coli* is the prevailing primary causative agent of UTIs (51). While this remains true, its prevalence (~50%) in adult females with UTI-like symptoms is considerably less than originally thought (70-90%) (52). Furthermore, studies utilizing EQUC and metagenomic techniques have found that detecting *E. coli* alone is a weak predictor of UTI symptoms (53).

As such, investigations have been establishing relationships between particular microbiome profiles and diseases including UTIs (54, 55, 56), UUI (57), urethritis (58), bladder

cancer (59), and others. These efforts have culminated in a greater appreciation of the diversity within the human urinary microbiome with nearly 200 different species identified so far (60). But establishing correlative relationships between species identities and urinary diseases does not get to the root of pathophysiology. Mechanistic studies identifying pathogenic microbial behaviors are needed to arrive at corresponding countermeasures and therapies.

Emerging Uropathogens

The concept of bacteria being integral to human health is relatively new and almost antithetical to the original founding ideas of germ theory – that microbes are the etiological agents of infectious disease. With the gut microbiome field leading the way, there has been a drastic paradigm shift away from assigning blame to any single microbe as the causative agent in a patient's sickness (61). Instead, an individual's microbiome is often compared to the role of any other human organ; individual microbes collectively contributing towards nutrition, immunity, regulation, and homeostasis (62). An imbalance in steady-state functioning or shift towards a pathogenic state is termed dysbiosis, and its defining composition varies by microbiome niche, person, and disease (63). Annotating the behaviors of bacteria that make up a healthy or diseased state has been a monumental task with a myriad of variables to consider.

Further complicating things is the observation that bacteria individually may have varying contextual behaviors. While some bacterial species may be described as pathogens, they may still be found in "subclinical" amounts in healthy individuals. At the same time, some bacteria may be considered normal flora, but can transition into opportunistic pathogens when conditions change (64, 65). Within the urinary microbiome, classifying organisms categorically as either "good" or "bad" in terms of human health can be a fool's errand due to this varying

nature of microbial behavior. Instead, researchers in this space cluster bacterial communities by association patterns often termed community state type (CST) or urotype to compare between individuals (66, 47, 67).

In this way, bacteria whose roles were previously completely unknown or unannotated are now able to be evaluated by association as more or less likely to be involved in certain disease states. Many of these species are labelled as “emerging uropathogens” with their poor representation in the literature often due to their rarity, difficulty in isolation, lack of surveillance, and/or poor identification standards (68, 69). Many of these understudied uropathogens belong to the bacterial families of Aerococcaceae, Actinomyceteceae, and Bifidobacteriaceae with some species only having been given names within the last few decades. One of the greatest emerging threats among these is a species called *Aerococcus urinae* (70, 71).

Aerococcus urinae **in Humans**

As the name would suggest, *Aerococcus urinae* is often implicated with the urine of humans. The circumstances and implications of how it ends up there, however, have so far remained a mystery. The natural reservoir of the bacterium (if not humans) is poorly described, and the circumstances in which it becomes pathogenic are uncharacterized. Indeed, the bacterium is often referenced as an emerging uropathogen, indicating its understudied status as a potential causative agent of disease (72). Currently, there is a demonstrative need for greater investigation as to *A. urinae*’s involvement in diseases, such as urinary tract infections, urge urinary incontinence, and invasive tissue infections. With increasing antibiotic resistance

annotated in clinical isolates, *A. urinae* poses a growing threat to the undiagnosed (and misdiagnosed) patient.

Initially described in 1992, the first isolates of *A. urinae* came from urines of patients diagnosed with urinary tract infections (73). Originally thought of as a rare cause of human infection, the bacterium has since seen a clear rise in diagnoses and case reports alongside improvements in culture techniques and identification technologies (74, 75, 76, 77). While lethal cases are rare, *A. urinae* has been identified in a variety of severe disease complications, such as soft tissue infections and bacteremia, all tracing to a urological origin (74, 78, 79, 80). Non-invasive infections are associated with cystitis and UUI in females (81, 82). Culturing of urine from healthy patients has also identified positive colonization of the bacterium, complicating *A. urinae* as an opportunistic pathogen (83). Monoculture of the bacterium from urine is uncommon with it often identified alongside several other species, contributing to its dismissal as a contaminant. Whether this bacterium works in concert with others or on its own in pathogenic settings is unclear.

Risk factors for invasive infections include older age and comorbid genitourinary diseases (77, 80, 84). In cases of bacteremia, the majority of infections are monomicrobial with significant risk for endocarditis and septic embolization (85, 86). In pediatric settings, *A. urinae* has been reported as a cause for extraordinary malodorous urine in boys with comorbid urogenital disorders as a risk factor (87, 88). Malodorous urine has been documented in adult cases as well, having been described as ammoniacal and “socially disabling” (74, 89). In all severe cases of infection, correct identification is vital whereas misidentification and lack of resistance testing can lead to fatality (90, 91).

A. urinae is easily missed on routine bacteriological tests and often misidentified as streptococci, staphylococci, or enterococci due to sharing many similar characteristics. Currently, the most common test for rapid identification in the clinical setting is via MALDI-TOF MS (92, 93). *Aerococcus* species share colony morphologies with other alpha-hemolytic streptococci with misidentification contributing to under-reporting in patient culture samples (94). This problem is so prevalent that Facklam and co-authors reported in 2003 that all sixteen conventional bacterial identification tests used by the Centers for Disease Control and Prevention (CDC) and the National Center for Streptococcus (NCS) failed to properly identify *A. urinae* reference strains (95).

Whole genome sequencing and phenotypic characterization of the organism has revealed substantial diversity within the *A. urinae* species designation such that subdivision has been suggested (96, 97, 98), although the clinical relevance of such divisions is unknown. As is found in other invasive uropathogens, *A. urinae* demonstrates the ability to form biofilms on catheters and heart tissue, as well as the ability to aggregate platelets (99, 100, 101). Analysis for virulence factors revealed genes whose homologs are associated with adhesion and anti-phagocytosis (97). Proteomic studies have supported this association, revealing an abundance of adhesive surface proteins expressed on the bacterium's surface (99, 102). Unfortunately, no genetic model currently exists to allow mechanistic studies into these virulence factors.

With proper identification and susceptibility testing, antibiotic therapy proves effective in treatment. Isolates from several studies have demonstrated susceptibilities to most antibiotics used against Gram-positive organisms; however, resistances have been indicated to fluoroquinolones, cephalosporins, trimethoprim-sulfamethoxazole, and tetracycline (103, 104,

105, 106, 107). There is concern that antibiotic resistance may be increasing as detection in wastewater samples indicates rising resistance in *Aerococcus* (108). As with other streptococci, staphylococci, and enterococci, the possibility of horizontal gene transfer of resistance genes may pose a significant risk for the future. Fellow genus member *Aerococcus urinaeequi* has been found to harbor a plasmid that carries a gene for tetracycline resistance and a transposable element that carries an operon for vancomycin resistance (109, 110). As such, prudent stewardship paired with proper identification is imperative in the diagnosis and treatment of *A. urinae* infections moving forward.

Hints of *Aerococcus urinae* Strain Diversity

In 1992, *Aerococcus urinae* was the second species to be named within the *Aerococcus* genus after *Aerococcus viridans*. It was described as Gram-positive, alpha-hemolytic, non-motile, cocci that appear in pairs or tetrads (73). The basis of the discovery was the genetic and phenotypic uniqueness of several *Aerococcus*-like organisms that differed from *A. viridans*. The isolates had come from urine samples obtained from patients diagnosed with UTIs, and it was theorized that *Aerococcus*-like organisms were the rare cause of human infection, partly due to how rarely the organisms were detected. Since then, seven additional *Aerococcus* species have been named, all of which have been isolated from either livestock or human hosts (**Figure 1**).

The rapid expansion of the genus can be attributed to the improvement in identification techniques and sequencing technologies that continue to be developed (83). Indeed, the rarity of *Aerococcus* isolates from human infections may have been overestimated in the past simply due to the inability to distinguish them from related genera using conventional methods (94). As more adept methods for isolation and identification of fastidious organisms have been

utilized, larger collections of *Aerococcus* strains have afforded a greater appreciation of the genetic heterogeneity both within extant species designations, as well as for those yet to be named.

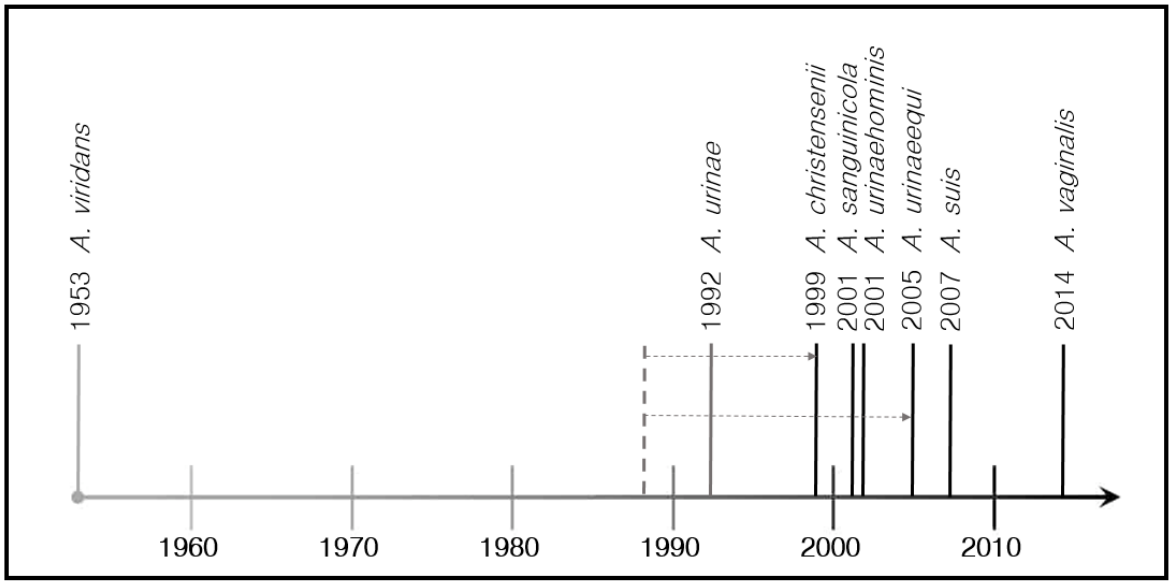


Figure 1. Timeline of Discovery of Novel *Aerococcus* Species. The genus began with *A. viridans* "to make green" by Williams et al. in 1953. After the discovery of *A. urinae* "of urine" by Collins et al. in 1992, the genus began rapid expansion: *A. sanguinicola* "blood dweller" by Lawson et al., *A. urinaehominis* "of a human urine" by Lawson et al., *A. suis* "of a hog" by Vela et al., *A. vaginalis* "of a <cow> vagina" by Tohno et al. Note: *Streptococcus acidominimus* and *Pediococcus urinaeequi* were both discovered in 1988 but were reclassified as *A. christensenii* in 1999 and as *A. urinaeequi* in 2005, respectively.

The five isolates used to name the *A. urinae* species in 1992, based on phenotypic and 16S rRNA gene sequence analysis data, came from the work of Dr. Jens Christensen and co-workers, who originally recognized the bacterium as a potential urinary tract pathogen (111). The strains were part of a Danish national study of 63 strains from patients of whom the majority presented with symptoms of UTI (112). Later in 1997, DDH studies and metabolic phenotype comparison on a collection of 22 international strains from the United States, Canada, and Denmark revealed the existence of a novel biotype within *A. urinae*, distinct from

that previously described in 1992 (96). These authors were the first to provide evidence that A. *urinae* may consist of two or more species. This was expanded upon in 2005 with phenotypic tests, ribotyping, and multilocus enzyme electrophoresis, again demonstrating the existence of a distinct group within the *A. urinae* designation (113). However, the lack of strains and inability to distinguish groups via 16S rRNA gene sequencing caused the authors to hesitate to formally propose a new species. This sentiment was echoed by Felis and co-workers in 2005, when *Pediococcus urinaeequi* was reclassified to *Aerococcus urinaeequi* with those authors noting that 16S rRNA gene sequence identity alone was not enough to guarantee species identity (114). Indeed, both Felis and Christensen and their co-authors found much greater discriminatory ability with the use of DDH methods. It would not be until 2016, when the type strains of *Aerococcus* species would be fully sequenced, that whole genome sequences could be compared (115). Accordingly, Zhou and co-authors were able to demonstrate that ANI analysis based on whole genome sequencing is a more effective means for discrimination between *Aerococcus* species, such as when comparing an *A. urinaeequi* strain and *A. viridans* strain, despite sharing 99.9% 16S rRNA gene sequence (110).

The first comparative genomics study of *A. urinae* strains was conducted in 2017, with whole genome sequencing revealing substantial genetic diversity within the species designation (97). The pan- and core-genomes of 40 strains of *A. urinae* were analyzed, indicating intra-species clustering by isolation year period. However, all 40 strains were isolated from patients in Denmark, and it remains to be seen whether the study's findings are representative of isolates from different countries. Perhaps the greatest hurdle thus far in assembling such

isolates has been the first step, identification. Only with the advent of MALDI-TOF MS has the ability to rapidly and reliably identify *Aerococcus* species been realized (92, 93).

As such, utilizing MALDI-TOF MS coupled with EQUC, a large collection of *A. urinae* isolates from patient samples were curated along with whole genome sequencing at Loyola University Chicago (116). In 2020, twenty-four of these strains were characterized, and along with publishing their genomes, Dr. Evann Hilt suggested that a potential association may exist between phylogeny and the ability to form flocking biofilms (98). In summary of her analysis, no correlation was found between patient lower urinary tract symptoms (LUTS) and cultured CFUs/mL or patient urine clinical markers (pH, turbidity, and color). As such, no strong inferences could be made on these parameters to differentiate between infectious and non-infectious *Aerococcus* strains. However, she hypothesized that *in vitro* phenotypes could instead serve to identify *Aerococcus* isolates that are involved in disease. She went on to describe three unique phenotypes that these isolates displayed including a 'hockey puck' phenotype relating to colony morphology, a 'pigmentation' phenotype involving production of a black pigment, and finally a 'flocking' phenotype describing a behavior of biofilm aggregates. She hypothesized that these behaviors were likely related to a strain's pathogenic potential involved in persistence, quorum sensing, and adherence, respectively.

Combining these strains and all publicly available *A. urinae* genomes, a total of 77 genomes were utilized to perform whole genome ANI analysis, from which we reported that the genomes could be clustered into five distinct clades (117). It is currently unknown what characteristics may distinguish these *Aerococcus* groupings and to what, if any, clinical significance they may have on human disease.

Summary of Introduction

There is certainly a sense of irony in the observation that some of the most unknown forms of life to humans are those that are most intimately intertwined in our daily functioning. Aristotle is credited as saying, “knowing yourself is the beginning of all wisdom,” and indeed, understanding our microbiome may be the key in unlocking our hidden potential as humans. But the mysteries of the human microbiome, particularly the urinary microbiome, remain uncharacterized and understudied.

The science of studying this field has come a long way. From Petri dishes to metagenomics, the ability to observe, isolate, and study members of the urinary microbiome has been a major bottleneck. The greatest advances have been due to the development of sequencing technologies, bypassing the need to isolate urinary organisms alive from inside the human body. Instead, DNA evidence has provided proof that the bladder is not sterile, leading to the appreciation of significant microbial diversity within this space.

The roles these microbes play within human health and disease have been difficult to parse, partly due to the difficulty in observing these organisms in the laboratory environment. Some of the most poorly annotated are those called emerging uropathogens whose significance have often only been determined by microbial community association or urotype. One species in particular, called *Aerococcus urinae*, has been implicated in a range of urological diseases with cases of fatalities increasingly reported. As hints emerge of this organism behavioral diversity, it has been hypothesized that *A. urinae* may actually represent several distinct groups of separate organisms that warrant further distinction and characterization.

Thus, the following experiments and analyses assesses this diversity, and focuses on disambiguating the species complexity in the context of taxonomy, phylogenetics, and clinical significance.

CHAPTER TWO

EXPERIMENTAL RESULTS

Meta-analysis of *Aerococcus urinae* Complex Identification in Patients with LUTS

Previous studies have observed increased frequencies in detection of *Aerococcus* in patients with LUTS, particularly in patients diagnosed with UUI (81, 118). However, these studies were focused on single patient-diagnosis groups. To understand the detection frequency of *Aerococcus* relative to other microbes in the broader context of patients with LUTS, I conducted a meta-analysis of eight clinical studies with a combined total of 1007 individuals. **Table 2** shows the detection of microbes in individuals broken down by LUTS diagnosis: UTI, UUI, stress urinary incontinence (SUI), interstitial cystitis / painful bladder syndrome (IC/PBS), and healthy control.

Although it is unknown if any disease etiologies of these participants can be attributed to any particular microbe(s) from this analysis, it is of significance to note the increased detection rate of certain microbes, both known uropathogens and emerging uropathogens. For example, we can observe that the detection rate of known uropathogen *Klebsiella sp.* at 11.51% for UTI patients is comparable to the detection rate of *Aerococcus sp.* at 13.49%. Within the UUI category, several genera appear elevated in detection including *Aerococcus* (36.36%), *Corynebacterium* (51.78%), *Lactobacillus* (56.13%), *Staphylococcus* (45.85%), and *Streptococcus* (63.64%). It is also important to note that *Aerococcus* detection is not limited to LUTS, but also

found in healthy individuals, indicating that *Aerococcus* may have a role in both health and disease states in humans.

Table 2. Frequency of Microbe Identification via Metaculturomics in Patients with LUTS

Cells are shaded red relative to the highest value (63.64%). Multiple members of the same species occurring within the sample are included as a single count. Percentages within each disease category are not cumulative to 100% due to polymicrobial samples.

Microbe	Total N=1007	**UTI** N=304	**UUI** N=253	**SUI** N=50	**IC/PBS** N=49	**Control** N=351
Acinetobacter	0.50%	0.00%	1.98%	0.00%	0.00%	0.00%
Actinobaculum	1.39%	0.99%	4.35%	0.00%	0.00%	0.00%
Actinomyces	10.13%	7.57%	22.53%	4.00%	6.12%	4.84%
Actinotignum	5.26%	3.95%	13.04%	6.00%	0.00%	1.42%
Aerococcus	16.29%	13.49%	36.36%	16.00%	10.20%	5.13%
***Aerococcus urinae* Complex**	**14.20%**	**11.18%**	**33.60%**	**14.00%**	**8.16%**	**3.70%**
Alloscardovia omnicolens	6.65%	4.93%	13.83%	8.00%	4.08%	3.13%
Bacillus	0.40%	0.33%	0.40%	0.00%	0.00%	0.57%
Bifidobacterium	5.06%	4.28%	9.88%	0.00%	8.16%	2.56%
Brevibacterium	3.28%	2.30%	8.70%	2.00%	2.04%	0.57%
Candida	3.18%	1.97%	7.11%	4.00%	2.04%	1.42%
Citrobacter	0.60%	1.64%	0.40%	0.00%	0.00%	0.00%
Corynebacterium	22.34%	15.13%	51.78%	26.00%	16.33%	7.69%
Cutibacterium	1.19%	0.99%	2.77%	0.00%	4.08%	0.00%
Dermabacter hominis	0.79%	0.33%	2.37%	2.00%	0.00%	0.00%
Enterobacter	1.79%	0.33%	5.53%	0.00%	0.00%	0.85%
Enterococcus	11.42%	8.88%	23.72%	8.00%	8.16%	5.70%
Enterococcus faecalis	11.12%	8.88%	22.92%	8.00%	8.16%	5.41%
Escherichia coli	24.83%	50.99%	25.69%	12.00%	8.16%	5.70%
Facklamia hominis	4.07%	1.32%	12.25%	2.00%	0.00%	1.42%
Gardnerella	14.10%	11.84%	20.95%	12.00%	2.04%	13.11%
Gemella	0.40%	0.00%	1.58%	0.00%	0.00%	0.00%
Globicatella	0.50%	0.00%	1.98%	0.00%	0.00%	0.00%
Haematomicrobium	0.30%	0.99%	0.00%	0.00%	0.00%	0.00%
Haemophilus	0.89%	0.33%	1.58%	4.00%	2.04%	0.28%
Klebsiella	5.76%	11.51%	6.72%	0.00%	2.04%	1.42%
Klebsiella pneumoniae	4.57%	8.22%	5.93%	0.00%	2.04%	1.42%
Kocuria	0.30%	0.00%	0.79%	0.00%	2.04%	0.00%
Lactobacillus	37.24%	35.53%	56.13%	42.00%	30.61%	26.21%
Micrococcus	3.38%	0.99%	7.11%	0.00%	2.04%	3.42%
Moraxella	0.30%	0.00%	0.79%	0.00%	0.00%	0.28%

Morganella	0.50%	0.33%	1.19%	0.00%	2.04%	0.00%
Neisseria	0.89%	0.66%	1.19%	6.00%	0.00%	0.28%
Oligella	1.19%	0.66%	3.56%	2.00%	0.00%	0.00%
Peptoniphilus	0.50%	0.00%	1.98%	0.00%	0.00%	0.00%
Prevotella	0.30%	0.00%	0.79%	2.00%	0.00%	0.00%
Proteus	2.38%	4.28%	3.95%	0.00%	0.00%	0.28%
Pseudoglutamicibacter	3.67%	1.64%	10.67%	0.00%	4.08%	0.85%
Pseudomonas aeruginosa	1.29%	1.97%	2.77%	0.00%	0.00%	0.00%
Rothia	1.79%	0.66%	2.77%	10.00%	0.00%	1.14%
Staphylococcus	22.44%	16.45%	45.85%	30.00%	20.41%	9.97%
Coagulase Negative	21.05%	14.47%	45.06%	24.00%	20.41%	5.41%
Coagulase Positive	2.38%	2.30%	3.56%	8.00%	0.00%	0.28%
Streptococcus	36.35%	28.29%	63.64%	40.00%	28.57%	24.22%
Streptococcus viridans grp.	13.90%	7.89%	22.92%	14.00%	14.29%	4.56%
Streptococcus anginosus grp.	23.24%	16.45%	48.62%	26.00%	12.24%	6.84%
Streptococcus agalactae	8.04%	7.89%	11.86%	10.00%	10.20%	4.84%
Trueperella bernardiae	2.09%	0.99%	5.53%	4.00%	0.00%	0.57%
Winkia neuii	9.43%	7.24%	21.74%	10.00%	4.08%	3.13%
Unknown	24.03%	17.76%	57.71%	18.00%	2.04%	9.12%

As with other diseases that involve microbes, it is dangerous to assign any single species or taxon as a lone etiological cause. For example, in the gut microbiome field, the theory of gut microbiome dysbiosis as a contributing factor towards complex diseases, such as Crohn's disease or inflammatory bowel disease, has been gaining favor over blaming any single species (119). Instead, it is more prudent to look for patterns in groupings of microbial taxa, sometimes referred to as community state types, that describes how strong or weak associations are for two or more microbes to co-occur. Within the meta-analysis, I analyzed samples for the species that most commonly co-occur with *Aerococcus* species. I found that members belonging to the Actinomycetaceae family were the most commonly co-identified, specifically of the genera *Actinotignum, Actinomyces,* and *Winkia* (**Figure 2**). 48% of all samples where members of *Aerococcus* species co-occurred with at least one species in the family Actinomycetaceae, indicating a strong relationship between these taxa.

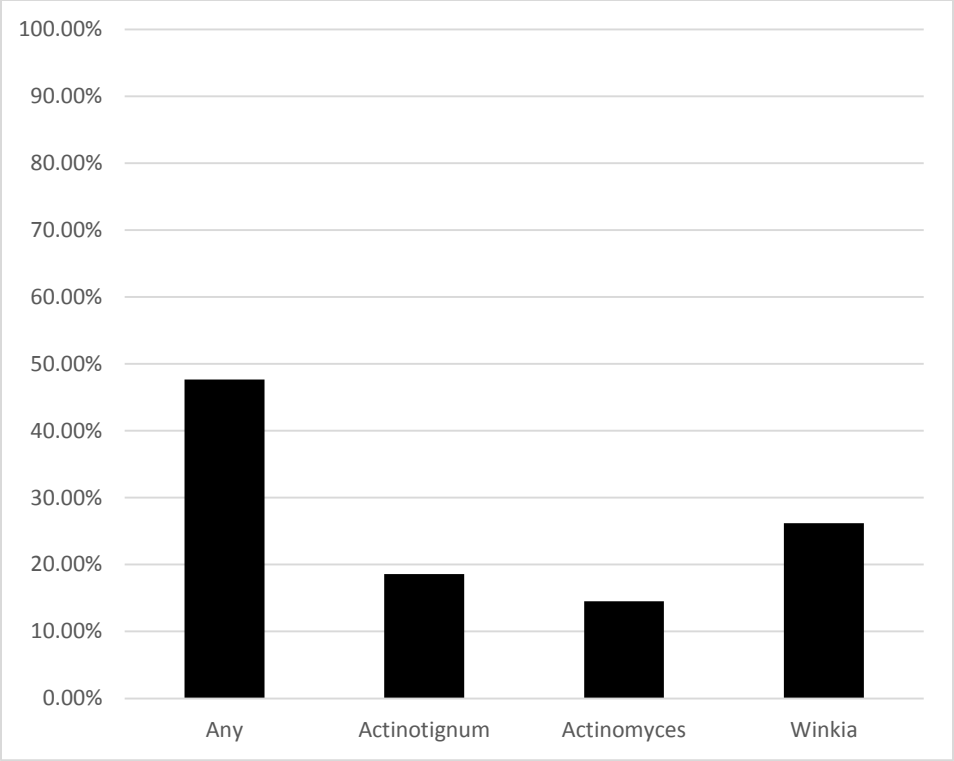


Figure 2. Co-identification Rate of *Aerococcus sp.* and Actinomycetaceae *sp.* Total n=172 samples with *Aerococcus sp.* identified.

As previously noted, although *Aerococcus* isolates are identified at higher frequencies in patients with LUTS, they are also found in asymptomatic individuals. Thus, *Aerococcus sp.* may be opportunistic, wherein an environmental stimulus can lead to a pathogenic behavioral shift (72). Additionally, certain strains may be more or less likely to lead to disease due to inherent bacterial genetic factors. A good example of this scenario is *Salmonella*, of which there are over 2500 serotypes but only one serotype that results in serious typhoid fever (120). The investigations that follow are structured around these two lines of reasoning, although not mutually exclusive, in querying the differences between *Aerococcus* strains.

Phylogenetic Investigation

The *Aerococcus* isolates from the studies that went into the previous meta-analysis are described in further detail by Dr. Evann Hilt (121). In sequencing 24 clinical *Aerococcus* isolates

identified as *A. urinae*, Dr. Evann Hilt found that those strains displaying the strongest flocking biofilm phenotype also clustered based on genotype, indicating a potential genetic link with the behavior. The ability to form biofilms in particular is commonly associated with pathogenesis, such as through tissue adherence, antibiotic resistance, and immune evasion (122). Additionally, several studies have described how biofilm behavior has been associated with disease pathologies, manifesting as adherence to catheter tips in catheter-associated UTIs (99) or valve adherence in infective endocarditis (100, 101). In collaboration with Dr. Nicole Gilbert, I assisted in identifying representatives from each of the aforementioned *A. urinae* genotype groups that persisted within a mouse model of bacteriuria where strains with greater biofilming ability persisted at higher CFU/mL in the mouse bladder compared to strains with lesser biofilming ability (117).

Using these same genetic groups as a basis, I conducted a larger scale investigation involving the whole genome sequencing of 115 strains, previously identified as *A. urinae,* to evaluate genomic phylogeny. A complete list of genomes analyzed is outlined in **Table 3** annotating genomic features, isolation source and location, and accession number in the NCBI database.

The most common method of analyzing bacterial phylogeny via DNA is by comparing 16S rRNA sequences. Comparison of representative strains from each of the proposed genomic groups shows some, but only faint relative distinction when arranged in a phylogenetic tree (**Figure 3A**). However, upon conducting average nucleotide identity analysis (ANI) of whole genomes, it was discovered that a large degree of diversity could be appreciated, such that major groupings of the genomes emerged.

Within those genomes labelled as *A. urinae*, six distinct taxonomic groups could be formed based on a cutoff standard of 95% ANI. *A. urinae* ATCC 51268^T strain, the type strain for the species, is contained in one clade, which is labelled as *A. urinae* (**Figure 3B, red**). These *A. urinae* strains are uniformly distinct from three other named clades: *A. mictus* (**Figure 3B, purple**), *A. tenax* (**Figure 3B, green**), and *A. loyolae* (**Figure 3B, orange**). A fourth and fifth taxonomic group from the *A. urinae* clade are also identified (**Figure 3B, dark and light blue**), but they are not named here as currently few genomes are available. For now, they have been labelled as *A. sp.* Group 1 and *A. sp.* Group 2. Two strains, UMB0553 and UMB8614 could not be definitively placed into any group and have been left unassigned. The average ANI within each of these clades is as follows: *A. tenax* = 98.79%, *A. mictus* = 98.37%, *A. loyolae* = 99.02%, and *A. urinae* = 99.96%. These values suggest that the *A. urinae* group demonstrates the highest homogeneity, while the *A. mictus* group demonstrates the highest heterogeneity.

Because the ANI percentages between these proposed genomic groups is at or below the 95% threshold, they meet bioinformatic standards for species distinction. Genomes belonging to the new *A. urinae* group have an average shared ANI percentage with the *A. mictus* group at 92.6%, the *A. tenax* group at 94.0%, the *A. loyolae* group at 92.2%, the unnamed species Group 1 at 93.7%, and the unnamed species Group 2 at 93.1%, all below the 95% threshold. As such, the three new named species have been published in the *International Journal of Systematic and Evolutionary Microbiology* (IJSEM), the official publication of the International Committee on Systematics of Prokaryotes (ICSP) which is the international forum for the publication of new prokaryote species (123). In compliance with rules, type strains have been deposited at international repositories under the following aliases: *A. tenax* UMB3669 =

CCUG 76531^{T} = NR-58630^{T} = ATCC TSD-302^{T} = DSM 115700^{T}, *A. mictus* UMB3440 = CCUG 76532^{T} = NR-58629^{T} = ATCC TSD-301^{T} = DSM 115699^{T}, and *A. loyolae* UMB0080 = CCUG 76533^{T} = NR-58628^{T} = ATCC TSD-300^{T} = DSM 115698^{T}. Together, these new groups form the *Aerococcus urinae* species complex, collectively referred to from now on in this manuscript as *A. urinae* complex (AUC).

Table 3. *Aerococcus* Genomes Analyzed

Taxonomic Designation	Strain Name	Accession No.	Genome Size	No. tRNA	No. rRNA	No. CDS	Isolation Site	Source Laboratory	Isolation location	BioProject
A. loyolae	AU3	GCF_001649715.1	1935027	34	4	1753	Blood	E Senneby, Lund, Sweden	Europe	PRJNA315093
A. loyolae	CDC-1515-U85	GCA_026694535.1	1975558	58	3	1786	Urine	R Facklam, CDC, Atlanta, USA	USA	PRJNA876651
A. loyolae	CDC-3352-U95	GCA_026694515.1	2017114	58	3	1831	Urine	R Facklam, CDC, Atlanta, USA	USA	PRJNA876651
A. loyolae	LUND-40-B12	GCA_026694525.1	1986294	58	3	1806	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. loyolae	HMSC075D05	GCF_001811135.1	1893491	38	2	1765	Urine	WUSM, St. Louis, United States	USA	PRJNA299938
A. loyolae	UMB0080	GCF_002871915.2	1975336	53	3	1793	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB0088	GCF_002884955.2	1966441	58	6	1794	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB0126	GCF_002847705.1	1977544	55	3	1801	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB0232	GCF_002847625.1	1975554	58	4	1795	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB0509	GCF_003286825.2	1915973	56	5	1740	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB2126	GCF_008726675.1	1911344	56	2	1745	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB5628	GCF_003286585.1	1957540	56	2	1798	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB7480	GCF_003286555.1	2205693	55	2	2111	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. loyolae	UMB8711	GCF_008726285.1	1921934	57	2	1750	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	SLA-40126-U13	GCA_026695825.1	2098146	58	2	1903	Urine	JJ Christensen, Slagelse, Denmark	Europe	PRJNA876651
A. mictus	SLA-43350-U13	GCA_026695805.1	2035364	58	4	1836	Urine	JJ Christensen, Slagelse, Denmark	Europe	PRJNA876651
A. mictus	SLA-43565-U13	GCA_026695785.1	2010684	58	2	1845	Urine	JJ Christensen, Slagelse, Denmark	Europe	PRJNA876651
A. mictus	SLA-45893-U13	GCA_026695765.1	2102359	58	5	1912	Urine	JJ Christensen, Slagelse, Denmark	Europe	PRJNA876651
A. mictus	SLA-48199-U13	GCA_026695745.1	2045284	58	3	1861	Urine	JJ Christensen, Slagelse, Denmark	Europe	PRJNA876651
A. mictus	SLA-48590-U13	GCA_026695725.1	2104893	59	4	1932	Urine	JJ Christensen, Slagelse, Denmark	Europe	PRJNA876651
A. mictus	LUND-01-B14	GCA_026695685.1	2088823	58	5	1881	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-12-B13	GCA_026695705.1	2083051	58	2	1877	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-14-B13	GCA_026695665.1	1965705	58	4	1763	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-02-B14	GCA_026695635.1	2074620	58	3	1877	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651

Taxonomic Designation	Strain Name	Accession No.	Genome Size	No. tRNA	No. rRNA	No. CDS	Isolation Site	Source Laboratory	Isolation location	BioProject
A. mictus	LUND-20-B13	GCA_026695625.1	2223583	57	2	2045	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-33-B12	GCA_026695585.1	2085192	58	3	1904	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-35-B12	GCA_026695535.1	1958223	58	4	1767	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-04-B14	GCA_026695605.1	2097617	58	4	1905	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-41-B12	GCA_026695525.1	2096923	58	3	1900	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	LUND-44-B12	GCA_026695025.1	2035337	58	4	1823	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. mictus	HMSC062B07	GCF_001809895.1	2150168	49	2	1990	Vaginal/Rectal	WUSM, St. Louis, United States	USA	PRJNA296287
A. mictus	Au-01-U13	GCF_905113045.1	2441113	55	2	2283	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-10-B10	GCF_905125205.1	2010711	53	2	1840	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-13-B13	GCF_905111125.1	2092009	51	2	1935	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-28-U13	GCF_905125235.1	2102795	51	2	1900	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-43-B13	GCF_905112965.1	2112677	58	3	1914	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-51-B15	GCF_905112975.1	1946251	58	3	1759	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-52-U15	GCF_905115165.1	1946262	58	3	1760	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-53-B14	GCF_905115105.1	2070278	58	3	1887	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-54-U14	GCF_905112985.1	2070039	58	3	1887	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-57-B15	GCF_905115155.1	2108572	58	3	1911	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-58-U15	GCF_905115195.1	2089562	58	3	1893	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-59-B15	GCF_905112995.1	2002979	58	4	1828	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	Au-60-U15	GCF_905115145.1	1999833	58	3	1821	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. mictus	UMB0071	GCF_003286525.1	2165045	56	1	2035	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB0072	GCF_002847665.1	2122266	56	5	1976	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB0072b	GCF_002884575.1	2127773	60	5	1967	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB0239	GCF_003286895.1	2198507	56	1	2048	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB0267	GCF_003286875.3	1993357	54	1	1832	Urine	Wolfe, Chicago, United States	USA	PRJNA316969

Taxonomic Designation	Strain Name	Accession No.	Genome Size	No. tRNA	No. rRNA	No. CDS	Isolation Site	Source Laboratory	Isolation location	BioProject
A. mictus	UMB0637	GCF_008726885.1	2018628	55	1	1855	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB1016	GCF_003286735.3	2214030	56	2	2051	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB2325	GCF_003286725.1	2622222	59	2	2726	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB2354	GCF_003286695.1	2122926	56	1	1965	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB2879	GCF_003286665.1	2182601	56	2	1998	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB3440	GCF_003286595.3	2100480	56	4	1917	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB6497	GCF_003286635.1	2092635	56	1	1925	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB7382	GCF_003286565.1	2601092	57	4	2759	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. mictus	UMB8662	GCF_008726315.1	1944087	55	2	1759	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. sp. Group 1	CDC-1656-U92	GCA_026695435.1	1970940	58	3	1822	Urine	R Facklam, CDC, Atlanta, USA	USA	PRJNA876651
A. sp. Group 1	CDC-1871-U94	GCA_026695425.1	2023179	58	2	1847	Urine	R Facklam, CDC, Atlanta, USA	USA	PRJNA876651
A. sp. Group 1	CDC-944-U94	GCA_026695485.1	1971043	58	3	1824	Urine	R Facklam, CDC, Atlanta, USA	USA	PRJNA876651
A. sp. Group 1	LUND-10-B14	GCA_026694505.1	1973647	55	2	1795	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. sp. Group 1	ACS-120-V-Col10a	GCF_000193205.1	2080974	60	12	1946	Urogenital tract	Belgium	Europe	PRJNA51073
A. sp. Group 2	NLD-060-U95	GCA_026694965.1	1938423	58	3	1733	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. sp. Group 2	LUND-16-B13	GCA_026694895.1	1968763	55	5	1760	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. sp. Group 2	LUND-36-B12	GCA_026694865.1	1957449	55	3	1766	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. sp. Group 2	LUND-06-B14	GCA_026694875.1	1937711	55	3	1759	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. sp. Group 2	Au-44-B14	GCF_905125285.1	1933310	58	3	1736	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sp. Group 2	Au-47-U14	GCF_905112955.1	1933470	58	3	1738	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. tenax	Au-06-U13	GCF_905113055.1	1933214	51	3	1766	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. tenax	Au-29a-U14	GCF_905125315.1	1981083	58	3	1804	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. tenax	Au-45-U14	GCF_905115075.1	1978409	58	3	1767	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. tenax	Au-49-B14	GCF_905115085.1	1950400	58	3	1748	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. tenax	Au-50-U14	GCF_905115115.1	1953851	58	3	1753	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767

Taxonomic Designation	Strain Name	Accession No.	Genome Size	No. tRNA	No. rRNA	No. CDS	Isolation Site	Source Laboratory	Isolation location	BioProject
A. tenax	UMB0337	GCF_003286845.2	2051035	56	1	1901	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. tenax	UMB1741	GCF_003286715.1	2403057	65	2	2456	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. tenax	UMB3669	GCF_003286645.3	2036703	56	3	1869	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. tenax	UMB7049	GCF_008726475.2	1943021	55	1	1756	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. tenax	UMB0970	GCF_008726845.1	2023870	55	2	1866	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. urinae	SLA-48243-U13	GCA_026694815.1	1950800	60	3	1809	Urine	JJ Christensen, Slagelse, Denmark	Europe	PRJNA876651
A. urinae	NLD-005-U95	GCA_026694675.1	1957390	58	3	1774	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	NLD-015-U95	GCA_026694665.1	1949922	58	3	1767	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	NLD-022-U95	GCA_026694725.1	1950281	58	3	1773	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	NLD-036-U95	GCA_026694685.1	1936021	58	3	1758	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	NLD-049-U95	GCA_026694645.1	1950789	58	3	1766	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	NLD-059-U95	GCA_026695445.1	1950453	58	3	1775	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	NLD-066-U95	GCA_026695565.1	1949953	58	3	1769	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	NLD-971-U95	GCA_026694485.1	1948855	58	3	1761	Urine	PMH Schuur, Tilburg, The Netherlands	Europe	PRJNA876651
A. urinae	LUND-25-B13	GCA_026695505.1	1958547	57	4	1764	Blood	E Senneby, Lund, Sweden	Europe	PRJNA876651
A. urinae	ATCC_51268	GCF_002087935.1	1949279	58	2	1767	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJNA379934
A. urinae	Au-02-B96	GCF_905115225.1	1942637	55	3	1751	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-03-U96	GCF_905113065.1	2016180	54	5	1825	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-04-B04	GCF_905113075.1	1942231	55	3	1765	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-07-B93	GCF_905111135.1	1942854	55	3	1758	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-08-B04	GCF_905125195.1	1942510	53	3	1765	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-09-B94	GCF_905115135.1	1944119	50	3	1763	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-11-U84	GCF_905125305.1	1946188	53	3	1762	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767

Taxonomic Designation	Strain Name	Accession No.	Genome Size	No. tRNA	No. rRNA	No. CDS	Isolation Site	Source Laboratory	Isolation location	BioProject
A. urinae	Au-12-B98	GCF_905125255.1	1942055	54	3	1757	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-15-B94	GCF_905125225.1	1954450	53	3	1781	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-16-B92	GCF_905115205.1	1943395	53	3	1786	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-17-B94	GCF_905125275.1	1942073	53	3	1759	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-18-B93	GCF_905125265.1	1943166	54	3	1767	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-19-H93	GCF_905125335.1	1945111	55	3	1770	Heart valve	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-20-U84	GCF_905111885.1	1944574	56	3	1760	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-21-B87	GCF_905125295.1	1942356	51	3	1767	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-23-B94	GCF_905115095.1	1942475	55	4	1762	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-26-U84	GCF_905125245.1	1948080	58	3	1761	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-30-B02	GCF_905111875.1	1945136	50	3	1775	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-31-B94	GCF_905125325.1	1934217	53	3	1759	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	Au-33-U84	GCF_905115125.1	1946683	53	3	1759	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. urinae	CCUG36881	GCF_001543175.1	1974262	60	5	1767	Urine	Christensen JJ, Korner B and Kjaegaard H,Denmark: Copenhagen	Europe	PRJNA308559
A. urinae	FDAARGOS_911	GCF_016026975.1	1974256	60	12	1771	Urine	FDA-ARGOS	USA	PRJNA231221
A. urinae	NBRC_15544_=_CCUG_36881_NBRC_15544	GCF_001544335.1	1946900	58	9	1755	Urine	Christensen JJ, Korner B and Kjaegaard H,Denmark: Copenhagen	Europe	PRJNA308559
A. urinae	UMB0621	GCF_003286515.1	1989746	57	2	1840	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. urinae	UMB0722	GCF_003286755.1	1949481	56	3	1766	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. urinaeequi	151250009-4-258-51	GCF_014050445.1	1980389	48	3	1891	International Space Station dining table	Jet Propulsion Laboratory, California Institute of Technology	USA:ISS	PRJNA649272
A. urinaeequi	151250015-1-258-55	GCF_014050435.1	1979209	48	3	1888	International Space Station cupola	Jet Propulsion Laboratory, California Institute of Technology	USA:ISS	PRJNA649272
A. urinaeequi	151250015-2-258-56	GCF_014050425.1	1979874	46	3	1891	International Space Station	Jet Propulsion Laboratory, California Institute of Technology	USA:ISS	PRJNA649272

Taxonomic Designation	Strain Name	Accession No.	Genome Size	No. tRNA	No. rRNA	No. CDS	Isolation Site	Source Laboratory	Isolation location	BioProject
A. urinaeequi	AV208	GCF_001719605.1	2227638	28	8	2101	Abdominal fluid	Nanjing University	China	PRJNA339485
A. urinaeequi	CCUG28094	GCF_001543205.1	2013339	55	5	1818	Horse Urine	D.Claus, DSM, Braunschweig, Germany	Europe	PRJNA308559
A. urinaeequi	DSM_20341_=_CCUG_28094	GCF_000425085.1	1992132	43	12	1781	Horse Urine	D.Claus, DSM, Braunschweig, Germany	Europe	PRJNA308559
A. urinaeequi	T43	GCF_014931015.1	2117928	57	19	1930	Pig nasopharynx	China:tianjin	China	PRJNA668452
A. urinaeequi	USDA-ARS-USMARC-56713	GCF_001518795.1	2054328	56	7	1870	Calf Nares	USA	USA	PRJNA281531
A. urinaehominis	CCUG42038B	GCF_001543245.1	1831400	57	6	1689	Urine	Sweden: Linkoping	Europe	PRJNA308559
A. urinaehominis	DSM_15634	GCF_900103385.1	1780457	37	4	1683	Urine	Sweden: Linkoping	Europe	PRJEB16196
A. viridans	ATCC_11563_=_CCUG_4311	GCF_000178435.1	2005853	41	2	1710	Environmental	United Kingdom: London	Europe	PRJNA308559
A. viridans	CCUG4311	GCF_001543285.1	2199877	55	7	2059	Environmental	United Kingdom: London	Europe	PRJNA308559
A. viridans	FDAARGOS_249	GCF_002083135.2	2003760	58	7	1841	Blood	FDA-ARGOS	USA	PRJNA231221
A. viridans	LL1	GCF_000262085.1	1994039	43	4	1789	Human host	Zhejiang University	China	PRJNA159617
A. viridans	NCTC7595	GCF_900445105.1	2315448	57	21	2218	Environmental	United Kingdom: London	Europe	PRJEB6403
A. viridans	NCTC8251	GCF_900445095.1	2265362	71	24	2144	Environmental	United Kingdom: London	Europe	PRJEB6403
A. viridans	UMB0240	GCF_002871935.1	1921807	55	6	1809	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. christensenii	CCUG28831	GCF_001543105.1	1634920	60	4	1588	Vagina	USA: Washington	USA	PRJNA308559
A. christensenii	DSM_15819_=_CCUG_28831	GCF_001466745.1	1589378	34	3	1563	Vagina	USA: Seattle	USA	PRJNA175732
A. christensenii	KA00635	GCF_001552755.1	1710611	36	3	1761	Vagina	USA: Washington	USA	PRJNA272079
A. christensenii	UMB0844	GCF_002861505.1	1664487	58	5	1678	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. suis	Aerococcus_suis_DSM_21500	GCF_900176325.1	1382195	49	2	1316	Swine brain	Madrid, Spain	Europe	PRJEB20326
A. sanguinicola	As-24-U13	GCF_905109795.1	2102729	60	3	1928	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sanguinicola	As-25-U14	GCF_905125215.1	2099886	60	3	1925	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sanguinicola	As-29b-U14	GCF_905113005.1	2067281	60	3	1856	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sanguinicola	As-34-B09	GCF_905115175.1	2115354	57	3	1947	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767

Taxonomic Designation	Strain Name	Accession No.	Genome Size	No. tRNA	No. rRNA	No. CDS	Isolation Site	Source Laboratory	Isolation location	BioProject
A. sanguinicola	As-41-B14	GCF_905113015.1	2067483	60	3	1864	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sanguinicola	As-46-U14	GCF_905113025.1	2064531	60	4	1863	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sanguinicola	As-55-B15	GCF_905113035.1	2101563	60	3	1919	Blood	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sanguinicola	As-56-U15	GCF_905109855.1	2101065	60	4	1915	Urine	D Carkaci, Slagelse, Denmark	Europe	PRJEB36767
A. sanguinicola	CCUG43001	GCF_001543145.1	2033849	62	4	1833	Blood	Truberg Jensen K, Christensen JJ and Facklam RR, Denmark	Europe	PRJNA308559
A. sanguinicola	UMB0139	GCF_002847725.1	2247792	58	5	2092	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
A. sanguinicola	UMB623	GCF_008726925.1	2138252	59	5	1947	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
Aerococcus sp.	UMB0553	GCF_003286805.1	2075067	58	3	1887	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
Aerococcus sp.	UMB8614	GCF_008726385.1	2071275	55	1	1933	Urine	Wolfe, Chicago, United States	USA	PRJNA316969
Aerococcus sp.	1KP-2016	GCF_002252085.1	2042438	43	6	2052	Blood	Pirogov Russian National Research Medical University	Russia	PRJNA383933
Aerococcus sp.	HMSC035B07	GCF_001836025.1	2087212	55	1	1951	Urine	WUSM, St. Louis, United States	USA	PRJNA296312
Aerococcus sp.	HMSC061A03	GCF_001813115.1	2176154	55	1	2038	Urine	WUSM, St. Louis, United States	USA	PRJNA300046
Aerococcus sp.	HMSC062A02	GCF_001809535.1	2135362	57	2	2032	Vaginal/Rectal	WUSM, St. Louis, United States	USA	PRJNA296233
Aerococcus sp.	HMSC06H08	GCF_001807425.1	2179205	37	1	2029	Urine	WUSM, St. Louis, United States	USA	PRJNA269898
Aerococcus sp.	HMSC072A12	GCF_001811965.1	2173025	46	3	2048	Urine	WUSM, St. Louis, United States	USA	PRJNA299983
Aerococcus sp.	HMSC10H05	GCF_001806975.1	1892427	36	1	1777	Urine	WUSM, St. Louis, United States	USA	PRJNA269853
Aerococcus sp.	HMSC23C02	GCF_001806805.1	2099444	47	1	1866	Urine	WUSM, St. Louis, United States	USA	PRJNA269827
Aerococcus sp.	SJQ22	GCF_003797145.1	2113287	29	4	1964	Soil	Shanghai Academy of Agricultural Sciences	China	PRJNA504512

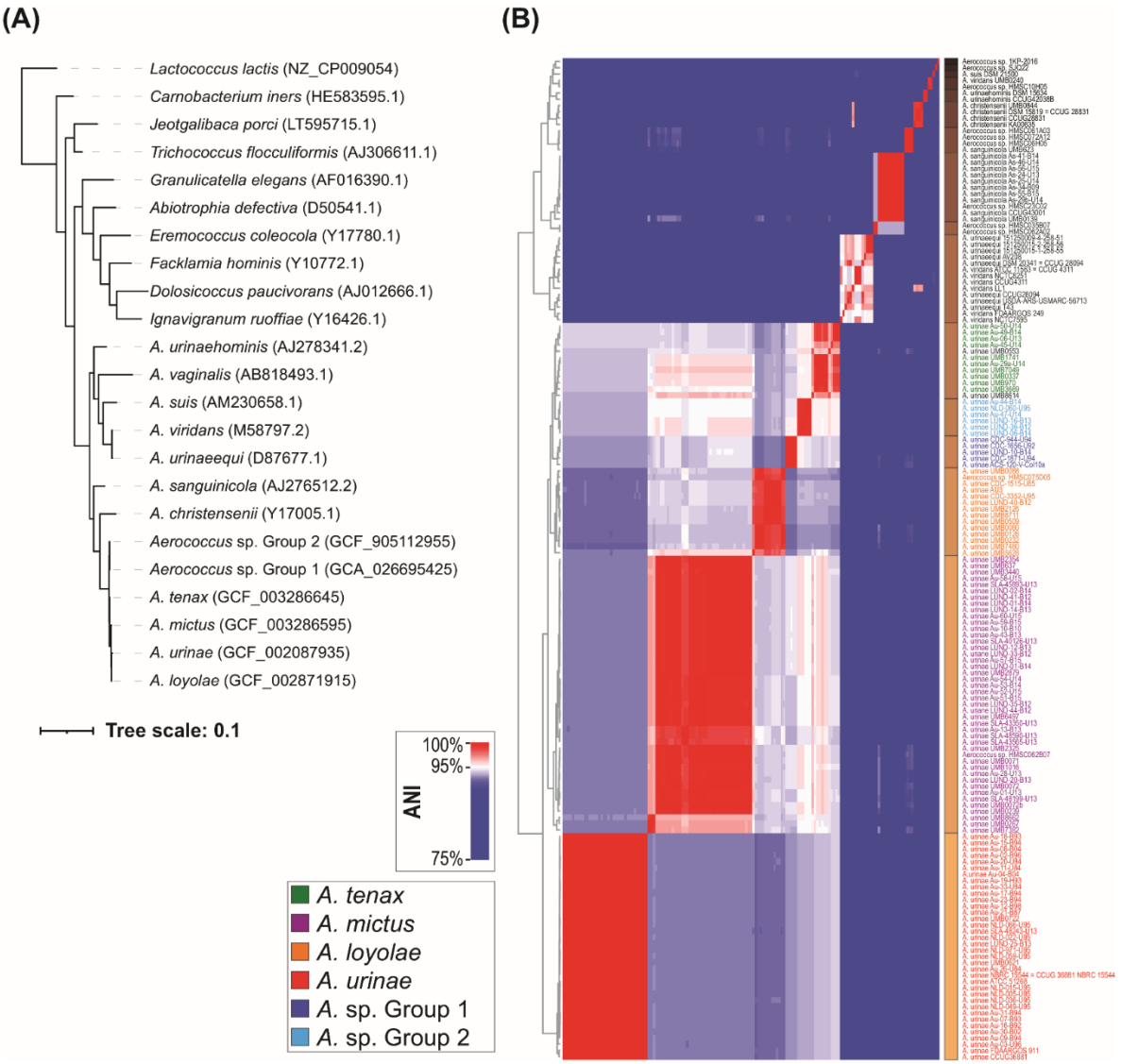


Figure 3. Comparison of *Aerococcus* Genomes. A) Phylogenetic tree based on 16s rRNA gene sequence comparison. B) ANI analysis of core genomes. X and Y axis are pairwise comparisons between *Aerococcus* genomes.

Biochemical Phenotype Investigation

Although the new AUC groups of *A. urinae, A. tenax, A. mictus,* and *A. loyolae* could be supported and identified by ANI analysis, it has yet to be determined if there are distinct behavioral differences of each group that can be used to distinguish between them and/or link any particular group to pathogenesis. The original characteristics of *A. urinae* defined by Aguirre

and Collins in 1992 was described as β-galactosidase negative, β-glucuronidase positive, pyrrolidonylarylamidase negative, leucine aminopeptidase positive, lactose fermentation negative, glycerol fermentation negative, and sorbitol fermentation positive with *A. viridans* possessing the opposite profile (73). With each successive discovery of a novel *Aerococcus* species, the respective authors conducted similar biochemical tests, and the tests required to differentiate each novel species grew longer each time. This culminated in the last emendation of *Aerococcus* biochemical descriptions in 2014 by Tohno et al. with the outlining of 46 biochemical tests to differentiate eight *Aerococcus* species (124). These descriptions, however, were based only on the biochemical phenotyping of the type strains of each respective species and did not account for robustness or variants. Since the new AUC groups represent undescribed variants, I hypothesized that the AUC groups would not fit in any of these species descriptions.

As such, I conducted 69 analytical profile index (API) biochemical tests on multiple representatives of each AUC group to characterize their ability to ferment select sugars and demonstrate select enzymatic activities. Summarized results of tests that showed discrepancies between AUC groups is shown in **Table 4**. From these tests, I confirmed that the original biochemical profile description of *A. urinae* is still true. Additionally, I was able to match the description of *A. tenax*'s biochemical profile with that of the *Aerococcus* biotype II as first described by Christensen et al. in 1992 (96). Indeed, the ability to ferment esculin and other aryl β-glucoside compounds appears to be a uniquely *A. tenax* ability amongst the AUC. As for other discerning traits, *A. mictus* appears uniquely able to ferment D-mannose, although this ability may not be found in all isolates (86%).

Table 4. Biochemical Test Results Determined by API 50 CH and API ZYM Systems
Parenthesized number after each species name indicates number of strains tested. Parenthesized percentages within the table indicate percentage of isolates tested that indicated a positive result.

Test	*A. urinae* (3)	*A. tenax* (5)	*A. loyolae* (7)	*A. mictus* (7)
Production of acid from:				
Esculin	– (0%)	+ (100%)	– (0%)	– (0%)
Salicin	– (0%)	+ (100%)	– (0%)	– (0%)
Amygdalin	– (0%)	+ (60%)	– (0%)	– (0%)
Arbutin	– (0%)	+ (60%)	– (0%)	– (0%)
D-cellobiose	– (0%)	+ (100%)	– (0%)	+/- (57%)
D-melezitose	– (0%)	+ (100%)	– (0%)	– (0%)
Gentiobiose	– (0%)	+ (100%)	– (0%)	– (0%)
D-mannose	– (0%)	– (20%)	– (0%)	+ (86%)
Enzymatic Activity:				
beta-glucuronidase	+ (100%)	+ (80%)	+/- (42%)	+ (100%)
Naphthol-AS-BI-phosphohydrolase	+ (100%)	– (20%)	+ (100%)	+ (100%)

With evidence of both genotype and phenotype differing between the AUC species, it was investigated if they could be discriminated by mass spectrometry. In doing so, a mass spectrometry library on AUC species could be created to be used in routine clinical microbe identification via MALDI-TOF MS. In collaboration with Dr. Jens J. Christensen and Rimtas Dargis, 29 representative strains of AUC and the unnamed species groups were analyzed via mass spectrometry for distinct spectra (123). When using traditional formic acid extraction methods, distinct mass spectra could be obtained for each of the AUC species but not the unnamed species groups. Only when the extraction method utilized an ethanol/formic acid technique was it then able to retrieve distinct mass spectra for all groups. Thus, it was demonstrated that, in addition to biochemical and ANI analysis, MALDI-TOF MS could be utilized to distinguish between AUC species.

Comparative Genomics of the *Aerococcus urinae* Complex

Rather than only characterizing differences based on phenotype, I also conducted analysis of genomic differences between AUC species via comparative genomics. To compare genomes directly, I utilized representative complete genome sequences from each of the AUC species as outlined in **Figure 4**, which displays a pangenome analysis totaling twelve representatives. Over half of the genomes appear solid in color, indicating a shared core genome. As seen with the ANI analysis, the *A. urinae* genomes are highly homogeneous and demonstrate little intra-species variation. However, the other three AUC species demonstrate variability within strains with the *A. mictus* strains demonstrating the highest heterogeneity.

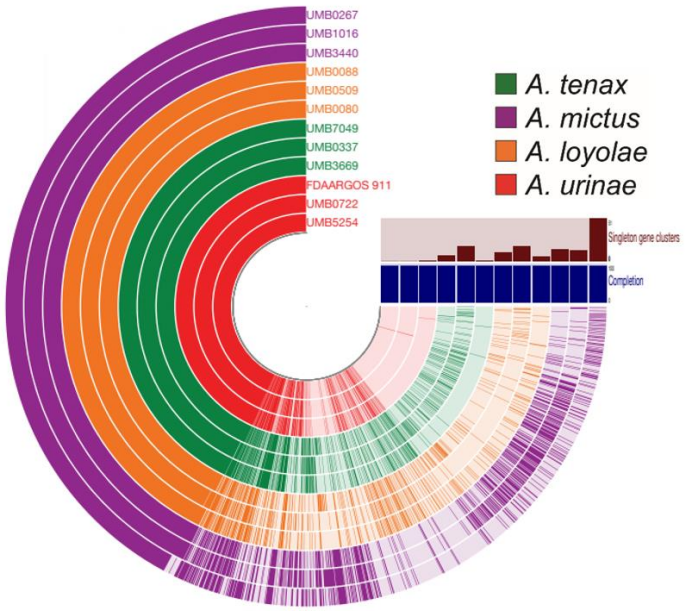


Figure 4. Pangenome Analysis of AUC Species Representatives. Comparison of shared (solid color) and unique (transparent color) genomic content of AUC species representatives.

The pangenome analysis demonstrates that there is a large portion of DNA that appears unique to each species. To investigate this DNA content, I conducted pairwise genome comparisons analyzing syntenic loci for differences. A combined visualization of genome

homology of type strains is shown in **Figure 5**. Here, it appears that high amounts of variation between genomes occurs at concentrated loci, or hypervariable hotspots, rather than being diffusely distributed. I hypothesize that these hotspots can explain the majority of speciation between AUC species.

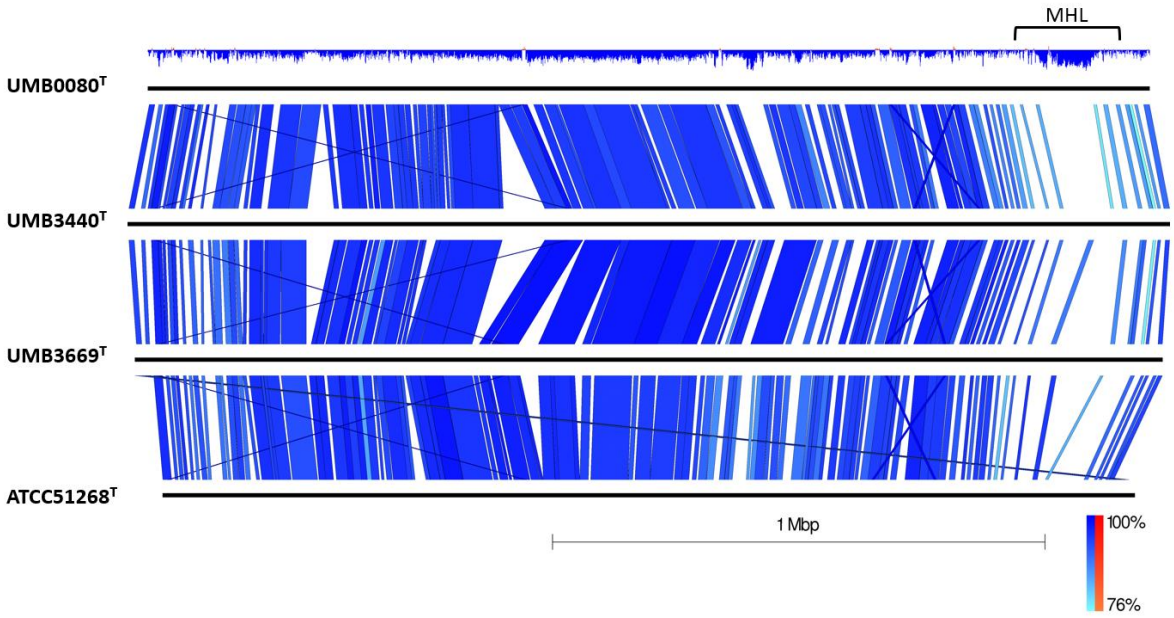


Figure 5. Genetic Homology between Aligned AUC Type Strains. Blue regions between aligned genomes represent BLAST homology (darker blue is closer to 100% homology). The top graph indicates GC percentage with blue indicating below 50% and red indicating above 50%. MHL = Main Hypervariable Locus

To define the genetic content that is unique to each species, I first annotated the genetic content that is variable within members of the same species, or intra-species variation. Unsurprisingly, the majority of this content can be attributed to mobile genetic elements, such as prophages, Integrative and Conjugative Elements (ICEs, also known as conjugative transposons), and transposable elements. These elements also occur primarily at defined hotspots. All twelve genomes analyzed contained at least one ICE.

As a case example, UMB3440 and UMB1016, both *A. mictus* strains have a difference of almost 100kb in DNA content amounting to about 120 extra open reading frames (ORFs) in UMB1016. This large DNA content amounts to about 5% of discrepant DNA, which would have put these two strains as different species were this content part of their core genome. Instead, this content can be almost entirely explained by differences attributed to the accessory genome consisting of three mobile genetic elements: one compound transposon, one prophage, and one ICE.

The two largest hypervariable regions between species are ones that contain a surface capsule biosynthesis operon (CPS) and one that contains a large collection of metabolic operons. Comparison of this first hotspot region is shown in **Figure 6**. Between the flanking homologous regions, a 30kb stretch of DNA is found in each type strain that is representative of their respective species. Although the first four ORFs of the CPS operon are conserved in all species, the latter portions are entirely species-specific, primarily consisting of predicted sugar transferase enzymes.

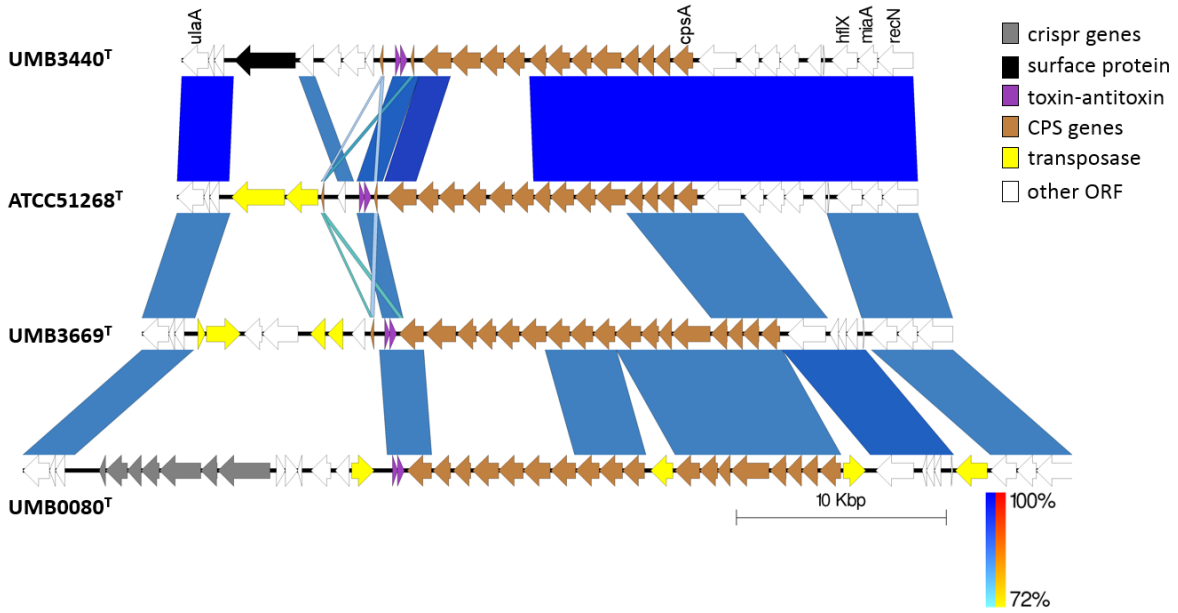


Figure 6. Comparison of CPS Operon Hotspot between Type Strains. Alignment of syntenic CPS loci with comparison of BLAST homology. ORFs color coded by predicted function.

I conducted an analysis of the same 115 AUC genomes in **Table 3** for the consistency of these species-specific CPS operons and found almost perfect agreement between CPS operon pattern and species designation. Interestingly, the *A. mictus* pattern demonstrated several exceptions as shown in **Figure 7**, where the *A. mictus* pattern appears in certain *A. tenax* and *A. loyolae* strains. Additionally, *A. loyolae* strains possess the greatest variability in this hotspot attributed to transposable elements (**Figure 8**). In this case example, UMB0088 appears to possess eight extra CPS ORFs compared to the type strain UMB0080. However, these extra genes are actually just the first eight CPS ORFs of the operon that have been duplicated and rotated. Furthermore, two rRNA genes have been inserted into this region, also presumably due to transposable element activity.

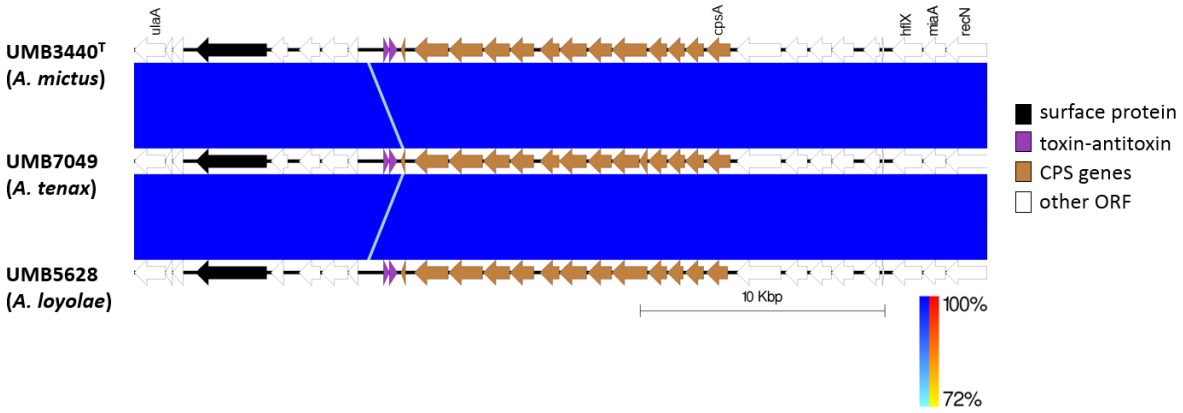


Figure 7. *A. mictus* CPS Operon Pattern Exceptions. Demonstration of similarity in CPS locus arrangement in other species.

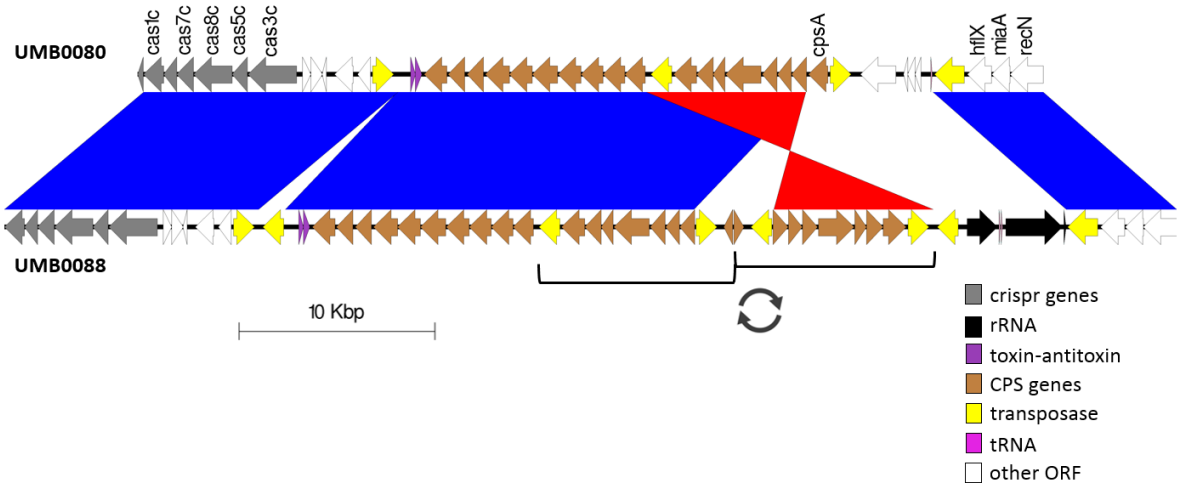


Figure 8. Variability within *A. loyolae* CPS Operon Hotspot. Demonstration of dissimilarity in CPS locus arrangement within the same species. Red homology coloring indicates genomic segment inversion.

Because such variability in this CPS operon occurs between species, I investigated whether polysaccharide biosynthesis pattern could be used to discriminate between isolates similar to how serotyping is conducted in *Streptococcus* and *Enterococcus* species (**Figure 9A**). Following the protocol to isolate surface polysaccharides from Gram positive bacteria, I was, however, unable to visualize any bands (**Figure 9B**). Although capsule polysaccharides and

teichoic acids could be visualized from the *Enterococcus* controls, only the teichoic acid bands could be visualized from *Aerococcus* strains. It is possible that AUC only produces surface polysaccharides under certain conditions not reproduced here or that it does not produce them at all. For this experiment, I was only able to conduct this assay twice, each time producing negative capsule polysaccharide bands. However, because the protocol was designed for *E. faecalis* strains, it is possible the methodology for investigating AUC strains may need to be further adapted.

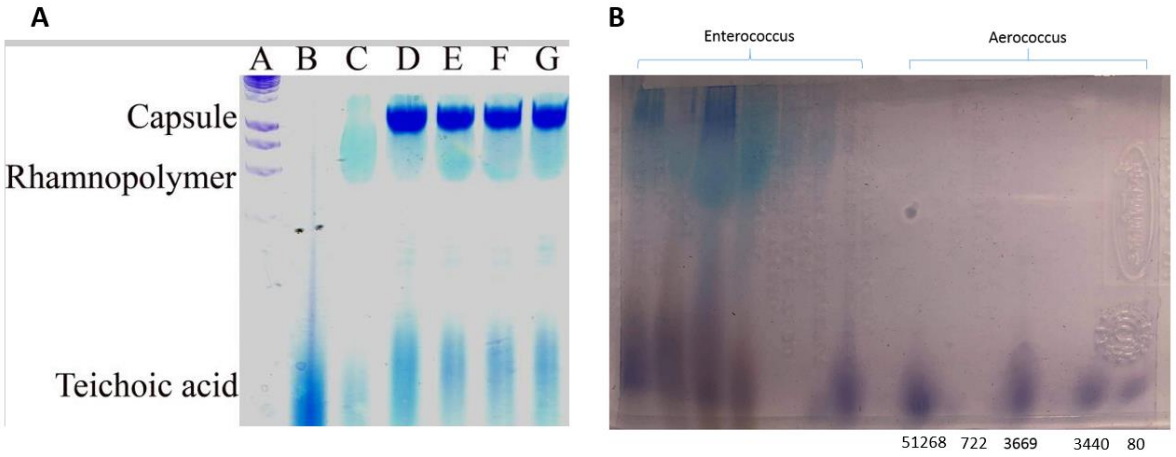


Figure 9. Isolation of Capsular Polysaccharides of *E. faecalis* and AUC Isolates. A) Capsular polysaccharide isolation gel key adapted from Thurlow et al. (125). B) Isolation of capsular polysaccharides of *E. faecalis* vs AUC strains.

The second hypervariable region is much larger than the first, ranging between 50kb to 100kb in size and has been labelled as the Main Hypervariable Locus (MHL) (**Figure 5**). This region varies widely between species with the *A. loyolae* pattern at 48 ORFs, *A. urinae* pattern at 66 ORFs, *A. tenax* at 79 ORFs, *A. mictus* secondary pattern at 71 ORFs, and the *A. mictus* primary pattern at 106 ORFs (**Figure 10**). The observation of *A. mictus* possessing two distinct patterns matches the phylogenetic analysis of the species with the secondary pattern describing the *A. mictus* minor branch. The MHL region heavily consists of predicted metabolic

ORFs, with the most common differences between species attributed to sugar phosphotransferase (PTS) transporter ORFs. These genes are involved with the importation and metabolism of specific sugars. I hypothesize that this region largely explains species-specific metabolic behaviors. For example, I analyzed all genomes for β-glucosidase genes, finding them only in *A. tenax* genomes within this MHL region. In fact, *A. tenax* strains possess up to four copies of β-glucosidase genes, which likely explains the *A. tenax*-unique β-glucoside metabolic activity (**Table 6**).

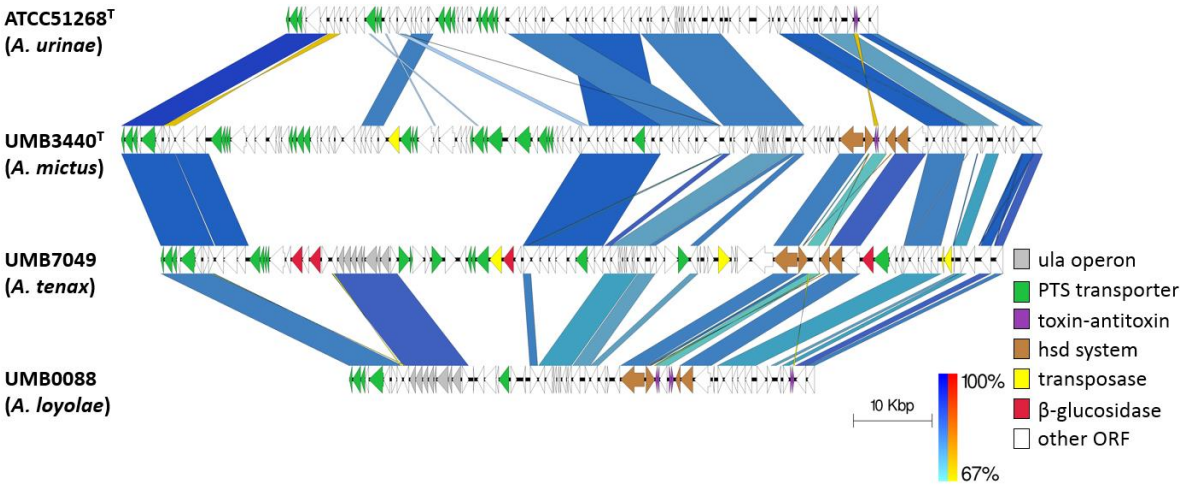


Figure 10. Comparison of Main Hypervariable Locus. BLAST homology comparison of syntenic MHL regions between selected strains of AUC species. ORFs are color coded by predicted function.

Because I predict that many of these metabolic genes within the MHL can explain species-specific behavior, the MHL comparison could be used to predict these behaviors. As seen in **Figure 10**, *A. tenax* and *A. loyolae* possesses 8 extra ORFs, which have been predicted to encode the full L-ascorbate degradation pathway (ula). As such, I tested three representatives from each AUC species for ability to acidify L-ascorbate and, indeed, only the *A. tenax* and *A. loyolae* strains were able to do so. Therefore, it is likely that many of the other PTS systems

found within the MHL could be exploited to further characterize metabolic differences between species.

However, not all species-specific metabolic behaviors can be traced to the MHL. As seen in **Table 4**, almost half of tested isolates were positive for β-glucuronidase behavior. I surveyed eight *A. loyolae* genomes for β-glucuronidase genes in their genomes and found hits in all eight. However, it was discovered that four genomes, UMB0126, UMB0232, UMB0080, and UMB0088, possessed intact copies of the gene while the other four genomes, UMB2126, UMB8711, UMB5628, and UMB0509, all possessed a copy of the same gene with a single C→A polymorphism introducing an early stop codon that is predicted to produce a truncated form of the protein within the first eight amino acids. As such, while the MHL and CPS regions may explain a large part of species differences, there is DNA content outside these loci that contributes as well.

Multiplex PCR for *Aerococcus urinae* Complex Species Discrimination

After the ANI analysis of the 115 strains that went into the original investigation to organize the AUC species was completed, 36 genomes were revealed to belong to the newly more narrowly defined *A. urinae* species designation (**Table 3**). However, of these 36 *A. urinae* strains, only two came from Loyola University Health Center, representing just 6% of the total 33 Loyola strains that went into the investigation. As such, I hypothesized that *A. urinae* strains are rare within the local isolate distribution. To test this hypothesis, I utilized the output from the comparative genomics analysis to create a multiplex PCR test that could rapidly discriminate between the AUC species.

Primer set 1 identifies the *A. tenax*–specific aryl β-glucosidase gene(s). *A. tenax* strains encode up to four polymorphic copies of β-glucosidase genes within the MHL, while all non-*A. tenax* species encode none. These genes are hypothesized to endow *A. tenax* strains with the ability to ferment aryl β-glucoside molecules, such as esculin and amygdalin, a distinctive phenotypic characteristic first described by Dr. Jens Christensen and colleagues as "biotype II" (96). The identifying PCR product size of this primer set can produce multiple bands within the 740 – 1000nt range most often appearing as a single band at either 740 or 1000 length size (**Figure 11**).

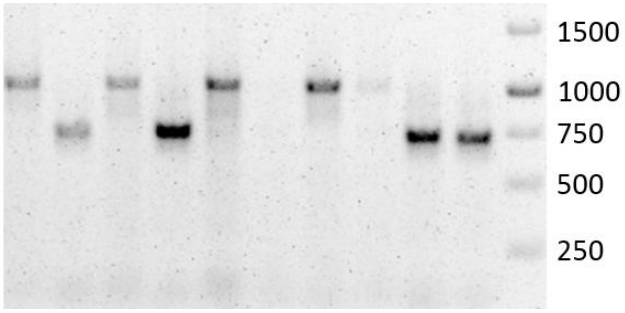


Figure 11. Band Sizes from Primer Set 1 Identifying *A. tenax* Strains

Primer set 2 identifies *A. urinae* and most *A. mictus* strains that encode the Type 1-E CRISPR system. *A. loyolae* and *A. tenax* strains either encode a different system (see primer set IV) or do not encode a CRISPR system at all. This primer set specifically identifies the helicase/endonuclease gene, Cas3, and positively identifies all *A. urinae* and most *A. mictus* strains with a single band size of 207. There is a small genetic outgroup of *A. mictus* strains that do not encode any CRISPR system and can be positively identified with the optional primer set (see primer set V). To distinguish *A. urinae* from *A. mictus* strains, primer set III is utilized with a product size of 359, producing a double band for all *A. urinae* strains (see primer set III) and only a single band for most *A. mictus* strains.

Primer set 3 identifies the *A. urinae*–specific *Aerococcus* surface protein gene, Asp1, as first described by Dr. Erik Senneby and co-authors (102). Asp1 is an LPxTG surface protein that ranges from 855 to 1449 nt in all *A. urinae* complex members. All forms of the protein contain a similar N-terminal secretion signal sequence and C-terminal sortase recognition motif (LPxTG) with a highly variable middle region of unknown function(s). This variable region of the protein is species-specific and can be used to positively identify strains of *A. urinae* identity. The identifying PCR product size of this primer set is always a single band at 359. Combined with primer set II, a positive identify of an *A. urinae* strain will produce a double band at sizes 207 and 359 (**Figure 12**).

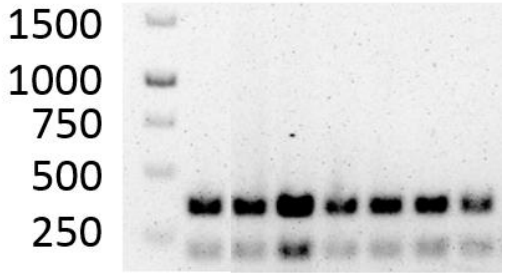


Figure 12. Band Sizes from Primer Set 2 and 3 Identifying *A. urinae* Strains

Primer set 4 identifies the *A. loyolae*-specific Type 1-C CRISPR system. This CRISPR system is directly adjacent to the CPS operon within *A. loyolae* strains (**Figure X**). This primer set specifically identifies the endonuclease gene, Cas1c, producing a single identifying band at size 511 (**Figure 13**).

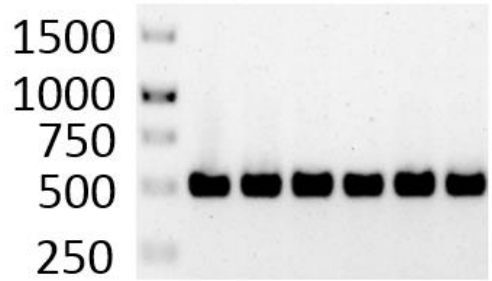


Figure 13. Band Size from Primer Set 4 Identifying *A. loyolae* Strains

The four core primer sets can be combined into a multiplex PCR experiment to then identify AUC isolates. Testing these primers on the type strains of each of the AUC species produces the gel output as seen in **Figure 14A**. The algorithm summarizing the features that the multiplex PCR tests for is outlined in **Figure 14B**.

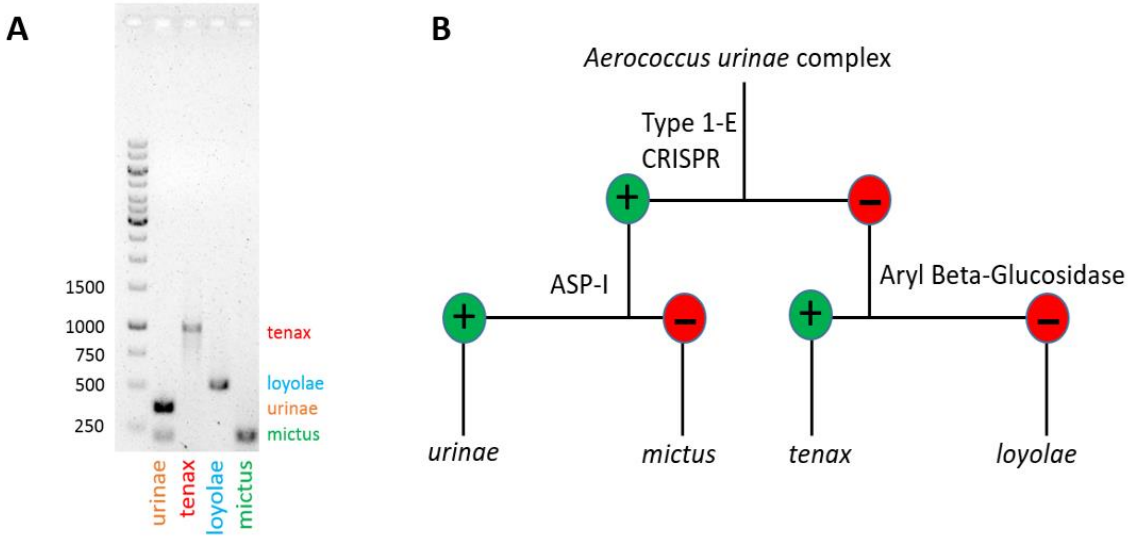


Figure 14. Multiplex PCR Gel Pattern and Identification Algorithm. A) Band key for identification of AUC species identity. B) Identification algorithm based on selected unique genomic features

An optional fifth primer set can be utilized to identify the secondary branch of *A. mictus* strains, which cannot be positively identified by primer set II, as they do not encode any CRISPR system. An alternative method for identifying *A. mictus* strains is via identifying the species-specific prophage. The prophage belongs to the Siphoviridae family and is 48,593 nt in length. This prophage is only found within strains of *A. mictus* and is also found in the genetic outgroup. It is not recommended to include this primer set with the other four primer sets in multiplex as this set produces variable and interfering results. When used in conjunction with primer set II, this primer set should ideally produce a positively identifying double band, one at 207 and one at 1632. However, for not fully understood reasons, only one of either band tends to dominate. Additionally, members belonging to the *A. mictus* outgroup sometimes produces a

secondary unknown band near 300 that can easily be confused with the *A. urinae* specific primer set III band at 359. As such, this PCR primer set is recommended to be utilized in a separate PCR when a signal is absent after using the four core multiplex primers.

To evaluate the efficacy of this multiplex PCR test, I conducted the test on 189 isolates identified as belonging to the AUC by MALDI-TOF MS (**Table 5**). These isolates were collected at Loyola University Health Center from urogynecology patients between 2017 and 2023. Of the 189 isolates, 166 could be positively identified yielding the following distribution: *A. loyolae* – 50 (26%), *A. mictus* – 95 (50%), *A. tenax* – 14 (7%), *A. urinae* – 7 (4%), and unidentifiable – 23 (12%) (**Figure 15**). The unidentifiable strains either produced unusual band sizes or no bands upon gel imaging. Ten of the 23 unidentifiable strains were analyzed via short-read sequencing to determine if these strains belonged to the unnamed species groups. After ANI analysis, three strains were identified as belonging to Species Group 1, two strains belonging to the *A. mictus* minor branch, three strains as unknown, one strain as *A. loyolae,* and one strain as *S. aureus*.

Table 5. *A. urinae* Complex Isolates Tested

pswab = perineal swab, vswab = vaginal swab, uswab = urethral swab, fswab = foreskin swab, CTIP = catheter tip, cath = transurethral catheter urine, void = voided urine, puswab = periurethral swab, stone = kidney stone. 1 = positive test value, 0 = negative test value

Isolate UMB	Sample collection date	AUC Species	Collection Method	Growth at 10µg/ml Azithromycin	Growth at 100µg/ml Azithromycin	*ermA* PCR
4427	2017	*loyolae*	void	1	0	0
4576	2017	*loyolae*	cath	1	0	0
4583	2017	*mictus*	cath	1	1	1
4604	2017	*mictus*	cath	1	0	0
4704	2017	*mictus*	void	1	1	1
4761	2017	*mictus*	void	1	1	1
4867	2017	*loyolae*	pswab	1	0	0
4870	2017	*loyolae*	vswab	1	0	0
4882	2017	*mictus*	cath	1	1	1
4893	2017	*mictus*	vswab	1	1	1
4904	2017	*mictus*	pswab	1	1	1

Isolate UMB	Sample collection date	AUC Species	Collection Method	Growth at 10µg/ml Azithromycin	Growth at 100µg/ml Azithromycin	*ermA* PCR
4909	2017	*mictus*	void	1	1	1
5025	2017	*loyolae*	vswab	1	0	0
5035	2017	*loyolae*	pswab	1	0	0
5042	2017	*mictus*	void	1	0	0
5090	2018	*mictus*	void	1	0	0
5098	2018	*mictus*	vswab	1	1	1
5254	2018	*urinae*	void	1	0	0
5258	2018	*loyolae*	cath	1	1	1
5266	2018	Unknown	cath	1	0	0
5293	2018	*mictus*	cath	1	1	1
5311	2018	*tenax*	void	1	0	0
5440	2018	*tenax*	cath	1	0	0
5445	2018	*tenax*	vswab	1	0	0
5455	2018	*mictus*	pswab	1	0	0
5467	2018	*mictus*	void	1	1	1
5548	2018	*mictus*	void	1	1	1
5567	2018	*loyolae*	void	1	0	1
5582	2018	*mictus*	cath	1	0	0
5628	2018	*loyolae*	cath	1	0	0
5682	2018	Unknown	vswab	0	0	0
5702	2018	*mictus*	cath	1	1	1
5710	2018	Unknown	cath	1	1	1
5792	2018	*tenax*	cath	1	0	0
5842	2018	*loyolae*	void	0	0	0
5863	2018	*urinae*	cath	1	0	0
5903	2018	*mictus*	pswab	1	0	0
5938	2018	*mictus*	void	0	0	0
5956	2018	*mictus*	vswab	1	0	0
5973	2018	*mictus*	cath	1	0	0
5975	2018	*urinae*	void	1	0	0
5980	2018	*mictus*	void	1	0	0
6039	2018	*loyolae*	vswab	0	0	0
6088	2018	*loyolae*	pswab	0	0	0
6151	2018	*tenax*	upper	1	0	0
6163	2018	*tenax*	vswab	1	0	0
6195	2018	*urinae*	cath	1	1	1
6232	2018	*mictus*	cath	0	0	0
6249	2018	*loyolae*	void	0	0	0
6289	2018	*mictus*	void	0	0	0

Isolate UMB	Sample collection date	AUC Species	Collection Method	Growth at 10µg/ml Azithromycin	Growth at 100µg/ml Azithromycin	*ermA* PCR
6296	2018	*mictus*	vswab	0	0	0
6444	2018	*urinae*	void	1	0	0
6497	2018	*mictus*	cath	0	0	0
6509	2018	*loyolae*	void	1	0	0
6511	2018	*loyolae*	pswab	0	0	0
6587	2018	*mictus*	void	1	1	1
6625	2018	*loyolae*	cath	1	1	1
6782	2018	*mictus*	cath	1	1	1
6852	2018	*mictus*	cath	1	1	1
6872	2018	*mictus*	pswab	0	0	0
6965	2018	Unknown	pswab	1	0	0
7049	2018	*tenax*	cath	0	0	0
7137	2018	*mictus*	vswab	1	1	1
7144	2018	*mictus*	void	1	1	1
7153	2018	*mictus*	pswab	1	1	1
7228	2018	*tenax*	cath	1	1	1
7235	2018	*mictus*	vswab	0	0	0
7245	2018	*mictus*	pswab	0	0	0
7248	2018	*mictus*	void	0	0	0
7288	2018	*tenax*	vswab	1	1	1
7377	2019	Unknown	void	1	0	0
7382	2019	*mictus*	cath	1	1	0
7395	2019	*mictus*	vswab	1	1	0
7412	2019	Unknown	cath	0	0	0
7480	2019	*loyolae*	cath	0	0	0
7535	2019	*mictus*	cath	0	0	0
7574	2019	*loyolae*	cath	0	0	0
7783	2020	*mictus*	void	0	0	0
7835	2020	*loyolae*	cath	1	1	1
7838	2020	*mictus*	cath	1	1	1
7968	2020	*mictus*	void	1	1	1
8048	2020	*urinae*	cath	1	1	1
8052	2020	Unknown	void	0	0	0
8065	2020	*loyolae*	cath	0	0	0
8115	2021	*loyolae*	cath	1	1	1
8147	2021	*mictus*	void	1	1	1
8196	2021	*loyolae*	cath	1	1	1
8255	2021	Unknown	void	0	0	0
8266	2021	*urinae*	cath	1	1	1

Isolate UMB	Sample collection date	AUC Species	Collection Method	Growth at 10µg/ml Azithromycin	Growth at 100µg/ml Azithromycin	*ermA* PCR
8287	2021	*mictus*	cath	1	1	1
8332	2021	*mictus*	cath	1	0	0
8345	2021	*mictus*	void	1	1	1
8377	2021	*loyolae*	cath	0	0	1
8400	2021	*mictus*	void	1	1	1
8442	2021	*mictus*	cath	1	0	0
8445	2021	*loyolae*	cath	0	0	0
8463	2021	*loyolae*	cath	0	0	1
8534	2021	*loyolae*	cath	0	0	0
8609	2021	*loyolae*	cath	1	1	1
8662	2021	*mictus*	cath	1	0	0
8711	2021	*loyolae*	cath	1	0	0
8790	2021	*mictus*	void	1	1	1
8834	2021	*mictus*	cath	1	1	1
8876	2021	*mictus*	cath	1	0	0
8877	2021	*mictus*	cath	1	0	0
8888	2021	*loyolae*	cath	0	0	0
8891	2021	*mictus*	cath	1	1	1
8922	2021	Unknown	cath	1	1	1
8965	2021	*mictus*	cath	1	0	0
8978	2021	*loyolae*	void	0	0	1
8997	2021	*mictus*	stone	1	1	1
9003	2021	*mictus*	cath	1	0	0
9047	2021	*mictus*	cath	1	1	1
9071	2021	*mictus*	cath	1	0	0
9075	2021	Unknown	cath	0	0	0
9092	2021	*loyolae*	void	0	0	0
9109	2021	Unknown	cath	1	0	0
9113	2021	Unknown	cath	1	0	0
9125	2021	*loyolae*	cath	1	0	0
9223	2021	*mictus*	void	1	1	1
9241	2021	*loyolae*	cath	0	0	0
9252	2021	*mictus*	cath	1	0	0
9291	2021	*tenax*	cath	1	0	0
9335	2021	*mictus*	cath	1	0	0
9597	2022	*loyolae*	cath	1	0	0
9609	2022	*mictus*	cath	1	0	0
9750	2022	*loyolae*	cath	1	1	1
9774	2022	*mictus*	cath	1	0	0

Isolate UMB	Sample collection date	AUC Species	Collection Method	Growth at 10µg/ml Azithromycin	Growth at 100µg/ml Azithromycin	*ermA* PCR
9866	2022	*mictus*	void	1	1	1
9904	2022	*loyolae*	cath	1	1	1
9966	2022	*loyolae*	cath	1	1	1
9988	2022	*mictus*	cath	1	1	1
9998	2022	*loyolae*	cath	0	0	0
10115	2022	*mictus*	void	1	1	1
10122	2022	Unknown	cath	1	0	1
10164	2022	*mictus*	cath	1	1	1
10229	2022	*mictus*	cath	1	1	1
10306	2022	Unknown	cath	1	0	0
10435	2022	*loyolae*	cath	1	1	1
10473	2022	*mictus*	void	1	0	0
10479	2022	*mictus*	vswab	1	0	0
10504	2022	Unknown	cath	1	1	1
10676	2023	Unknown	cath	1	0	0
10699	2023	*tenax*	cath	1	0	0
10710	2023	*tenax*	uswab	1	1	1
10718	2023	*loyolae*	puswab	1	0	0
10724	2023	*tenax*	CTIP	1	1	1
10849	2023	*mictus*	cath	1	1	1
10889	2023	*loyolae*	cath	1	1	1
10891	2023	*loyolae*	vswab	1	1	1
10991	2023	Unknown	puswab	1	0	0
11311	2023	*mictus*	uswab	1	1	1
11484	2023	*loyolae*	cath	1	1	1
11494	2023	*loyolae*	vswab	1	1	1
11740	2023	Unknown	void	0	0	0
11781	2023	*mictus*	CTIP	1	0	0
11920	2023	*loyolae*	void	0	0	0
12094	2023	Unknown	void	0	0	0
12271	2023	Unknown	void	0	0	0
12459	2023	*mictus*	cath	1	1	1
12466	2023	*mictus*	void	1	1	1
12475	2023	*mictus*	puswab	1	1	1
12569	2023	*loyolae*	cath	1	1	1
12582	2023	*loyolae*	void	1	1	1
12589	2023	*mictus*	puswab	0	0	0
12697	2023	*mictus*	vswab	1	1	1
12698	2023	*mictus*	void	1	1	1

Isolate UMB	Sample collection date	AUC Species	Collection Method	Growth at 10µg/ml Azithromycin	Growth at 100µg/ml Azithromycin	*ermA* PCR
12699	2023	Unknown	void	1	1	1
12804	2023	*mictus*	vswab	1	1	1
12807	2023	*mictus*	cath	1	1	1
12811	2023	*mictus*	void	1	1	1
12828	2023	*mictus*	vswab	1	0	0
12866	2023	*mictus*	vswab	1	1	1
12903	2023	*mictus*	void	1	1	1
12930	2023	*mictus*	cath	1	1	1
12935	2023	*mictus*	vswab	1	1	1
12943	2023	*loyolae*	puswab	1	1	1
12944	2023	*tenax*	cath	1	1	1
12945	2023	Unknown	vswab	1	1	1
12976	2023	*mictus*	cath	1	0	0
12980	2023	*mictus*	puswab	1	0	0
12993	2023	*mictus*	vswab	1	1	1
13000	2023	Unknown	vswab	1	0	0
13003	2023	*loyolae*	cath	1	0	0
13074	2023	*mictus*	cath	1	1	1
13120	2023	*mictus*	cath	1	0	0
13136	2023	*loyolae*	void	1	1	1
13145	2023	*mictus*	void	1	1	1
13182	2023	*mictus*	void	1	0	1

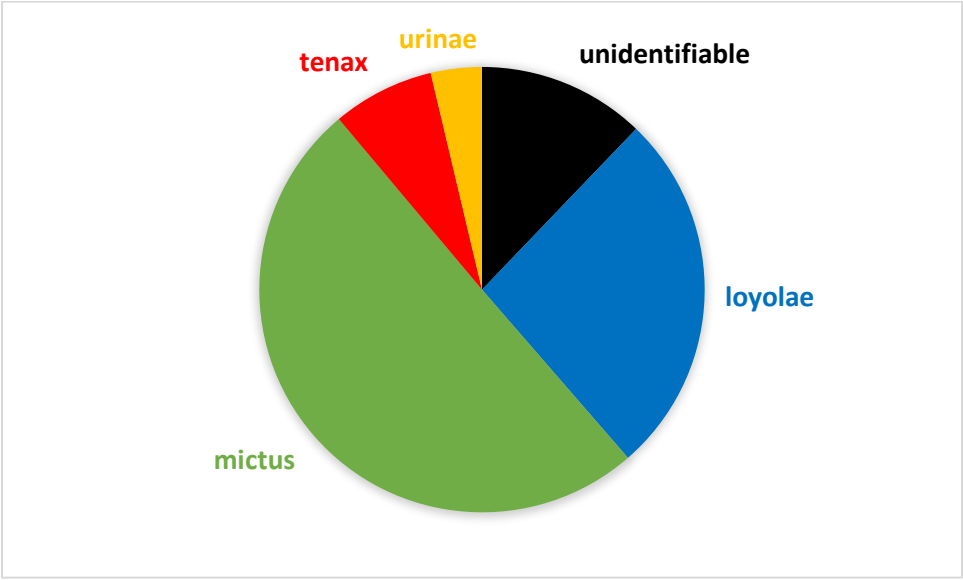


Figure 15. AUC Species Distribution Identified via Multiplex PCR. Pie graph of species distribution of 189 AUC isolates

Within this analysis, it was observed that the same human individual can be colonized by multiple AUC species within and between different body sites, including bladder, urethra, perineum, and vagina. No patterns could be determined between species for body site, lower urinary tract disease, or year of isolation primarily due to the low representation of *A. tenax* and *A. urinae* isolates. As such, this supported the original hypothesis that *A. urinae* exists as a minority within the local Loyola University Health Center distribution.

Genetic Features of *Aerococcus urinae* Complex

While sequencing, assembling, and analyzing the complete genomes of the AUC, a couple novel features were discovered including an *Aerococcus* plasmid and a predicted quorum sensing system. Only one *Aerococcus* plasmid has been annotated so far, found in *Aerococcus urinaeequi* that confers resistance to tetracycline (109). As such, the discovery of a plasmid found in *A. tenax* and *A. mictus* strains represents the first plasmid annotated in the AUC.

The plasmid size ranges from 7642bp to 9244bp with variation due to a single insertion sequence element (**Figure 16**). The transposase found within this insertion element typically reflects the dominant transposase that is found on the strain's chromosome. Within UMB3669, this is the IS1182 family transposase IS663, and within UMB3440, it is the IS3 family transposase ISEfa10. The rest of the plasmid outside of the insertion element is conserved between strains and contains two predicted ORFs, the larger of which may encode a putative replication protein possessing a helix-turn-helix motif at its C-terminal end. However, no matches to any known plasmid replication proteins were found in any plasmid databases, potentially representing a novel plasmid family. The only other DNA match that could found on

the plasmid was for a 214bp sequence identified as a Streptococcal small RNA (sRNA) with unknown function. So far, within the AUC, plasmids have only been found in *A. tenax* and *A. mictus* strains.

Figure 16. *A. tenax* and *A. mictus* Plasmid. Predicted ORFs within identified plasmids visualized via SnapGene

Dr. Evann Hilt previously hypothesized that a quorum sensing system may exist within *Aerococcus sp.* when she described a density-dependent pigmentation phenotype that also depended on sugar availability (121). As such, I investigated if any genetic features within the AUC genomes could be predicted components of a quorum sensing system. The first evidence I found was a lipoprotein sex pheromone precursor, predicted by BLAST homology. This ORF is predicted to encode a lipoprotein about 400 amino acids long with a leader sequence, the last seven residues of which encode a linear heptapeptide (**Figure 17**). The translated precursor protein is hypothesized to be cleaved by a lipoprotein signal peptidase to release the heptapeptide from its signal sequence and out into the extracellular space. The predicted AUC sex pheromone shares many similarities with other Firmicute/Bacillota pheromones, including those from *Streptococcus gordonii, Enterococcus faecalis, and Staphylococcus aureus* (bolded in

Figure 17). The pheromone sequence itself is extremely hydrophobic, contains a conserved antepenultimate leucine residue, and is predicted to be cleaved before a conserved cysteine residue. All species within the AUC encode the same predicted pheromone heptapeptide; however, other *Aerococcus* species encode different predicted pheromones.

```
A. urinae (ALL):  MKKRIQHMIVIGLA SLSLLVA CGRDAQPKSTAGQTQ
A. christensenii: MKKRKQVITGLALS MLCLLSA CQSASLSKSVAGQSQTTNE
A. sanguinicola:   MKRLHHKLGLLVC LALALAG CQGPDQAEEGQAQESSGQASQAE


S. gordonii:         MKKIYTLALLVF SVFILAA CSSQEAWLNGTWKGEKNK

E. faecalis:       MLKKPFLLFFSLLG AIFILAS CGIGKDAVTDTKYKVSLQQAAE   (cAM373)
E. faecalis:     MKVNKFVKGFAAIA LFSLVLAG CGADKKDNTTNSS            (cAD1)
                               FLVMFLSG                           (cPD1)
                               VAVLVLGA                           (cOB1)

S. aureus:             MKRTLVLLIT AIFILAA CGNHKDDQAGKDN
S. epidermidis:        MKRTIFLLMS ILLLLTA CGDGHKQTSSDKEQSEHKDNHNKNQVKQ
S. haemolyticus:       MKRTLFLIIT SIVLLTA CGNNNDNKNNQSVNKQETKTDS

L. crispatus:        MKKYLQVMALAG IALTLSG CGKLKDSSLANNA
L. acidophilus:      MKKYLQIMALAG IAITLTG CGRLKDSSLANNATTTSTTKKKSYQTTNT
L. johnsonii:        MKRFLQIALLLA TGLSLSA CGNLKNSDLANNPTTSTTKKKSY
L. delbrueckii B:              MAS CSLLLAA CGNLKNSDLASNSTTTTSEAKKYETTSSTDG
L. reuteri:      MKRKAKKIAVSAAVLM CTVLLAS CGFGEKSSSKN

B. cereus:             MKKMALSFAV VSLLLGA CSNNSNTISKKDEVIQKDTKEKS
```

Figure 17. Predicted and Previously Annotated Sex Pheromone Peptides. Peptides in bold have been annotated previously by other studies. Peptides underlined have been confirmed experimentally to be involved in conjugative mating.

In *Enterococcus* species, the pheromone is part of a pheromone-responsive plasmid system, facilitating horizontal gene transfer between and across species (126). In *S. gordonii*, this pheromone can induce the conjugative mating response in *E. faecalis* to donate its plasmid not only to *S. gordonii* recipients, but also to non-pheromone producing *Streptococcus* species in the vicinity (127). During this mating response, *E. faecalis* expresses clumping factor to tightly bind with recipient cells, leading to dense biofilm aggregates (128). The host specificity of these

pheromones is poorly understood, and it is unknown whether any bacterial pheromones have any physiological significance within the bladder microenvironment.

To investigate any potential impact this pheromone may have on AUC behavior, I incubated the type strains of AUC with synthetic predicted heptapeptide in NYCIII media. However, over the course of 48 hours, no changes were observed in growth patterns, biofilm behavior, or bacterial survival. Even with cell-free supernatant, no discernable changes could be observed, which had been echoed by Dr. Evann Hilt in her supernatant experiments characterizing the pigmentation phenomenon. In *Streptococcus* species, the short hydrophobic peptide (SHP) is unable to induce responses when grown in peptide-rich media due to competition with non-specific peptides present in high concentrations (129, 130). As such, I incubated the strains again with synthetic heptapeptide in CDM that is absent of free peptides. But once again, no discernable impact could be observed on growth patterns, biofilm behavior, or bacterial survival.

I also investigated whether this pheromone had any impact on inducing competence. As is the case with *Streptococcus* species, addition of synthetic pheromone can promote the uptake of exogenous DNA to allow for genetic manipulation of strains. I first investigated whether AUC strains possess the necessary machinery for competence. Indeed, I was able to find homologous ORFs predicted to encode for all the necessary operons for late competence genes including the *comG*, *comE,* and *comF* operons as well as the genes *recA, ssbB,* and *dprA* as found in *Streptococcus* species (131). Furthermore, I analyzed the untranslated regions upstream of these operons and genes and was able to detect a homologous combox, the promoter sequence that is recognized by the sigma factor ComX to transcribe the late

competence genes (**Figure 18A**)(132). Besides the late competence genes, a reverse BLAST search of this promoter sequence found hits upstream of a large predicted LPxTG surface protein in *A. loyolae* and *A. mictus* strains and before a mobile genetic element in *A. tenax* and *A. urinae* strains. The predicted AUC combox is conserved in all AUC members and is highly similar to combox sequences of related Lactobaciliaceae (**Figure 18B**). With the potential for competence induction established, I once again incubated the type strains with synthetic pheromone in CDM with both linear and plasmid DNA to evaluate potential bacterial transformation. However, when following the transformation protocol for *Streptococcus pneumoniae*, no DNA uptake was observed for any strains with any of the tested DNA.

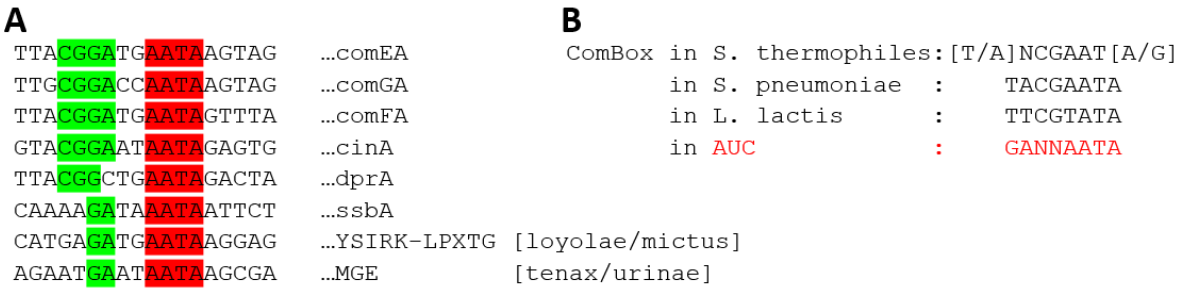


Figure 18. Predicted Combox Sequence in AUC and Related Bacteria. A) Upstream untranslated regions of late competence genes with conserved sequences. B) Comparison of selected species Combox sequences

Besides transformation, other efforts to introduce foreign DNA into AUC were met with similar rounds of failure including attempts at an ICE conjugation system (133) and electroporation of streptococcal plasmids. This resistance to genetic manipulation is a common phenomenon of non-laboratory bacterial strains and was especially true for *S. aureus* two decades ago. However, one strain RN4220 was found to accept plasmids by electroporation because of a mutation in its restriction-modification system, which has since led to the creation of many more genetically manipulate-able *S. aureus* strains (134). As such, I discovered that

AUC species maintain a similar restriction-modification system that can be found within the MHL (**Figure 10**). The AUC restriction enzyme gene *hsdR* shares 30% identity and 52% similarity in amino acid sequence with the *S. aureus hsdR*. I conducted further comparative analysis of the AUC restriction-modification system and uncovered several patterns. First, the specificity subunit, *hsdS*, which determines the recognition sequence for the restriction enzyme, comes in a pair (two different copies) within strains (**Figure 19**). These two copies are always split by at least one toxin-antitoxin pair. Second, the *hsdS* genes are unique to each species, sharing little homology across species (**Figure 10**). Third, large ICEs will sometimes insert in the middle of the modification gene, *hsdM*. Finally, some strains are missing the restriction-modification operon entirely, which also means no *hsdM* gene for ICEs to insert into. I hypothesize that disabling this system may allow for the genetic manipulation of AUC strains, similar to how it was accomplished in *S. aureus.*

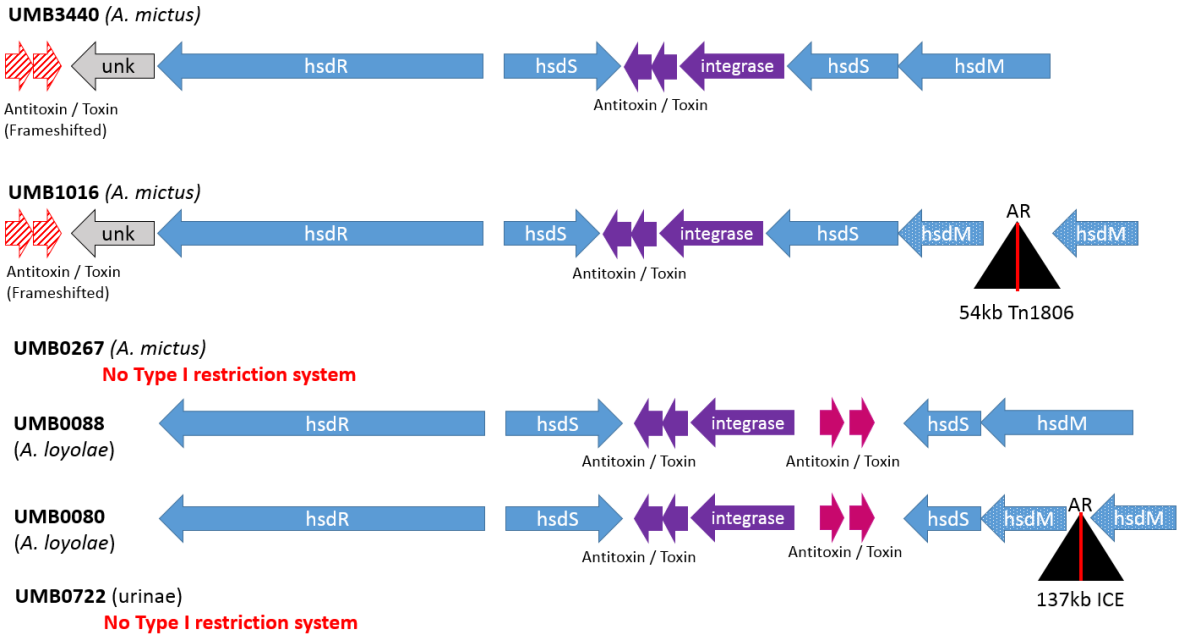


Figure 19. AUC Restriction Modification System. Comparison of syntenic regions containing the restriction modification system within the MHL. AR = antibiotic resistance gene

Integrative and Conjugative Elements

Although an AUC strain amenable to DNA manipulation in the laboratory remains out of reach, it does not mean AUC isolates do not engage in DNA exchange themselves. As mentioned before, strains of the same species can vary in their DNA content in large part due to mobile genetic elements, such as prophages, transposons, and ICEs. These ICEs in particular can introduce large amounts of extra DNA content capable of changing the bacterium's behavior (135). As seen in **Figure 19**, these ICEs can reach as large as 130kb in size, bringing a substantial number of ORFs as cargo into the host chromosome. I hypothesized that these mobile genetic elements may be changing the behavior of AUC strains, and I investigated whether the types of ICEs were species-specific.

Again, analyzing the same genomes from **Table 3**, I found ICEs were very common with most genomes possessing at least one partial or full ICE. These ICEs were not diffuse throughout the chromosome, but instead confined to distinct loci likely due to integration being a site-specific phenomenon. Analysis of all ICEs revealed that they belonged to the Tn916 and Tn1806 ICE superfamily commonly found in *Streptococcus* and *Enterococcus* species where they were first annotated (136, 137). These vectors are infamous for carrying genes that confer antibiotic resistance, such as to tetracycline and aminoglycosides. Indeed, comprehensive antibiotic resistance database (CARD) analysis detected the antibiotic resistance gene *ermA*, predicted to confer resistance to macrolides, in many of the AUC ICEs.

To evaluate whether strains were resistant to macrolides, I tested 189 clinical AUC isolates for their susceptibility to low (10μg/ml) or high (100μg/ml) azithromycin in NYCIII media (**Table 5**). 150 (79.4%) isolates demonstrated resistance at 10 μg/ml azithromycin.

However, when evaluated at the high concentration of 100 µg/ml azithromycin, 85 (45.0%) isolates demonstrated resistance (**Figure 20A**). Further analyzing the strains that survived at the high concentration, isolates were compared based on isolation year with resistances ranging between 29% and 64% (**Figure 20B**). To evaluate a genetic basis for the high rate of macrolide resistance, a PCR test identifying the *ermA* gene was conducted on the same 189 AUC isolates. The PCR primers were designed based on the *ermA* sequence detected by the CARD analysis. *ermA* was detected in 89 (47.1%) isolates, which was similar to the 100 µg/ml azithromycin resistance result of 85 (45.0%). Comparing these two experiments, 95.8% of isolates matched both the resistance phenotype and *ermA* genotype (**Figure 20C**). Of the eight isolates that did not agree, two demonstrated a positive phenotype resistance but a negative *ermA* genotype and six demonstrated a negative phenotype resistance but a positive *ermA* genotype. When stratified by species identity, *ermA* presence was detected in 44% of *A. loyolae*, 56% of *A. mictus,* 36% of *A. tenax*, and 43% of *A. urinae*.

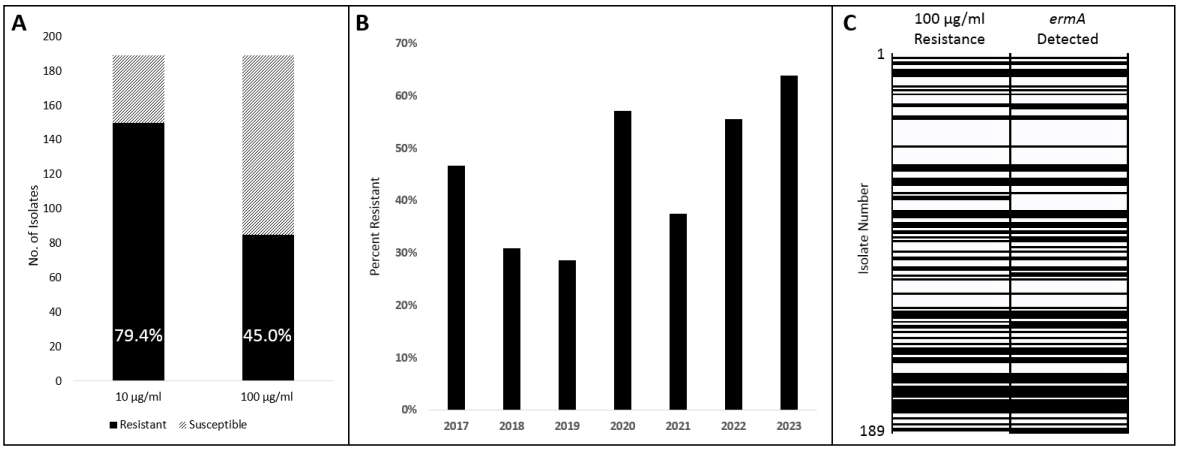


Figure 20. Macrolide Susceptibility and *ermA* Gene Presence. Azithromycin susceptibility and *ermA* presence among AUC isolates. A) Susceptibility of AUC isolates out of a total of 189 tested at 10 µg/ml and 100 µg/ml. B) Susceptibility of AUC isolates tested at 100 µg/ml compared by isolation year. C) Concordance of phenotype (100 µg/ml) with genotype (*ermA* detection) of all 189 isolates. Isolates descending by isolation order with isolate No. 1 earliest. Solid black indicates a positive result from either test.

Comparative genomic analysis revealed that ICEs containing the resistance gene integrated into AUC genomes at three distinct loci (**Figure 21A**). ICEs inserting at syntenic loci generally shared a high degree of nucleotide identity; however, very little identity was observed between ICEs of non-syntenic loci. As one example, all ICEs found to insert within the *hsdM* type 1 restriction enzyme gene (locus 1) shared nearly 100% gene identity in ICE structural genes. But these ICEs only shared three open reading frames with ICEs inserted at locus 2 and only the *ermA* gene with ICEs at locus 3. Within the ICEs, the *ermA* gene often was inherited in combination with an upstream leader peptide and a downstream aminoglycoside 3'-phosphotransferase (APH (3')) resistance gene (**Figure 21B**).

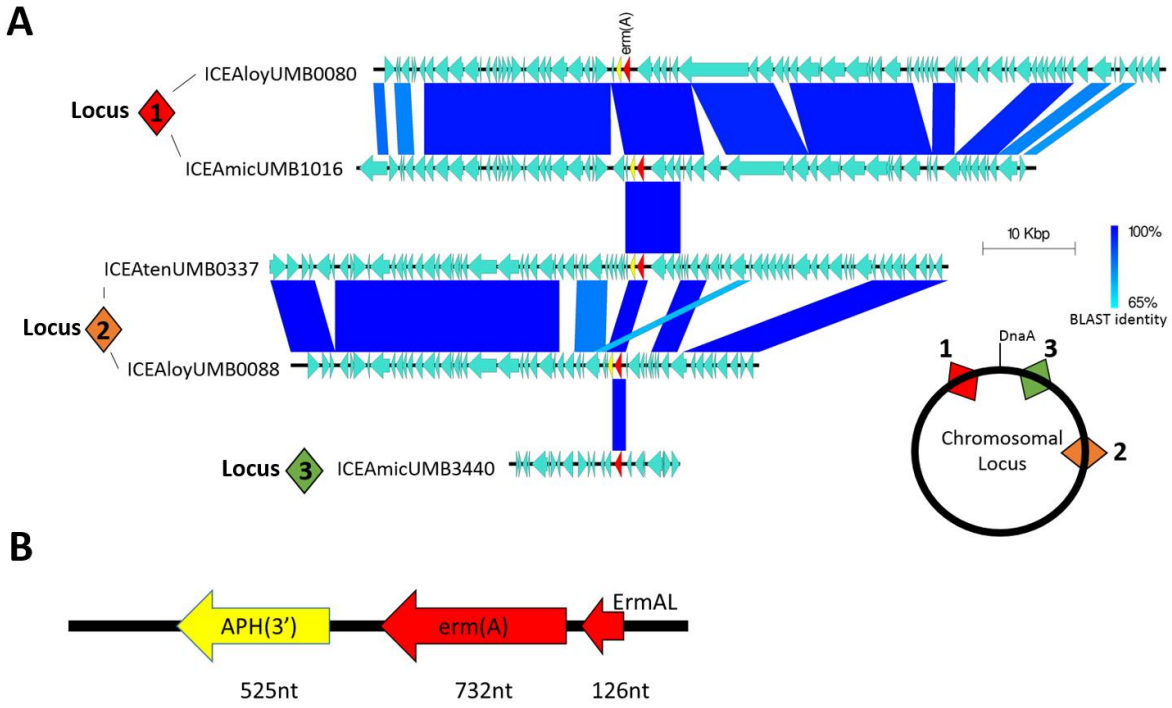


Figure 21. Macrolide Resistance Encoded in ICEs. A) Homology comparison between different ICEs found in AUC isolates. Numbered square diamonds indicate syntenic loci for chromosomal insertion of each ICE. B) Commonly inherited group of genes containing the *ermA* resistance gene.

Taken together, all these findings would suggest that a very strong association exists between horizontal gene transfer and macrolide antibiotic resistance inheritance. Due to the nature of this type of ICEs, it is possible that they were inherited either from other *Aerococcus* species or even from different species (138). One ICE in particular, ICEAmicUMB3440, was found to be 97% identical to ICESag066, an ICE first documented in *Streptococcus agalactiae* (139). As for species patterns, there does not appear to be any major biases for a particular AUC species, making macrolide resistance a commonly shared characteristic within the AUC.

Biofilm Phenotype of the *Aerococcus urinae* Complex

In looking for major behavioral differences between species that may be directly linked to pathogenesis, I investigated potential determinants of biofilm formation in the AUC. As Dr. Evann Hilt had previously annotated, AUC isolates demonstrate different forms of biofilm activity, ranging from dense flocking aggregates to no aggregation at all (98). Furthermore, she found that these behaviors could be organized by genomic group that are now referred to as the different AUC species. Thus, I hypothesized that the AUC species groups that possess stronger biofilm behavior have unique biofilm genetic features that may enable greater pathogenic potential through adherence.

Although Dr. Evann Hilt's work primarily characterized AUC strain biofilm behavior *in vitro*, other groups had found similar biofilm behavior *in vivo* both in the context of infective endocarditis and UTI. Using fluorescent probes, Yaban et al. visualized "impressive" biofilms within patient heart valve tissue (100). Using proteomics, Yu et al. analyzed biofilms isolated from urethral catheters (99). However, I was not able to replicate these characterizations of such biofilm behavior in the laboratory environment under standard growth conditions used for

non-fastidious microbes. I hypothesized that the lack of robust AUC growth was due to missing environmental requirements that would better simulate the internal human environment.

Clinical and Laboratory Standards Institute (CLSI) guidelines recommend growing *Aerococcus* isolates in CAMHB supplemented with LHB (5% v/v). As such, I found that the supplementation of horse blood greatly increased bacterial growth (**Figure 22A**). However, blood renders media opaque preventing the measurement of bacterial growth via spectrophotometry. It had been noted that plasma was sufficient to induce biofilm formation in *Aerococcus* by Shannon and co-authors (101), and serum-based mediums could robustly support *in vitro* growth and biofilm formation of fastidious vaginal microbes (140). As such, I evaluated the use of NYCIII, a horse serum-based medium, as a laboratory-friendly growth medium to support biofilm formation in AUC isolates. As seen in **Figure 22A**, NYCIII yielded even greater CFU/ml in *A. tenax* UMB3669, the "hyper"-flocculating biofilm species, than in CAMHB-LHB. Because NYCIII is transparent, spectrophotometry in 96-well format could evaluate the growth patterns of isolates, demonstrating a clear improvement in growth of *A. tenax* UMB3669 in NYCIII (**Figure 22C**) compared to the medium Dr. Evann Hilt used as her standard (i.e., Brain-Heart-Infusion; BHI) (**Figure 22B**).

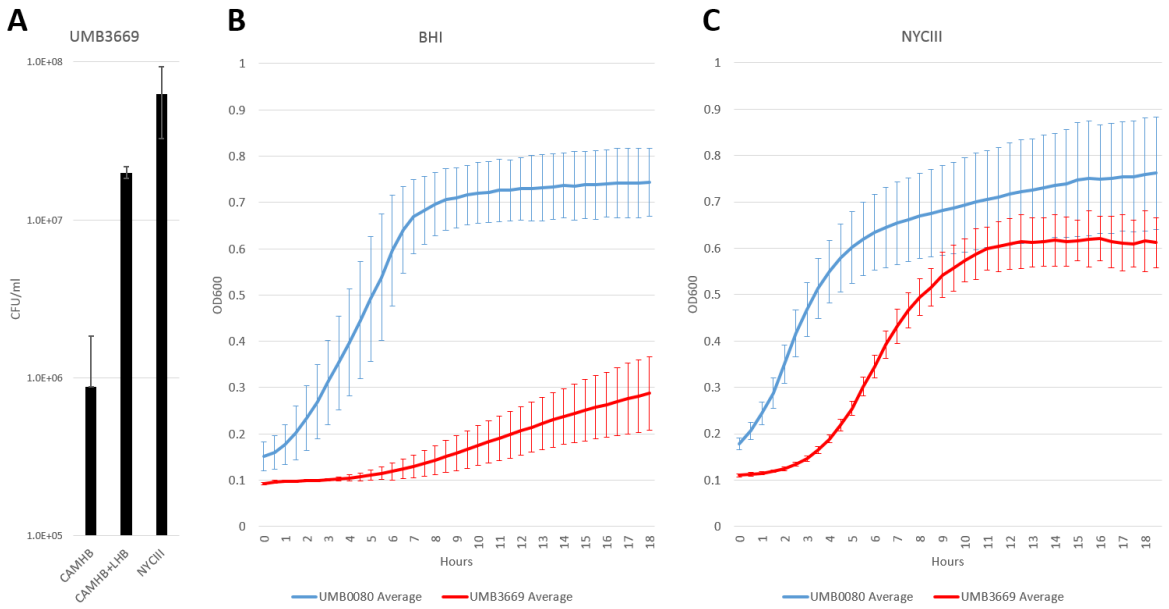


Figure 22. Growth Comparisons of UMB3669 *A. tenax* in Different Media. A) Comparison of UMB3669 growth in CFU/ml after 48 hours in selected media at 37˚C with shaking at 5% supplemented CO_2. B) Growth of UMB0080 and UMB3669 over 18 hours in BHI medium at 37˚C with shaking at 5% supplemented CO_2 compared with C) the same strains in same conditions but grown in NYCIII medium.

When studied in a larger volume, 6-well format, *A. tenax* forms dense aggregates that fall out of solution when grown in NYCIII. In this format, it can be appreciated that UMB3669 grown in NYCIII can reach higher growth turbidities at earlier times before rapidly falling out of solution when compared to growth in BHI (**Figure 23A**). In this visualization, individual lines representing a single well indicate the increase in medium turbidity as the bacteria grows; however, the sudden drop off of in turbidity (indicated by the red line) demarcates the common point at which the bacteria rapidly fall out of solution to form aggregates as seen in the photo (**Figure 23B**). The chaotic lines past the red demarcation line are due to the spectrometer laser either hitting or missing these aggregates that move around freely due to the shaking of the 6-well plate.

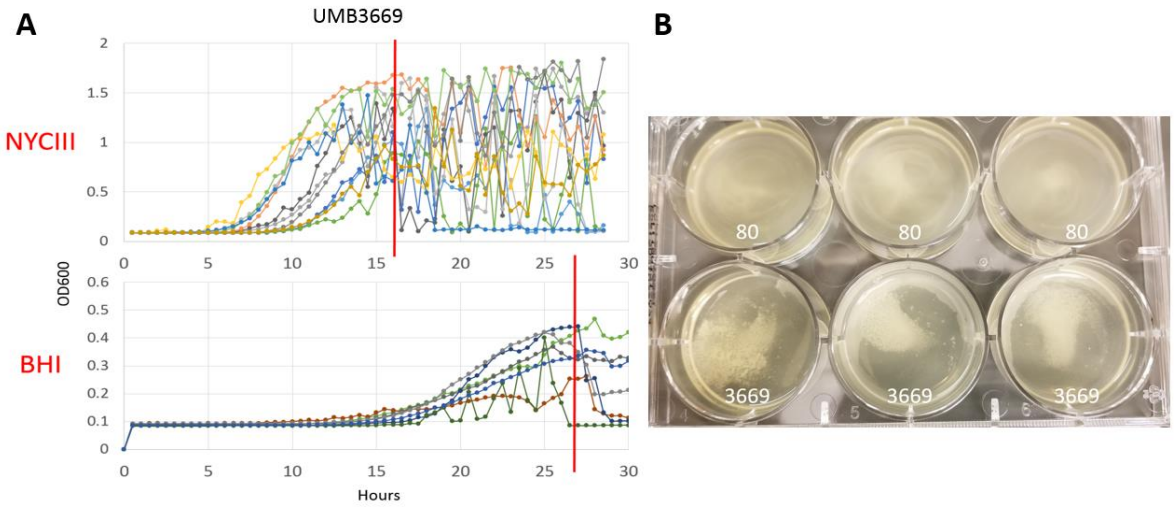


Figure 23. Biofilm Behavior of UMB3669 in NYCIII Medium. A) Growth curves of individual UMB3669 cultures grown in 6-well plates with selected media at 37˚C with shaking at 5% supplemented CO_2. Red line indicates aggregation point. B) Picture of 6-well plate culture of UMB0080 compared to UMB3669 at 48 hours.

These aggregations are even better appreciated when the bacteria are grown in large volumes such as in 100ml of NYCIII within 250ml Erlenmeyer flasks (**Figure 24**). These aggregations are less than 1mm in size and resist dissolving back into solution even with vigorous shaking. For such drastically different behaviors to be observed between the different AUC species strongly suggests a genetic basis for the behavior. Additionally, there appears to be some component within NYCIII absent in BHI that allows for more robust growth and biofilm production.

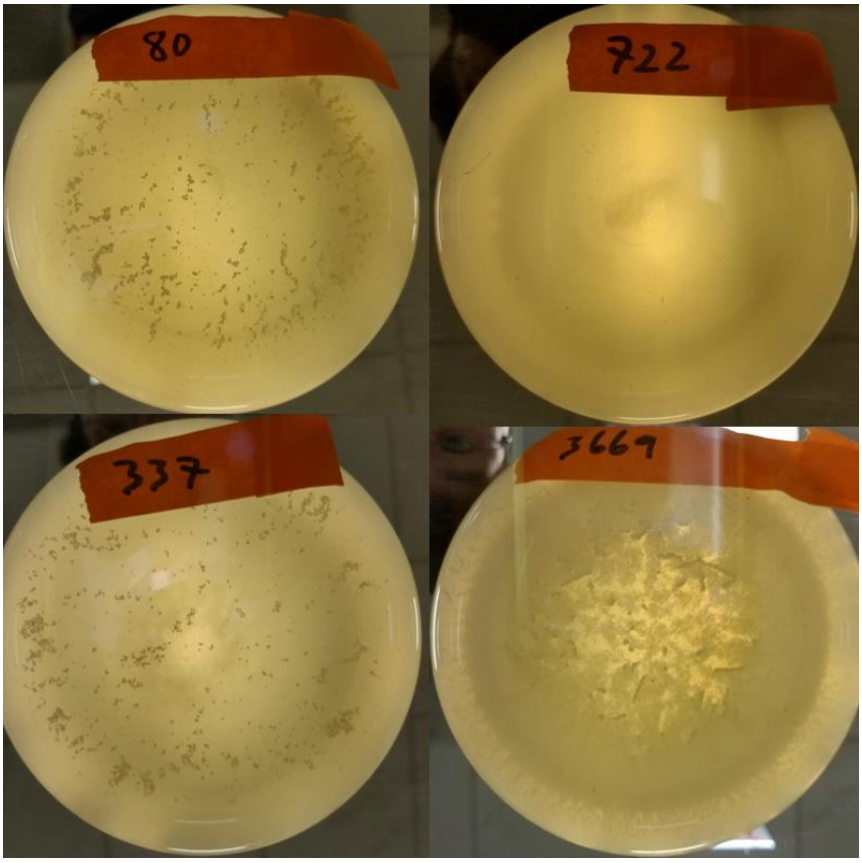


Figure 24. Flocculating Biofilm of AUC in Flasks. Bottom-up images of 250mL Erlenmeyer flasks of UMB0080, UMB0722, UMB0337, and UMB3669 comparing flocking aggregates after 10 hour growth in 100mL NYCIII at 37˚C with shaking at 5% supplemented CO_2.

Aerococcus urinae Complex Surface Proteins

To investigate the mechanism of the different biofilm behaviors between AUC species, I assessed the comparative genomics analysis for species-specific genes contributing towards biofilm behavior. Although I had previously annotated that the capsular surface polysaccharide operon appeared unique to each species, I was unable to resolve any polysaccharides through extraction methods (**Figure 9**). Additionally, electron microscopy from previous studies did not visualize any capsule around cells (117). Thus, I hypothesized that the strong aggregating behavior demonstrated in AUC species is protein-based.

When analyzing all of the *Aerococcus* predicted surface proteins, the one that stands out the most is the predicted LPxTG protein belonging to the accessory Sec operon. LPxTG proteins are a class of cell-wall anchored adhesins commonly annotated in Gram-positive pathogens that confer the bacterium with host-surface adhering abilities (141). This operon is dedicated to the assembly and secretion of very large surface proteins that are often post-translationally modified with sugar moieties at serine residues (142). These surface proteins have been found to be virulence factors in pathogens such as *Streptococcus pneumoniae*'s Pneumococcal serine-rich repeat protein (PsrP) that confers the bacterium with the ability to adhere to host-surfaces (143). Serine-rich repeat proteins follow a common structural pattern with a C-terminal wall anchoring motif, a large serine-rich repeat (SRR) region, a substrate-specific binding region, a secondary serine-rich repeat region, and an N-terminal secretion signal (**Figure 25)**.

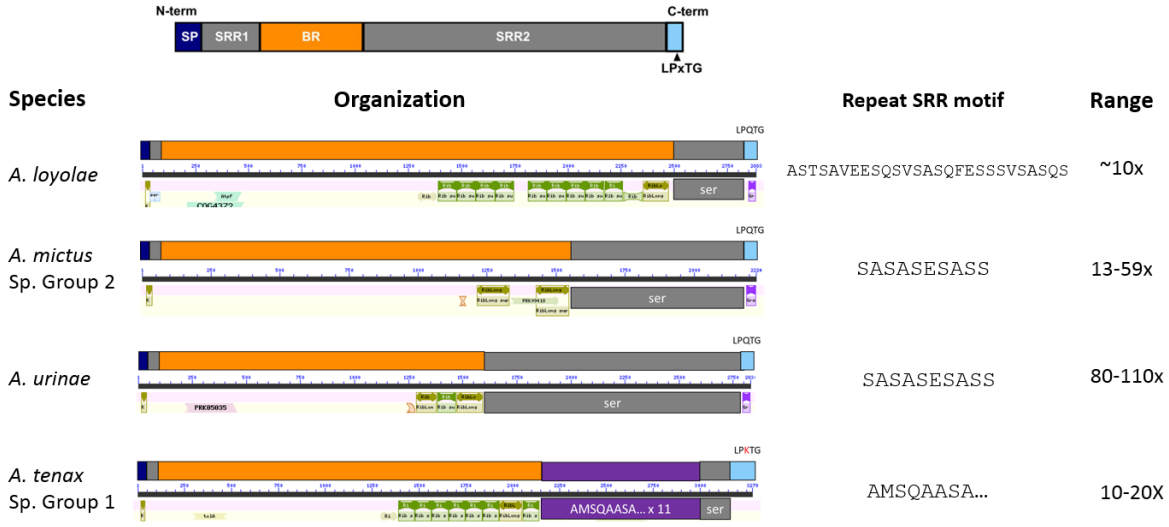


Figure 25. AUC Serine Rich Repeat Surface Protein. Comparison of AUC PsrP predicted functional regions: SRR [grey] = serine rich repeat, BR [orange] = binding region, SP [dark blue] =signal peptide, *A. tenax*-unique binding region [purple]. Protein lengths not to scale.

The SRR protein within AUC isolates is species-specific and can be differentiated by their SRR motif, as well as the unique binding region. The number of times the SRR motif is repeated

varies between strains of the same species. For example, the decapeptide motif found in *A. mictus* strains has been annotated to repeat as few as 13 times and as high as 59 times. The binding regions are also species-specific and possess varying numbers of Rib-domains. These domains have been commonly found in MSCRAMM (microbial surface components recognizing adhesive matrix molecules) proteins that confer substrate-specific binding to host-surfaces (144). The largest SRR protein is found in *A. tenax* strains, surpassing 3500 amino acids in length. The *A. tenax* SRR is also the most distinct compared to the other AUC species with a very different SRR motif that possesses fewer serine residues.

The organizational arrangement and makeup of the accessory Sec operon itself also varies by species (**Figure 26**). *A. tenax* possesses species-specific glycosyltransferases that are different from the other AUC species. Oddly, *A. loyolae* strains possess only one predicted glycosyltransferase within this region and, even then, it is often predicted to be defective with an internal stop codon. Thus, I hypothesize that the glycosylation patterns of the SrrP protein are also species-specific.

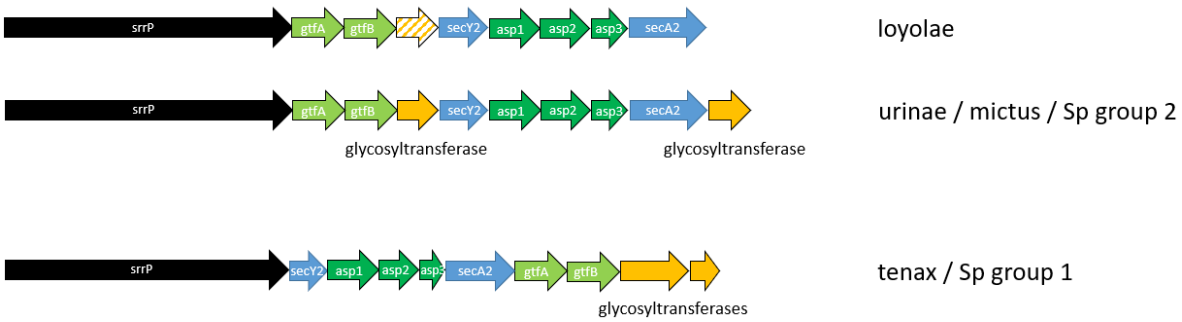


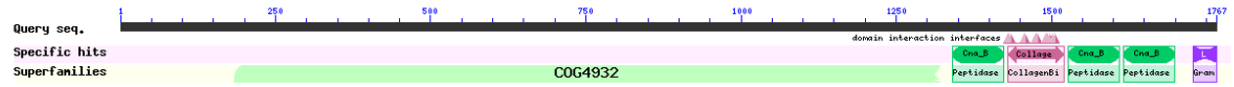
Figure 26. Organization of Accessory Sec Operon by AUC Species. asp refers to "accessory secretory protein" and not to be confused with ASP of "*Aerococcus* Surface Protein". Shaded arrow indicates predicted non-functional gene product.

Although the SRR protein is indeed likely to be involved in surface attachment, the LPxTG surface protein I hypothesize to be most likely involved in flocking self-aggregative

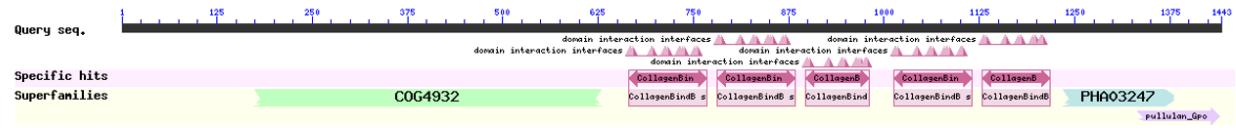
behaviors is one of much smaller length, ranging from 900 to 1500 amino acids (**Figure 27**). This particular protein has high domain homology with self-aggregative proteins of other Firmicute/Bacillota species that colonize humans. The arrangement of functional domain COG4932 with collagen-binding domains is very similar to the plasmid-encoded aggregation promoting factor AggE of *Enterococcus faecium* (145) and the *Lactococcus* aggregation protein AggL (146). Thus, to follow this naming theme, I will refer to this protein within AUC as the *Aerococcus* aggregation protein AggA.

As to be expected, the *aggA* gene is AUC species-specific with variation in the predicted binding domains near the C-terminal end of the protein (**Figure 27**). Within *A. tenax*, the gene encodes five copies of a collagen binding domain, while the *A. mictus* gene only encodes one. The *A. loyolae* strain UMB0080^{T} is predicted to possess a defunct form of the protein with an internal stop codon.

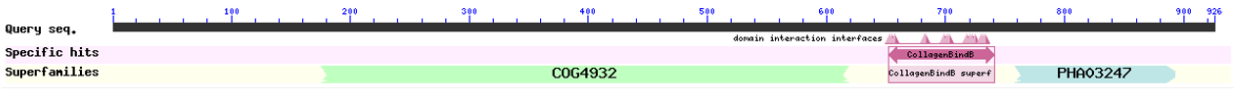
L. lactis subsp. *lactis* BGKP1 on pKP1



A. tenax UMB3669



A. mictus UMB3440



A. loyolae UMB0080

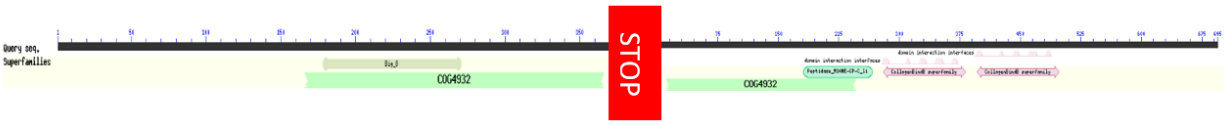


Figure 27. Comparison of Predicted Aggregating Promoting Factors. Comparison of predicted functional domains. Protein lengths not to scale

To test the aggregative abilities of this protein, I sought to conduct heterologous expression of the *A. tenax* UMB3669 *aggA* gene in an *aggL*$^{-}$ strain of *Lactococcus lactis* via the expression plasmid pNZ7021 (MoBiTec). To heterologously express *aggA* on the surface of the cell, I replaced the N-terminal secretion signal with the *L. lactis* secretion signal USP45 (147) and inserted the cloned gene into pNZ7021 via restriction digest. The resulting plasmid was about 7.5kb, and I attempted 18 times to electroporate the vector into *L. lactis* strain NZ1330. I was ultimately unsuccessfully despite testing a variety of electroporation conditions. Each attempt included at least 15 electroporation events in separate 0.2cm cuvettes varying ratios of plasmid DNA (µg) to bacterial cells of 1:1000, 1:500, 1:200, 1:100, 1:50, and 1:20. With constant settings of 25µF and 200Ω, varying voltages were also tried at 0.5kV, 1.2kV, 1.8kV, and 2kV. The electroporation buffer was kept constant in all experiments. The entire electroporated 50µl

volumes were plated on BHI agar plates with 100µg/ml chloramphenicol and incubated for 120hours. 10µg/ml chloramphenicol BHI plates were also attempted with negative results. The empty vector was able to be consistently taken up to yield 5-10 viable colonies per plate with a DNA to bacteria cell ratio of 1:200 and electroporation voltage of 1.2kV. Further attempts may yield successful colonies with the experimental vector, but I was forced to halt due to time constraints.

As a final experiment to evaluate the contribution of the AUC surface proteome to adhesive behaviors, I tested bacterial binding to host surface substrates via crystal violet binding assay (**Figure 28**). Substrates tested included host factors fibronectin, fibrinogen, and collagen, bladder glycosaminoglycans hyaluronic acid and chondroitin sulfate, and HBLAK urothelial cells. It was expected that binding would decrease when the bacteria were pretreated with proteinase K, indicating protein-involvement in substrate binding. This was largely true for all AUC species and substrates tested except for *A. loyolae* which consistently demonstrated increased binding after proteinase K treatment. As for binding abilities, *A. tenax* consistently had the highest binding activities of the AUC species. Besides the urothelial cells, the strongest of these affinities was for fibrinogen.

Figure 28. Binding Assay Results. Y-axis represents OD570 measurements of crystal violet staining of adherent bacterial cells.

CHAPTER THREE

DISCUSSION

Distribution of AUC Species

Through the sequencing and comparison of AUC genomes, the nuances in both genetic and phenotypic diversity are now better appreciated. Indeed, the advent of accessible sequencing technologies has significantly expanded the genus of *Aerococcus* as a whole. However, it is still important to note the past contributing efforts of those such as Dr. Jens Christensen and Dr. Evann Hilt whose initial characterizations prompted such investigations. With clearer classification and identification criteria, further elucidation of the new AUC species' behaviors is now possible, particularly as to their role in human health and disease.

When evaluating the results of the meta-analysis, *Aerococcus* detection by EQUC are revealed to be at rates higher than that previously reported. This was particularly true for the detection rate within individuals diagnosed with UTI, where 11.18% was substantially higher than the 0.8% (111) or 0.4% (148) reported by older studies. It was also true for UUI patients, where the finding of increased *Aerococcus* detection has been supported by other studies (98). One, assessing the urethral microbiome, found that females with UUI had an *Aerococcus* detection frequency of 36% (n = 34) in their urethral microbiome (149), similar to the meta-analysis finding of 33.6%. The significance of this rate is unknown as the pathogenesis of UUI is still undetermined; whether the presence of certain microbes is a contributing factor or merely a byproduct of altered bladder environment is poorly understood amongst urinary tract

diseases. Similarly, the significance of *Aerococcus* species being co-isolated with Acinomycetaceae species is also unknown. There may exist a synergistic relationship between the two involving cross-feeding metabolism and/or biofilm behavior. Within the bladder, biofilms are commonly polymicrobial communities, particularly in association with indwelling catheters (150, 99). Thus, it would be of clinical significance to investigate AUC species' relationship with Actinomycetaceae for any pathogenic synergisms.

The ANI analysis of 115 *Aerococcus urinae* genomes has led to the announcement and acceptance of three new species, *A. tenax*, *A. mictus*, and *A. loyolae* along with the emendation of the species description of *A. urinae* (123). Not all genomes were able to be categorized into one of the four AUC species, requiring the creation of two placeholder groups, Species group 1 and Species group 2. After using the newly developed multiplex PCR test on 189 unknown *Aerococcus* isolates, three more strains could be identified as belonging to one of these placeholder groups. Thus, it is possible that members of these placeholder groups are exceedingly rare. Alternatively, groups may be restricted by geographic sampling region. It is apt to recall that the initial investigation into biotype II (now *A. tenax*) by Dr. Jens Christensen was prompted by his comparisons of European *A. urinae* strains with North American strains in 1997 (96). Thus, there may exist more novel groups in unsampled regions of the world.

As such, some of the remaining unknown isolates could not be identified by the PCR test or ANI may represent novel *Aerococcus* groups not identified in the ANI study. As seen in the study of the variable CPS locus, rare exceptions to the species patterns were found where the *A. mictus* CPS operon pattern was observed in strains identified as *A. tenax* and *A. loyolae* by ANI (**Figure 7**). These outliers may have arisen through recombination events via horizontal

gene transfer, indicating that basing species identification on certain species-specific loci over others may result in problematic results. Evidence of horizontal gene transfer playing a role in speciation can be observed in the GC percentage graph in **Figure 5**. Regions of non-homology map to lower-than-average GC percentages, particularly within the MHL, suggesting this region is of foreign origin. Horizontal gene transfer appears to be an actively dynamic phenomenon within AUC species as seen with the ICE analysis, indicating the potential for future changes in species taxonomy (151).

The concept of time and place playing a role in distribution and/or evolution of AUC species is one that has been previously explored by Dr. Carkaci and colleagues, where they were able to generate a phylogenetic tree based on isolation period of 40 *Aerococcus* isolates within Denmark (97). These same strains were included in the ANI analysis, and I found that their phylogenetic clades matched with the species groupings of *A. urinae, A. tenax*, and *A. mictus*, emerging in that order chronologically. Interestingly, their study did not identify any *A. loyolae* strains. Because their study only assessed isolates up to the year 2015, it is possible that *A. loyolae* as a species emerged after this time period. Alternatively, it also possible that *A. loyolae* strains are exceedingly rare in Denmark and more common in the United States. Larger clinical studies investigating species distributions across geographical sites are warranted.

Research on this topic of the relationship between geography and bacterial evolution is sparse, but it has been proposed that formation of ecologically distinct populations, or cladogenesis, can rapidly arise due to local adaptation (152). Extrapolating upon this theory, it is possible the local lifestyles and diet of United States individuals are different enough from European individuals such that they foster different AUC species. In animal studies, the

geographic impact on gut microbiota has been identified as one of the most important factors on composition (153, 154, 155). Whether this is also true for the human urinary microbiome remains to be seen.

Despite the greater appreciation of *Aerococcus* diversity explored in this study, there still remain much larger ecological questions to be addressed such as to the natural reservoir for these bacteria, how they colonize the urogenital tracts of mammals, and the specificity of host selection. If it is true that the mammalian host is actually the primary reservoir of these organisms, then it needs to be investigated whether host-to-host transmission is possible, such as is the case with other urogenital pathogens. As more environmental and host samples are analyzed through metagenomic sequencing techniques, these questions may be answered in the near future.

Metabolic Behaviors of AUC Species

In the gut, a large factor for microbiota composition is nutrient availability which is ultimately shaped by the diet, genetics, and lifestyle of the host (156, 157). This association between diet and gut microbes is more obvious since there is a direct route for the food an individual consumes to reach the hosted bacteria. However, this association is not as clear when it comes to the urinary microbiome because urine, by its own nature, is primarily composed of the ammoniacal waste products thrown away by the host. The kidneys are highly specialized for the reuptake of sugars with the upper limit of normal urine glucose levels only at 25mg/dL (158, 159). Thus, it is not obvious as to the primary carbon source for many of the urinary microorganisms.

Because urine presents as a nutrient-poor growth environment, urinary bacteria may be forced to specialize in metabolism of very specific substrates. In uropathogenic *E. coli*, studies have demonstrated that the bacteria can survive on urine amino acids and peptides during urinary tract infection (160). A separate host-excreted metabolite, D-mannose, has been suggested by some studies as inhibitory to *E. coli* while promoting the growth of protective bacteria (161). Alternatively, rather than deriving their nutrients from the urine, it has also been hypothesized that bladder bacteria may derive their carbon from the highly glycosylated urothelium surface glycocalyx and glycosaminoglycans (162, 163). Thus, whether the carbon source is urine-derived or host surface-derived, it may be that bacteria must utilize specialized metabolic systems to successfully colonize parts of the human urinary tract.

This theory may explain the species-specific sugar phosphotransferase systems found in the comparative genomic analysis of the AUC species. With *A. tenax* specializing in aryl β-glucosides and *A. mictus* specializing in D-mannose, the different species may be better adapted towards urinary tract niches where these nutrients are available in the local environment. Another finding was the L-ascorbate degradation operon found only in *A. tenax* and *A. loyolae*, potentially indicating another favored carbon source. Since water-soluble vitamins such as vitamin C (ascorbic acid) are excreted in urine (164, 165), it is possible that these two species may be better adapted to colonize individuals with higher vitamin C intake. Thus, future studies to elucidate the full substrate specificity of each AUC species may describe the respective preferred environmental niche within the host.

Biofilm and Adherence Behaviors of AUC Species

The biofilm behaviors exhibited by *Aerococcus* strains have been hypothesized to be directly related to pathogenicity in diseases such as UTIs and infectious endocarditis. With the work conducted in this study, it was revealed that biofilm behavior is species-specific with *A. tenax* strains forming hyper-flocking aggregations, particularly when grown in NYCIII. As demonstrated in **Figures 22** and **23**, it is the addition of serum as a medium component that greatly enhances the growth and biofilm behavior of *A. tenax* UMB3669. As such, particular growth factors within serum may be responsible for promoting these aggregative behaviors. Alternatively, the stimulus for biofilm formation may be density-based, and the nutritional boost that serum provides may be what ultimately leads to reaching quorum faster. Ultimately, identifying the serum component(s) that are responsible for this stimulus may explain the pathogenic ability of *Aerococcus* to advance from UTI to endocarditis.

In *A. viridans*, 75–150 nm 'hairs' or fibrils have been described on the cell wall that are sensitive to trypsin (166). Additionally, Dr. Shannon and colleagues demonstrated that biofilm formation and platelet aggregation in *A. urinae* could be abrogated with trypsin treatment (101). Lastly, two studies utilizing mass spectrometry detected high levels of LPxTG surface proteins expressed in biofilm-associated *A. urinae* isolates (99, 102). These authors hypothesized that these surface proteins could be *Aerococcus* virulence factors involved in adhering to host surfaces and indwelling catheters.

Dr. Senneby and colleagues noticed that two of these surface proteins, which they named *Aerococcus* Surface Protein (ASP)-1 and ASP-2, had strain-dependent gene arrangements (102). As alluded to in my development of the multiplex PCR test, these ASPs are

species-specific, and I was able to confirm that the four patterns of ASP genes that Dr. Senneby and colleagues found align with the new species designations of the AUC. *A. mictus*, *A. urinae*, and *A. loyolae* all encode for their own species-specific copy of both ASP-1 and ASP-2, while *A. tenax* only possesses ASP-1. The two named ASP proteins are not the only surface-anchored proteins that are encoded by AUC genomes, and since *A. tenax* demonstrates the greatest biofilming behavior yet is missing one ASP, I hypothesized that a different LPxTG protein may be responsible for the hyper-aggregating capabilities. As such, I choose to investigate the AUC surface protein AggA.

The behavior of forming aggregations is often associated with quorum sensing and horizontal gene transfer in related species such as *Streptococcus* and *Enterococcus* (167, 168). For conjugation to successfully occur, donor and recipient must be closely bound together usually via a type VI secretion system that relies on the expression of surface proteins for adhesion (169). Thus, it is possible that the aggregating behavior demonstrated in AUC species is also implicated in horizontal gene transfer utilizing the adhesive properties of surface adhesins. As seen in **Figure 27**, the predicted motifs of the particular LPxTG surface protein gene I referred to as *aggA* is highly similar to the aggregation factors of other Gram-positive species. Proteins such as AggA of *L. lactis* and AggE of *E. faecium* are also involved in horizontal gene transfer and confer self-aggregating abilities to the bacteria when expressed (170, 171). Of note, both of these proteins are plasmid-encoded within their corresponding organisms while *aggA* is found on the chromosome within AUC strains. As such, the regulation and expression of this gene may differ from that of its plasmid-borne cognates.

Since I could not obtain any viable colonies when cloning the gene into a *L. lactis* strain via a heterologous expression vector, it is possible that overexpression of *aggA* is fatal. In a study of *Lactobacillus gasseri*, it was determined that overexpression of aggregation-promoting factor presented a metabolic burden leading to high rates of plasmid instability (172). Thus, success may be more likely in future experiments if the *aggA* cloned gene is kept at lower or under inducible expression. It is more likely, however, that the conditions I used for electroporation were not favorable enough to yield efficient transfer of the constructed vector. The efficiency of transformation with the empty vector alone is very low, and so it is reasonable that this efficiency is further reduced when attempting to transform with a vector that is several times larger in size. A larger scale reattempt in investigating the characteristics of a cloned *aggA* gene is warranted.

Because *aggA* is only one protein within a suite of surface-anchored adhesins predicted to be expressed by AUC strains, it is highly likely that these other proteins have their own binding substrates involved in biofilm and surface adhesion. As seen in the crystal violet binding assay, each of the AUC species have different binding affinities for the tested substrates. For example, *A. tenax* had the strongest binding affinity for fibrinogen and fibronectin and this affinity was abrogated with proteinase K treatment. In *Staphylococcus aureus*, this binding ability is conferred by a specific LPxTG protein, Fibronectin binding protein A which is highly involved in pathogenesis (173). Thus, it would be of clinical significance for future studies to annotate each of the AUC surface proteins to determine their role in pathogenesis. It is possible the species-specific forms of these proteins may confer tissue tropisms leading to some AUC strains being more or less likely to cause disease. Indeed, our previous study in a mouse model

of urinary tract infection demonstrated that *A. tenax* UMB3669 and *A. loyolae* UMB0080 had stronger tropisms for kidney infection compared to *A. urinae* UMB0722 or *A. loyolae* UMB5628 (117).

Future Prospective Studies

Many questions remain, such as the wider distribution of AUC species in humans, their disease manifestations and accompanying risk factors, and their roles within the urinary microbiome. Future investigations will use the insights and techniques developed by this study to reach greater understandings of this mysterious relationship between AUC isolates and their host.

Utilizing the multiplex PCR test developed within this study, prospective studies on AUC isolates in other regions around the world will help resolve the question as to whether there is a geographic component in AUC species distribution. Prospective studies evaluating AUC isolates from bloodborne infections will also determine whether certain AUC species are more likely to cause bacteremia and endocarditis. An ongoing collaboration with Dr. Bo Nilson is seeking to resolve exactly this question in Sweden. With the complete genomes of the AUC species type strains published on NCBI, metagenomics studies can now accurately identify these species when sequencing urine samples. Such studies will contribute to determining AUC species detection frequencies in urinary tract diseases and asymptomatic bacteriuria.

Future mechanistic studies are required to elucidate the virulence factors of AUC species involved in pathogenesis. Among these genes are likely specific surface proteins involved in host tissue and catheter adherence, metabolic genes conferring tissue tropisms, and immune evasion genes. Because *Aerococcus* strains have been isolated in asymptomatic

individuals, it is likely that AUC species follows the behaviors of other opportunistic pathogens, only expressing virulence behaviors under the right conditions. Identifying what environmental stimuli trigger these pathogenic shifts is of particular importance for the clinician.

To facilitate these mechanistic studies, the development of a genetically manipulatable strain of each AUC species is vital. I hypothesized that deactivation of the restriction-modification system might enable foreign DNA uptake as was conducted in *S. aureus*. Such a task may require mutagenesis of the strains with the goal of introducing a nonsense mutation in the *hsdR* gene. An alternative approach may be exploiting the *Aerococcus* plasmid discovered to exist natively in some AUC strains. A shuttle vector could be created to express cloned genes or even introduce a CRISPR silencing system. As seen in **Figure 16**, large portions of the plasmids are unannotated via the PGAP database, potentially corresponding to regions of unknown function. These plasmids may represent members of a novel plasmid family as the plasmid replication was not readily identifiable within the PlasmidFinder database. As such, further characterization of *Aerococcus* plasmids is also warranted.

Another method of introducing DNA into AUC strains may be to artificially induce competence by exploiting the competence sigma factor. This has been conducted in *L. lactis* where the stimulus for inducing competence is still unknown; however, inducing the competence regulator ComX allows for the production of competence machinery and consequent foreign DNA uptake (174). As seen in **Figure 18**, there is strong evidence that the machinery for competence exists within AUC strains despite the competence regulator being unidentified. A future study finding this regulator may be the key for unlocking horizontal gene transfer and aggregating behaviors in AUC.

Wider Implications

The findings on the speciation and behaviors of the *Aerococcus* species investigated in this study are not limited to just these species but are likely to play out similarly in other emerging uropathogens that are poorly annotated. For example, other members of the *Aerococcaceae* family such as species of the genera *Globicatella*, *Facklamia*, and *Abiotrophia* that have also been identified in the human urinary microbiome may be better understood with similar investigations into their phenotypic behaviors and genomic features as conducted in this study.

For emerging uropathogens in general, it remains to be definitely determined whether their role in pathogenesis is one of primary culprit or minor bystander. Due to the opportunistic nature of many of these bacteria, it may be worthwhile to explore the wider collective functions the bacteria perform as a mixed community. Just as in *Aerococcus*, there may be some environmental stimulus that results in a shift of the urinary community to head towards dysbiosis. In such case, it would be more worthwhile to study how to avoid or prevent the factors that culminate in dysbiosis rather than focusing solely on eradication strategies.

Indeed, it is now well-documented that urinary bacteria antibiotic resistance is on the rise with several antibiotic classes including macrolides having little to no effectiveness in AUC species. Furthermore, due to these antibiotic resistance genes (ARGs) residing on mobile genetic elements, AUC species can potentially serve as antibiotic resistance reservoirs within a microbiome community spreading these resistance genes cross-species.

This implication also holds potential for non-human urinary microbiomes as well. As mentioned, *Aerococcus* species primarily inhabit the urogenital tract of humans and

domesticated animals. Species such as *A. vaginalis* in cows or *A. suis* in pigs may be behaving similarly to AUC species in regard to horizontal gene transfer. Agricultural practices often expose domesticated animals to antibiotics at much higher rates than humans experience, leading to high rates of resistance. In fact, a study assessing dispersal of ARGs in agricultural farmland systems found high levels of macrolide resistance genes in livestock waste (175). Thus, investigation and monitoring of emerging uropathogens not under regular surveillance may be critical in slowing the evolution of superbugs. Wider studies into the potential mobile genetic elements involved (referred to as the mobilome) will be necessary (176).

Conclusion

With the species diversity of *A. urinae* now disambiguated into the species of *A. tenax*, *A. mictus*, and *A. loyolae*, more meticulous studies of these bacteria and their contribution to human disease can take place. I hope that these results will lead to the development of better identification, detection, and diagnostic practices that may ultimately result in treatment strategies entirely preventing any *Aerococcus* infection from exacerbating to fatality. The knowledge learned in this study can be used as a template to investigate other poorly understood bacteria, particularly uncharacterized emerging uropathogens, broadening our understanding of the human urinary microbiome.

CHAPTER FOUR

MATERIALS AND METHODS

Meta-analysis of Microbe Identification Frequency of *Aerococcus*

Eight clinical studies involved in the characterization of the urinary microbiome in females with and without LUTS were analyzed for microbial detection frequencies (81, 177, 178, 179, 180, 181, 182, 183). The microbes were isolated from transurethral catheter urine samples from a combined total of 1007 urinary microbiome study participants with the EQUC protocol combined with MALDI-TOF MS. Participants were stratified into five clinical groups: UTI (N=304), UUI (N=253), SUI (N=50), IC/PBS (N=49), and asymptomatic females (N=351). The data set was analyzed with the organizational assistance of Jacob Kaindl and Caroline Gonzalez.

Phylogenetic Analysis

Table 3 lists the strains used for this study and their origins. 159 *Aerococcus* genomes were analyzed, including sequences deposited at the National Center for Biotechnology Information (NCBI) as *A. christensenii* (n=4), *A. sanguinicola* (n=11), *Aerococcus* sp. (n=11), *A. suis* (n=1), *A. urinaeequi* (n=8), *A. urinaehominis* (n=2), *A. viridans* (n=7) and *A. urinae* (n=114). Of the *A. urinae* genomes, 33 strains were isolated from urogynecology patients at Loyola University Chicago in the United States, 5 strains were from Dr. R. Facklam at the CDC, Atlanta, USA, 17 strains were from Dr. Erik Senneby in Lund, Sweden, 41 strains were from Dr. Jens J.

Christensen and Dr. D. Carkaci in Denmark, 9 strains were from PMH Schuur in The Netherlands, and 10 strains have no documented origin. All genomes are publicly available and can be retrieved from GenBank via the accession numbers listed in **Table 3**. Sequencing and assembly information is annotated in the strains' respective BioProjects at NCBI.

16S rRNA sequences from the annotated *Aerococcus* genomes were used to conduct phylogenetic analysis of *Aerococcus* species and related genera. Sequences for each of the genes were aligned using MAAFT v7.388 (184). Phylogenetic trees were derived using maximum-likelihood via FastTree v2.1.11 (185), using default parameters, and visualized in iTOL v6.3 (186). Whole genome sequences from the annotated *Aerococcus* genomes were used to conduct ANI analysis. The single-copy core genome for the *Aerococcus* strains was identified using Anvi'o v7 [187]. ANI scores were computed using pyANI v.0.2.11 (188). Analysis and visualization were conducted with the assistance of Adriana Ene and Dr. Catherine Putonti.

Biochemical Phenotypic Analysis

Aerococcus strains were examined using API 50CH and API ZYM strip tests (bioMérieux), according to the manufacturer's instructions. Multiple strains were tested for each species, and a positive test was reported if greater than 50% of strains expressed the indicated phenotype.

For assessing fermentation ability of L-ascorbate, strains were first grown overnight in 10mL NYCIII media (Proteose Peptone No. 3, yeast extract, heat inactivated horse serum, heat inactivated neonatal bovine calf serum, glucose, HEPES buffer (1M), NaCl, and water) at 37°C with 5% CO_2 and shaking at 200rpm. Cells were pelleted and washed twice with PBS before being resuspended into 100µl API GP medium (L-cystine, Tryptone, sodium chloride, sodium sulfite, phenol red, and water) pH 7.4. Suspensions were incubated anaerobically for 72 hours

at 37°C before being assessed for color change (yellow color indicated positive for fermentation).

Comparative Genomics

Three representative strains from each of the *Aerococcus urinae* complex species were sequenced for comparative genomics analysis (**Table 6**). DNA extraction and sequencing of isolates was performed by SeqCenter (Pittsburgh). For short-reads, Qiagen UltraClean Microbial Kit for extraction and Nextera DNA Flex Library Prep kit for library preparation were used before being run on the Illumina MiSeq or NovaSeq platform. For long-reads, Zymo DNA Miniprep kit for extraction and Oxford Nanopore Technology Ligation Sequencing kit V14 for library preparation were used before being run on the MiniION platform. Hybrid assembly was performed by combining filtered short- and long-sequence reads using SPAdes v3.15.4 (189), Flye v2.9 (190), and/or Canu v1.5 (191). Genomes were polished with Pilon v1.24 (192), validated with QUAST v5.2.0 (193) and circularized using Circlator v1.5.5 (194). Open reading frames were annotated via PGAP v6.6 (195). *Aerococcus* plasmids were confirmed with PlasmidFinder (196). Complete genomes were deposited and are publicly available at BioProject PRJNA316969. Assembling and depositing of genomes was completed with the assistance of Dr. Melline Fontes Noronha.

Table 6. Strains for Complete Genome Assembly

Species	Type Strain	Second Strain	Third Strain
A. tenax	UMB3669^T	UMB0337	UMB7049
A. mictus	UMB3440^T	UMB0267	UMB1016
A. loyolae	UMB0080^T	UMB0088	UMB0509
A. urinae	ATCC51268^T	UMB0722	FDAARGOS 911

To carry out the comparative genomic analysis, both software alignment and manual comparison of ORFs was conducted. Anvi'o v.7.2 was used to annotate and identify the pangenome within the *A. urinae* complex species (187). Alignment and comparison of syntenic chromosomal homology was conducted in EasyFig version 2.2.5 with default parameters (197). Pairwise comparisons within strains of the same species identified intraspecies variation and the accessory genome. Pairwise comparisons between strains of different species identified interspecies variation. Components that were found to be unique to each species representatives (while not being part of the accessory genome) were compared against BLAST searches of the publicly available *Aerococcus* genomes for consistency.

Isolation of Capsular Polysaccharides

Bacteria were grown and pelleted after 10 hours incubation in 100ml NYCIII supplemented with glucose at 37°C with 5% CO_2 and shaking at 200rpm. Cells were harvested and treated with lysozyme and mutanolysin for 16 hours at 37°C. After centrifugation, the supernatant was treated with RNase (100µg/ml) and DNase (10U/ml) at 37°C for 4 hours before treatment with proteinase K (50µg/ml) for an additional 16 hour incubation. Impurities were removed via chloroform extraction before carbohydrates were precipitated in 75% ethanol. After air-drying, the resulting visible pellets were resuspended in sterile water. Acrylamide gels

were then loaded with 25µL of sample and stained with Stains-All dye (Sigma) subsequent to electrophoresis. Capsule preparations of strains with known serotypes were used as controls for comparison of staining regions. Experiment was conducted twice.

Development of Multiplex PCR for *Aerococcus urinae* Complex Discrimination

Analysis of results from the interspecies comparative genomics were used to identify species-specific features for use in discrimination. The primers used in the multiplex PCR protocol are outlined in **Table 7**. The strains tested are outlined in **Table 5**. These isolates were collected at Loyola University Health Center from male urology and female urogynecology patients between 2017 and 2023 from various clinical studies and procured from the IRB-approved biorepository (LU 215192).

Table 7. Primers used in *A. urinae* Complex Multiplex PCR

Primer Set	Primer Sequence
Primer Set I - *A. tenax* Specific	LEFT PRIMER: 5'-CGCTACCGCTGCTAATCAAC-3'
	RIGHT PRIMER: 5'-TGTGGTTTGGAGATGAGGGT-3'
Primer Set II – *A. mictus* and *A. urinae* Specific	LEFT PRIMER: 5'-GTCCCGCTTTTCAGGCTATG-3'
	RIGHT PRIMER: 5'-GGGCTCCAACAATAGAAGCG-3'
Primer Set III – *A. urinae* Specific	LEFT PRIMER: 5'-TGCATTCCCTAACCTAGATCCT-3'
	RIGHT PRIMER: 5'-GTTCGCCAGTTTCAGGGTTT-3'
Primer Set IV – *A. loyolae* Specific	LEFT PRIMER: 5'-GGTGATGGAGCTCGGACTTA-3'
	RIGHT PRIMER: 5'-AGTAGTCAATGTCTCCACGGA-3'
Primer Set V – *A. mictus* Specific (Optional)	LEFT PRIMER: 5'-GCCATCACAATGTCAGCGAT-3'
	RIGHT PRIMER: 5'-TCAAAACTCAACGGGTGCTG-3'

The multiplex PCR master mix was prepared by combining the four core primer sets in the following recipe per 50µl reaction volume: 10x Taq Buffer (Thermo) (5µl), 25 mM $MgCl_2$

(3µl), template DNA (3µl), 10µM Primer Set I L (1µl), 10µM Primer Set I R (1µl), 10µM Primer Set II (1µl), 10µM Primer Set II R (1µl), 10µM Primer Set III L (1µl), 10µM Primer Set III R (1µl), 10µM Primer Set IV L (1µl), 10µM Primer Set IV R (1µl), 10mM dNTPs (1µl), Taq Polymerase (Thermo) (0.5µl), and H_2O (29.5µl). The PCR reactions were performed in a SimpliAmp thermocycler (Applied Biosystems). Thermocycling conditions were as follows: initial denaturation 95°C (10 min), 30 cycles of 94°C (30 secs) + 54°C (1 min) + 72°C (1 min), and final extension at 72°C (7 min). PCR products were analyzed by separation on 1.0% agarose gels. All PCR reactions and respective gels were conducted twice.

Aerococcus Pheromone Experiments

Aerococcus pheromone SLSLLVA was synthesized at >75% purity (Lifetein, LLC) and resuspended in acetonitrile. Aliquots were stored at -80°C and subsequent dilutions for working stocks were stored at -20°C. *A. urinae* strains UMB0080^T, UMB3440^T, UMB3669^T, and ATCC 51268^T were grown for 48 hours in 25mL NYCIII or chemically defined medium (CDM) supplemented with either 1µM pheromone or spent media. CDM was prepared as described by Chang and co-authors (198). CFUs were measured every hour for 8 hours starting at the 40th hour of incubation.

For assessing competence, the same strains were incubated at the same conditions in either NYCIII or CDM with 1µM addition of various DNA constructs containing antibiotic selective markers including pNE1 (199), pMU1328 (200), pKV36 and pKV47 (201), and pJMP1335 (202). Plasmid extraction was conducted via plasmid mini kit (Qiagen) according to manufacturer's instructions.

Conjugation with and without the *Aerococcus* pheromone was attempted with the engineered mini ICE conjugation system XPORT (133). This involved growing and co-culturing various ratios of donor *Bacillus subtilis* and recipient AUC strains to allow for transfer of DNA before selecting for growth on kanamycin (100µg/ml) BHI plates. Conjugation was allowed to proceed for up to eight hours incubation. Ratios of 1:1, 1:2, 1:4, 1:10, and those opposites were tested, but did not yield viable colonies. A total of 40 conjugation attempts were conducted. A similar approach was also attempted with the mobile CRISPRi system (202) with the donor plasmid of pJMP1335. Donor and recipient strains were co-cultured, and transconjugant colonies were selected on kanamycin (100µg/ml) plus tetracycline (40µg/ml) BHI plates. Conjugation was allowed to proceed for up to eight hours incubation. A total of 24 conjugation attempts were conducted with similar donor/recipient ratios without yielding viable colonies.

Integrative and Conjugative Elements and Macrolide Resistance

The identification and classification of ICEs was conducted by the same method used for identification in *Streptococcus* species (203, 204). Briefly, genomes were queried via BLAST comparison against reference ICE proteins (e.g. integrases, transposases, Type IV secretion system proteins, and relaxases). The reference set of ICE proteins of known *Streptococcus* ICE families was utilized via ICEberg 2.0 database (205). Putative ICEs were delineated by syntenic comparison. The resistance gene *ermA* within ICEs was detected with VRprofile2 (206). Homology of ICEs and *ermA* was visualized with EasyFig version 2.2.5 (197).

For surveying strains for ICEs and ARGs, the same 159 publicly available AUC draft genomes from **Table 3** were retrieved from NCBI and analyzed with the comprehensive

antibiotic resistance database (CARD) version 3.2.6 (207). Analysis of results was limited to perfect hits with default settings.

Testing of macrolide resistance and PCR testing for *ermA* gene presence was assayed in the same 189 isolates outlined in **Table 5**. Macrolide susceptibility phenotypes of all 189 isolates were determined by broth microdilution technique in microtiter plates with 10-fold antibiotic dilutions. Overnight bacterial growth was inoculated in NYCIII media containing azithromycin with maximum antibiotic concentration at 100 µg/ml. The microtiter plates were then incubated within a BioTek Epoch 2 Microplate Spectrophotometer in aerobic conditions with 5% supplemented CO2 at 37˚C with shaking at 200 rpm for 24hrs. Optical density was measured every 30 minutes at 600nm, and minimum inhibitory concentration (MIC) was determined from the growth curves. For controls, PBS or bacteria was inoculated in NYCIII without antibiotics. All tests were conducted in duplicate.

To isolate genomic DNA from AUC isolates, 10 ml bacterial cultures were grown overnight in NYCIII media before being treated with lytic enzyme (lysozyme) at 37°C for 60 minutes. Cells were lysed using the Wizard Genomic DNA Purification kit (Promega). The final DNA pellet was rehydrated in nuclease-free water overnight at 4°C. Extracted bacterial DNA was stored at −20°C until further processing.

Primers were developed to the *ermA* gene utilizing Primer3 v4.1.0 (208) producing following sequences: Forward 5’-ACA TGA TAT TCC CTG TTT ACC CA-3’, Reverse 5’-TGG AAA TGA GTC AAC GGG TG-3’. The ICE found within UMB1016 that bears the *ermA* gene was utilized as the template. Presence of bacterial macrolide resistance gene *ermA* was detected by PCR. Each PCR reaction was carried out in a final volume of 50 µl consisting of molecular grade

nuclease-free water (35µl), 10x Taq buffer (5µl) (Thermo), 25 mM $MgCl_2$ (3µl), DNA template (3µl), 10 mM dNTPs (1µl), forward primer (1µl), reverse primer (1µl), and Taq polymerase (1µl) (Thermo). The PCR reactions were performed in a SimpliAmp thermocycler (Applied Biosystems). Thermocycling conditions were as follows: initial denaturation 94°C (10 min), 30 cycles of 94°C (30 secs) + 55°C (1 min) + 72°C (1 min), and final extension at 72°C (7 min). PCR products were analyzed by separation on 1.0% agarose gels. A PCR reaction without template DNA was used as a negative control.

Evaluation of Biofilm Formation and Substrate Binding

In 96-well format, type strains of *Aerococcus urinae* complex were incubated in 200µl of cation-adjusted Muller-Hinton Broth (CAMHB) (Thermo), CAMHB supplemented with lysed horse blood (5% v/v) (CAMHB-LHB), brain heart infusion (BHI) (Thermo), or NYCIII media for the evaluation of growth and biofilm formation. Growth was evaluated via BioTek Epoch 2 Microplate Spectrophotometer in aerobic conditions with 5% supplemented CO_2 at 37˚C with shaking at 200 rpm for 24hrs. Similar conditions were repeated for 6-well format but in a volume of 3mL of media and in 250mL Erlenmeyer flasks in a volume of 100mL media.

Binding of host factors was quantified via crystal violet binding assay. To prepare substrate binding plates, target substrate was suspended in PBS at 5mg/mL and allowed to bind to 96-well plastic plates (Corning) overnight at 10°C. Plates were then blotted with 0.5% BSA and washed with PBS. Two groups of bacteria were prepared from overnight cultures in NYCIII after being washed twice in PBS; one group was treated with proteinase K (1mg/mL) for 30mins at 37˚C before being washed again while the other group remained untreated. 100µl of bacteria were added to wells standardized to an OD of 1.0 and allowed to incubate in aerobic conditions

with 5% supplemented CO_2 at 37˚C with shaking at 200rpm. Non-adherent cells were then washed out gently with PBS before the plates were stained with 0.005% crystal violet. After incubation for 1 hour, the plates were washed twice with PBS before being analyzed in the BioTek Epoch 2 Microplate Spectrophotometer at 570nm absorbance. Substrates tested included fibronectin (Sigma), hyaluronic acid (Sigma), chondroitin sulfate (Sigma), fibrinogen from bovine plasma (Sigma), fibrinogen from human plasma (Sigma), HBLAK urothelial cells (CELLnTEC), and collagen from bovine achilles tendon (Sigma). *Staphylococcus aureus* USA300 and *Lactococcus lactis* NZ1330 were utilized as comparative controls. Experiments were conducted in triplicate.

Heterologous Cloning of Surface Protein AggA

Identification and analysis of *Aerococcus* LPxTG surface proteins was identified via PGAP v6.6 annotation (195). Functional domain prediction and comparison of LPxTG proteins was conducted and visualized in the NCBI CD-search function via the conserved domain database (208). *Aerococcus aggA* gene was identified via comparisons with reference aggregation proteins in related species of protein domain homology via CDART (209).

The *aggA* gene sequence was synthesized in a ColE1 plasmid (TWIST) and propagated in *E. coli* DH5α. Plasmid extraction was conducted via plasmid Mini Kit (Qiagen) according to manufacturer's instructions. The *aggA* gene sequence was restriction digested with SphI and SacI and ligated into the *L. lactis* expression vector pNZ7021 (MoBiTec). This formed a plasmid of about 7.5kb in size and was size-selected via electrophoresis. Plasmid was extracted and purified with the QIAquick Gel Extraction Kit (Qiagen).

This new construct was electroporated into *E. coli* DH5α for propagation and amplification before attempting to electroporate into *L. lactis* NZ1330. Electroporation was conducted in a Gene Pulser Xcell System (Bio-Rad) at varying ratios of *L. lactis* NZ1330 cells to plasmid DNA suspended in electroporation buffer (0.5 M sucrose + 10% glycerol) with the settings of 2000 V, 25 µF, and 200 Ω. Transformants were selected on 10µg/mL or 100µg/mL chloramphenicol BHI plates incubated at 37˚C for 72 hours. Empty pNZ7021 vehicle was used as a positive control.

REFERENCE LIST

1. Lahiri R, Adams LB. Cultivation and Viability Determination of *Mycobacterium Leprae*. International Textbook of Leprosy, American Leprosy Missions, 18 Sept. 2016. doi:10.1489/itl.5.3.
2. Zha Y, Chong H, Yang P, Ning K. Microbial Dark Matter: from Discovery to Applications. Genomics, Proteomics & Bioinformatics. 2022 Oct;20(5):867-81.
3. Staley JT, Konopka A. Measurement of in situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. Annu Rev Microbiol. 1985 Oct;39(1):321-46.
4. Reguera G. "The Great Plate Count Anomaly" That Is No More. In the Company of Microbes: Ten Years of Small Things Considered. 2016 Apr. Ch. 71
5. Bodor A, Bounedjoum N, Vincze GE, Erdeiné Kis Á, Laczi K, Bende G, et al. Challenges Of Unculturable Bacteria: Environmental Perspectives. Rev Environ Sci Biotechnol. 2020 Mar;19(1):1-22.
6. Bonnet M, Lagier J, Raoult D, Khelaifia S. Bacterial Culture through Selective and Non-Selective Conditions: The Evolution of Culture Media in Clinical Microbiology. New Microbes and New Infections. 2020 Mar;34:100622.
7. Lagier J, Hugon P, Khelaifia S, Fournier P, La Scola B, Raoult D. The Rebirth of Culture in Microbiology through the Example of Culturomics to Study Human Gut Microbiota. Clin Microbiol Rev. 2015 Jan;28(1):237-64.
8. Bondeson J. Spontaneous Generation. In: The Feejee Mermaid and Other Essays in Natural and Unnatural History. Ithaca, NY: Cornell University Press; 2014. p.193-249.
9. Strick J. Darwinism and the Origin of Life: The Role of H.C. Bastian in the British Spontaneous Generation Debates, 1868--1873. Journal of the History of Biology. 1999;32(1):51-92.
10. Slack HJ. Bastian and Pasteur on Spontaneous Generation. The Monthly Microscopical Journal. 1876 Oct;16(4):165-8.
11. Tyndall J. Essays on the Floating-Matter of the Air in Relation to Putrefaction and Infection. Nature. 1881 Nov;25(627):6.
12. Asscher A, Sussman M, Waters W, Davis R, Chick S. Urine as a Medium for Bacterial Growth. The Lancet. 1966 Nov;288(7472):1037-41.

13. Roberts W. On the Occurrence of Micro-Organisms in Fresh Urine. BMJ. 1881 Oct 15;2(1085):623-5.
14. Thomas-White K, Brady M, Wolfe AJ, Mueller ER. The Bladder Is Not Sterile: History and Current Discoveries on the Urinary Microbiome. Curr Bladder Dysfunct Rep. 2016 Mar;11(1):18-24.
15. Shama G. The Role of the Media in Influencing Public Attitudes to Penicillin during World War II. Dynamis. 2015;35(1):131-52.
16. Kass EH. Asymptomatic Infections of the Urinary Tract. Trans Assoc Am Physicians. 1956;69:56-64.
17. Roberts JA. Urinary Tract Infections. American Journal of Kidney Diseases. 1984 Sep;4(2):103-17.
18. Nicolle L. The Paradigm Shift to Non-Treatment of Asymptomatic Bacteriuria. Pathogens. 2016 Apr 19;5(2):38.
19. Maskell R, Allen J, Pead L. The Puzzle of "Urethral Syndrome": A Possible Answer?. The Lancet. 1979 May;313(8125):1058-9.
20. Maskell R. The Natural History of Urinary Tract Infection in Women. Medical Hypotheses. 2010 May;74(5):802-6.
21. Drabu Y, Sanderson P. Urine Culture in Urethral Syndrome. The Lancet. 1980 Jan;315(8158):37-8.
22. Seal D, Cuthbert E. Doubtful Significance of Fastidious Bacteriuria in the Urethral Syndrome. The Lancet. 1982 Jan;319(8263):115.
23. Woese CR. Bacterial Evolution. Microbiol Rev. 1987 Jun;51(2):221-71.
24. Woese C, Stackebrandt E, Macke T, Fox G. A Phylogenetic Definition of the Major Eubacterial Taxa. Systematic and Applied Microbiology. 1985 Sep;6(2):143-51.
25. Nam N, Do H, Loan Trinh K, Lee N. Metagenomics: An Effective Approach for Exploring Microbial Diversity and Functions. Foods. 2023 May 25;12(11):2140.
26. Hoffman C, Siddiqui NY, Fields I, Gregory WT, Simon HM, Mooney MA, et al. Species-Level Resolution of Female Bladder Microbiota from 16S rRNA Amplicon Sequencing. mSystems. 2021 Oct 26;6(5)
27. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun Metagenomics, from Sampling to Analysis. Nat Biotechnol. 2017 Sep;35(9):833-44.
28. Price TK, Wolff B, Halverson T, Limeira R, Brubaker L, Dong Q, et al. Temporal Dynamics of the Adult Female Lower Urinary Tract Microbiota. mBio. 2020 Apr 28;11(2)
29. Onywera H, Chambuso R, Benjamin DJ, Hilt EE, Thomas-White K. Editorial: An Outlook on Urobiome: Advances in Understanding the Role of Urobiome in Urological Health and Disease and its Potential in Biotherapeutics. Front Urol. 2024 Feb 23;4

30. Franco-Duarte R, Černáková L, Kadam S, S. Kaushik K, Salehi B, Bevilacqua A, et al. Advances in Chemical and Biological Methods to Identify Microorganisms—From Past to Present. Microorganisms. 2019 May 13;7(5):130.

31. Dixon B. What's in a Name?. The Lancet Infectious Diseases. 2001 Oct;1(3):212.

32. Lan R, Reeves PR. When Does a Clone Deserve a Name? A Perspective on Bacterial Species Based on Population Genetics. Trends in Microbiology. 2001 Sep;9(9):419-24.

33. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kämpfer P, Maiden MCJ, et al. Report of the ad hoc Committee for the Re-Evaluation of the Species Definition in Bacteriology. International Journal of Systematic and Evolutionary Microbiology. 2002 May 1;52(3):1043-7.

34. Ward BB. How Many Species of Prokaryotes Are There?. Proc Natl Acad Sci USA. 2002 Aug 6;99(16):10234-6.

35. Wayne LG, Good RC, Bottger EC, Butler R, Dorsch M, Ezaki T, et al. Semantide- and Chemotaxonomy-Based Analyses of Some Problematic Phenotypic Clusters of Slowly Growing Mycobacteria, a Cooperative Study of the International Working Group on Mycobacterial Taxonomy. International Journal of Systematic Bacteriology. 1996 Jan 1;46(1):280-97.

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

37. Richter M, Rosselló-Móra R. Shifting the Genomic Gold Standard for the Prokaryotic Species Definition. Proc Natl Acad Sci USA. 2009 Nov 10;106(45):19126-31.

38. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. J Clin Microbiol. 2012 Apr;50(4):1355-61.

39. Cho J, Tiedje JM. Bacterial Species Determination from DNA-DNA Hybridization by Using Genome Fragments and DNA Microarrays. Appl Environ Microbiol. 2001 Aug;67(8):3677-82.

40. Fox GE, Wisotzkey JD, Jurtshuk P. How Close Is Close: 16S rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. International Journal of Systematic Bacteriology. 1992 Jan 1;42(1):166-70.

41. Palmer M, Steenkamp ET, Blom J, Hedlund BP, Venter SN. All ANIs are Not Created Equal: Implications for Prokaryotic Species Boundaries and Integration of ANIs into Polyphasic Taxonomy. International Journal of Systematic and Evolutionary Microbiology. 2020 Apr 1;70(4):2937-48.

42. Ursell LK, Metcalf JL, Parfrey LW, Knight R. Defining the Human Microbiome. Nutrition Reviews. 2012 Aug;70:S38-S44.

43. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, FitzGerald M, et al. Evidence of Uncultivated Bacteria in the Adult Female Bladder. J Clin Microbiol. 2012 Apr;50(4):1376-83.

44. Tang J. Microbiome in the Urinary System—a Review. AIMS Microbiology. 2017;3(2):143-54

45. Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS. Assessing Diversity of the Female Urine Microbiota by High Throughput Sequencing of 16S rDNA Amplicons. BMC Microbiol. 2011 Dec;11(1)

46. Thapaliya J, Khadka P, Thapa S, Gongal C. Enhanced Quantitative Urine Culture Technique, a Slight Modification, in Detecting Under-Diagnosed Pediatric Urinary Tract Infection. BMC Res Notes. 2020 Jan 3;13(5)

47. Price T, Hilt E, Thomas-White K, Mueller E, Wolfe A, Brubaker L. The Urobiome of Continent Adult Women: a Cross‐Sectional Study. BJOG. 2020 Jan;127(2):193-201.

48. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, et al. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. J Clin Microbiol. 2014 Mar;52(3):871-6.

49. Barnes HC, Wolff B, Abdul-Rahim O, Harrington A, Hilt EE, Price TK, et al. A Randomized Clinical Trial of Standard versus Expanded Cultures to Diagnose Urinary Tract Infections in Women. Journal of Urology. 2021 Nov;206(5):1212-21.

50. Deen NS, Ahmed A, Tasnim NT, Khan N. Clinical Relevance of Expanded Quantitative Urine Culture in Health and Disease. Front Cell Infect Microbiol. 2023 Aug 1;13

51. Zagaglia C, Ammendolia MG, Maurizi L, Nicoletti M, Longhi C. Urinary Tract Infections Caused by Uropathogenic Escherichia coli Strains—New Strategies for an Old Pathogen. Microorganisms. 2022 Jul 14;10(7):1425.

52. Price TK, Hilt EE, Dune TJ, Mueller ER, Wolfe AJ, Brubaker L. Urine Trouble: Should We Think Differently about UTI?. Int Urogynecol J. 2018 Feb;29(2):205-10.

53. Garretto A, Miller-Ensminger T, Ene A, Merchant Z, Shah A, Gerodias A, et al. Genomic Survey of E. coli From the Bladders of Women With and Without Lower Urinary Tract Symptoms. Front Microbiol. 2020 Sep 4;11

54. Moustafa A, Li W, Singh H, Moncera KJ, Torralba MG, Yu Y, et al. Microbial Metagenome of Urinary Tract Infection. Sci Rep. 2018 Mar 12;8(1)

55. Zandbergen LE, Halverson T, Brons JK, Wolfe AJ, de Vos MGJ. The Good and the Bad: Ecological Interaction Measurements Between the Urinary Microbiota and Uropathogens. Front Microbiol. 2021 May 10;12

56. Kim DS, Lee JW. Urinary Tract Infection and Microbiome. Diagnostics. 2023 May 31;13(11):1921.

57. Joyce C, Halverson T, Gonzalez C, Brubaker L, Wolfe AJ. The Urobiomes of Adult Women With Various Lower Urinary Tract Symptoms Status Differ: A Re-Analysis. Front Cell Infect Microbiol. 2022 Jun 9;12

58. Frølund M, Wikström A, Lidbrink P, Abu Al-Soud W, Larsen N, Harder CB, et al. The Bacterial Microbiota in First-Void Urine from Men with and without Idiopathic Urethritis. PLoS ONE. 2018 Jul 27;13(7):e0201380.

59. Heidar NA, Bhat TA, Shabir U, Hussein AA. The Urinary Microbiome and Bladder Cancer. Life. 2023 Mar 17;13(3):812.

60. Du J, Khemmani M, Halverson T, Ene A, Limeira R, Tinawi L, Hochstedler-Kramer BR, Noronha MF, Putonti C, Wolfe AJ. Cataloging the Phylogenetic Diversity of Human Bladder Bacterial Isolates. bioRxiv [Preprint]. 2023 May 23:2023.05.23.541916. doi: 10.1101/2023.05.23.541916.

61. Belizário JE, Faintuch J. Microbiome and Gut Dysbiosis. Exp Suppl. 2018;109:459-476. doi: 10.1007/978-3-319-74932-7_13.

62. Baquero F, Nombela C. The Microbiome as a Human Organ. Clinical Microbiology and Infection. 2012 Jul;18:2-4.

63. Ogunrinola GA, Oyewale JO, Oshamika OO, Olasehinde GI. The Human Microbiome and Its Impacts on Health. International Journal of Microbiology. 2020 Jun 12;2020:1-7.

64. Wu Y, Wang Y, Yang H, Li Q, Gong X, Zhang G, et al. Resident Bacteria Contribute to Opportunistic Infections of the Respiratory Tract. PLoS Pathog. 2021 Mar 19;17(3):e1009436.

65. Dey P, Ray Chaudhuri S. The Opportunistic Nature of Gut Commensal Microbiota. Critical Reviews in Microbiology. 2023 Nov 2;49(6):739-63.

66. Gottschick C, Deng Z, Vital M, Masur C, Abels C, Pieper DH, et al. The Urinary Microbiota of Men and Women and its Changes in Women During Bacterial Vaginosis and Antibiotic Treatment. Microbiome. 2017 Dec;5(1)

67. Ammitzbøll N, Bau BPJ, Bundgaard-Nielsen C, Villadsen AB, Jensen A, Leutscher PDC, et al. Pre- and Postmenopausal Women have Different Core Urinary Microbiota. Sci Rep. 2021 Jan 26;11(1)

68. Haley E, Luke N, Mathur M, Festa R, Wang J, Jiang Y, et al. The Prevalence and Association of Different Uropathogens Detected by M-PCR with Infection-Associated Urine Biomarkers in Urinary Tract Infections. RRU. 2024 Jan;Volume 16:19-29.

69. Kline KA, Lewis AL. Gram-Positive Uropathogens, Polymicrobial Urinary Tract Infection, and the Emerging Microbiota of the Urinary Tract. Microbiol Spectr. 2016 Mar 25;4(2)

70. Zhang Q, Kwoh C, Attorri S, Clarridge JE. *Aerococcus urinae* in Urinary Tract Infections. J Clin Microbiol. 2000 Apr;38(4):1703-5.

71. Higgins A, Garg T. *Aerococcus urinae*: An Emerging Cause of Urinary Tract Infection in Older Adults with Multimorbidity and Urologic Cancer. Urology Case Reports. 2017 Jul;13:24-5.

72. Moreland RB, Choi BI, Geaman W, Gonzalez C, Hochstedler-Kramer BR, John J, et al. Beyond the Usual Suspects: Emerging Uropathogens in the Microbiome Age. Front Urol. 2023 Jul 26;3

73. Aguirre M, Collins MD. Phylogenetic analysis of some *Aerococcus*-like organisms from Urinary Tract Infections: Description of *Aerococcus urinae* sp. nov. J Gen Microbiol. 1992 Feb;138(2):401-5.

74. Greco M, Rúa-Figueroa Í, Ghiglione S, Santana Medina E, Quevedo Abeledo JC, Rodríguez-Lozano C. Musculoskeletal Infections Caused by *Aerococcus urinae*: a Case-Based Review. Clin Rheumatol. 2018 Sep;37(9):2587-94.

75. Narayanasamy S, King K, Dennison A, Spelman DW, Aung AK. Clinical Characteristics and Laboratory Identification of *Aerococcus* Infections: An Australian Tertiary Centre Perspective. Int J Microbiol. 2017;2017:5684614.

76. Senneby E, Petersson AC, Rasmussen M. Epidemiology and Antibiotic Susceptibility of Aerococci in Urinary Cultures. Diagn Microbiol Infect Dis. 2015 Feb;81(2):149-51.

77. Sihvonen R, Turunen M, Lehtola L, Pakarinen L, Grönroos JO, Rantakokko-Jalava K, et al. Clinical and Microbiological Characterization of *Aerococcus urinae* Bacteraemias at Helsinki Metropolitan Area, Finland. Eur J Clin Microbiol Infect Dis. 2022 May;41(5):751-60.

78. Lyagoubi A, Souffi C, Baroiller V, Vallee E. *Aerococcus Urinae* Spondylodiscitis: An Increasingly Described Localization. EJIFCC. 2020 Jun;31(2):169-73.

79. Forsvall A, Wagenius M, Rasmussen M. Perigenital necrotizing soft tissue infection caused by *Aerococcus urinae*. IDCases. 2019;18:e00590.

80. Tai DBG, Go JR, Fida M, Saleh OA. Management and Treatment of *Aerococcus* Bacteremia and Endocarditis. Int J Infect Dis. 2021 Jan;102:584-9.

81. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, et al. The Female Urinary Microbiome: a Comparison of Women with and without Urgency Urinary Incontinence. mBio. 2014 Jul 8;5(4):e01283-14.

82. Joyce C, Halverson T, Gonzalez C, Brubaker L, Wolfe AJ. The Urobiomes of Adult Women With Various Lower Urinary Tract Symptoms Status Differ: A Re-Analysis. Front Cell Infect Microbiol. 2022;12:860408.

83. Rasmussen M. *Aerococcus*: an Increasingly Acknowledged Human Pathogen. Clin Microbiol Infect. 2016 Jan;22(1):22-7.

84. Sahu KK, Lal A, Mishra AK, Abraham GM. *Aerococcus*-Related Infections and their Significance: A 9-Year Retrospective Study. J Microsc Ultrastruct. 2021;9(1):18-25.

85. Senneby E, Göransson L, Weiber S, Rasmussen M. A Population-Based Study of Aerococcal Bacteraemia in the MALDI-TOF MS-era. Eur J Clin Microbiol Infect Dis. 2016 May;35(5):755-62.

86. Sunnerhagen T, Nilson B, Olaison L, Rasmussen M. Clinical and Microbiological Features of Infective Endocarditis Caused by Aerococci. Infection. 2016 Apr;44(2):167-73.

87. Rast D, Evers KS, Egli A, Rudin C, Goischke A, Ritz N. *Aerococcus urinae* - Significance of Detection in the Paediatric Urinary Tract: a case series. Eur J Pediatr. 2023 Feb;182(2):749-56.

88. Lenherr N, Berndt A, Ritz N, Rudin C. *Aerococcus urinae*: a Possible Reason for Malodorous Urine in Otherwise Healthy Children. Eur J Pediatr. 2014 Aug;173(8):1115-7.

89. Geeraedts F, Stoffers C, Smidt H, Schijffelen M. Foul Smelling Urine in an Adult Caused by *Aerococcus urinae*. IDCases. 2023;31:e01657

90. Akinboboye O, Babalola TO, Onoriode K, Appiah-Pippim J. Fatal *Aerococcus urinae* Aortic Valve Endocarditis with Severe Regurgitation. Am J Case Rep. 2022 Oct 17;23:e937596.

91. Higgins A, Garg T. *Aerococcus urinae*: An Emerging Cause of Urinary Tract Infection in Older Adults with Multimorbidity and Urologic Cancer. Urol Case Rep. 2017 Jul;13:24-5.

92. Senneby E, Nilson B, Petersson A, Rasmussen M. Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry Is a Sensitive and Specific Method for Identification of Aerococci. J Clin Microbiol. 2013 Apr;51(4):1303-4.

93. Christensen JJ, Dargis R, Hammer M, Justesen US, Nielsen XC, Kemp M. Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry Analysis of Gram-Positive, Catalase-Negative Cocci Not Belonging to the *Streptococcus* or Enterococcus Genus and Benefits of Database Extension. J Clin Microbiol. 2012 May;50(5):1787-91.

94. Rasmussen M. Aerococci and Aerococcal Infections. Journal of Infection. 2013 Jun;66(6):467-74.

95. Facklam R, Lovgren M, Shewmaker PL, Tyrrell G. Phenotypic Description and Antimicrobial Susceptibilities of *Aerococcus sanguinicola* Isolates from Human Clinical Samples. J Clin Microbiol. 2003 Jun;41(6):2587-92.

96. Christensen JJ, Whitney AM, Teixeira LM, Steigerwalt AG, Facklam RR, Korner B, et al. *Aerococcus urinae*: Intraspecies Genetic and Phenotypic Relatedness. Int J Syst Bacteriol. 1997 Jan;47(1):28-32.

97. Carkaci D, Højholt K, Nielsen XC, Dargis R, Rasmussen S, Skovgaard O, et al. Genomic Characterization, Phylogenetic Analysis, and Identification of Virulence Factors in *Aerococcus sanguinicola* and *Aerococcus urinae* Strains Isolated From Infection Episodes. Microb Pathog. 2017 Nov;112:327-40.

98. Hilt EE, Putonti C, Thomas-White K, Lewis AL, Visick KL, Gilbert NM, et al. *Aerococcus urinae* Isolated from Women with Lower Urinary Tract Symptoms: In Vitro Aggregation and Genome Analysis. J Bacteriol. 2020 Jun 9;202(13):e00170-20.

99. Yu Y, Tsitrin T, Bekele S, Thovarai V, Torralba MG, Singh H, et al. *Aerococcus urinae* and *Globicatella sanguinis* Persist in Polymicrobial Urethral Catheter Biofilms Examined in Longitudinal Profiles at the Proteomic Level. Biochem Insights. 2019;12:1178626419875089.

100. Yaban B, Kikhney J, Musci M, Petrich A, Schmidt J, Hajduczenia M, et al. *Aerococcus urinae* - A Potent Biofilm Builder In Endocarditis. PLoS One. 2020;15(4):e0231827.

101. Shannon O, Mörgelin M, Rasmussen M. Platelet Activation and Biofilm Formation by *Aerococcus urinae*, an Endocarditis-Causing Pathogen. Infect Immun. 2010 Oct;78(10):4268-75.

102. Senneby E, Sunnerhagen T, Hallström B, Lood R, Malmström J, Karlsson C, et al. Identification of Two Abundant *Aerococcus urinae* Cell Wall-Anchored Proteins. Int J Med Microbiol. 2019 Nov;309(7):151325.

103. Skov R, Christensen JJ, Korner B, Frimodt-Møller N, Espersen F. In vitro Antimicrobial Susceptibility of *Aerococcus urinae* to 14 Antibiotics, and Time-Kill Curves for Penicillin, Gentamicin and Vancomycin. J Antimicrob Chemother. 2001 Nov;48(5):653-8.

104. Humphries RM, Hindler JA. In vitro Antimicrobial Susceptibility of *Aerococcus urinae*. J Clin Microbiol. 2014 Jun;52(6):2177-80.

105. Scholtz SL, Faron ML, Buchan BW, Ledeboer NA. Comparison of Methods for Determining the Antibiotic Susceptibility of *Aerococcus* Species in a Clinical Setting. Am J Clin Pathol. 2022 May 4;157(5):781-8.

106. Lupo A, Guilarte YN, Droz S, Hirzel C, Furrer H, Endimiani A. In vitro Activity of Clinically Implemented B-Lactams against *Aerococcus urinae*: Presence of Non-Susceptible Isolates in Switzerland. New Microbiol. 2014 Oct;37(4):563-6.

107. Hirzel C, Guilarte YN, Hirzberger L, Furrer H, Marschall J, Endimiani A. In vitro Susceptibility of *Aerococcus urinae* Isolates to Antibiotics Used for Uncomplicated Urinary Tract Infection. J Infect. 2015 Sep;71(3):395-7.

108. Siew SW, Musa SM, Sabri N', Farida Asras MF, Ahmad HF. Evaluation of Pre-Treated Healthcare Wastes during COVID-19 Pandemic Reveals Pathogenic Microbiota, Antibiotics Residues, and Antibiotic Resistance Genes against Beta-Lactams. Environ Res. 2023 Feb 15;219:115139.

109. Yang Q, Zhu Y, Schwarz S, Wang L, Liu W, Yang W, et al. A Novel Plasmid from *Aerococcus urinaeequi* of Porcine Origin Co-Harboring the Tetracycline Resistance Genes tet(58) and tet(61). Vet Microbiol. 2021 Jun;257:109065.

110. Zhou W, Gao S, Zheng J, Zhang Y, Zhou H, Zhang Z, et al. Identification of an *Aerococcus urinaeequi* Isolate by Whole Genome Sequencing and Average Nucleotide Identity Analysis. J Glob Antimicrob Resist. 2022 Jun;29:353-9.

111. Christensen JJ, Korner B, Kjærgaard H. *Aerococcus*-like Organism – an Unnoticed Urinary Tract Pathogen. APMIS. 1989 Jan;97(1-6):539-46.

112. Christensen JJ, Vibits H, Ursing J, Korner B. *Aerococcus*-like Organism, a Newly Recognized Potential Urinary Tract Pathogen. J Clin Microbiol. 1991 May;29(5):1049-53.

113. Christensen JJ, Kilian M, Fussing V, Anderson K, Blom J, Korner B, et al. *Aerococcus urinae*: Polyphasic Characterization of the Species. APMIS. 2005 Aug;113(7-8):517-25.

114. Felis GE. Reclassification of *Pediococcus urinaeequi* (ex Mees 1934) Garvie 1988 as *Aerococcus urinaeequi* comb. nov.. International Journal Of Systematic And Evolutionary Microbiology. 2005 May 1;55(3):1325-7.

115. Carkaci D, Dargis R, Nielsen XC, Skovgaard O, Fuursted K, Christensen JJ. Complete Genome Sequences of *Aerococcus christensenii* CCUG 28831 T, *Aerococcus sanguinicola* CCUG 43001 T, *Aerococcus urinae* CCUG 36881 T, *Aerococcus urinaeequi* CCUG 28094 T, *Aerococcus urinaehominis* CCUG 42038 B T, and *Aerococcus viridans* CCUG 4311 T. Genome Announc. 2016 Apr 28;4(2)

116. Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, et al. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. J Clin Microbiol. 2016 May;54(5):1216-22.

117. Gilbert NM, Choi B, Du J, Collins C, Lewis AL, Putonti C, et al. A Mouse Model Displays Host and Bacterial Strain Differences in *Aerococcus urinae* Urinary Tract Infection. Biology Open. 2021 Aug 15;10(8)

118. Pearce MM, Zilliox MJ, Rosenfeld AB, Thomas-White KJ, Richter HE, Nager CW, et al. The Female Urinary Microbiome in Urgency Urinary Incontinence. American Journal of Obstetrics and Gynecology. 2015 Sep;213(3):347.e1-347.e11.

119. Hrncir T. Gut Microbiota Dysbiosis: Triggers, Consequences, Diagnostic and Therapeutic Options. Microorganisms. 2022 Mar 7;10(3):578.

120. Coburn B, Grassl GA, Finlay BB. Salmonella, the Host and Disease: a Brief Review. Immunol Cell Biol. 2007 Feb;85(2):112-8.

121. Hilt EE. Aerococcus Urinae: Establishing the Pathogenesis of an Emerging Uropathogen. Dissertations. 2019 Aug. 3339. https://ecommons.luc.edu/luc_diss/3339

122. Vestby LK, Grønseth T, Simm R, Nesse LL. Bacterial Biofilm and its Role in the Pathogenesis of Disease. Antibiotics. 2020 Feb 3;9(2):59.

123. Choi BI, Ene A, Du J, Johnson G, Putonti C, Schouw CH, et al. Taxonomic Considerations on *Aerococcus urinae* with Proposal of Subdivision into *Aerococcus urinae*, *Aerococcus tenax* sp. nov., *Aerococcus mictus* sp. nov., and *Aerococcus loyolae* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2023 Sep 27;73(9)

124. Tohno M, Kitahara M, Matsuyama S, Kimura K, Ohkuma M, Tajima K. *Aerococcus vaginalis* sp. nov., Isolated from the Vaginal Mucosa of a Beef Cow, and Emended Descriptions of *Aerococcus suis*, *Aerococcus viridans*, *Aerococcus urinaeequi*, *Aerococcus urinaehominis*, *Aerococcus urinae*, *Aerococcus christensenii* and *Aerococcus sanguinicola*. International Journal of Systematic and Evolutionary Microbiology. 2014 Apr 1;64(Pt_4):1229-36.

125. Thurlow LR, Thomas VC, Hancock LE. Capsular Polysaccharide Production in *Enterococcus faecalis* and Contribution of CpsF to Capsule Serospecificity. J Bacteriol. 2009 Oct 15;191(20):6203-10.

126. Sterling AJ, Snelling WJ, Naughton PJ, Ternan NG, Dooley JSG. Competent but Complex Communication: The Phenomena of Pheromone-Responsive Plasmids. PLoS Pathog. 2020 Apr 2;16(4):e1008310.

127. Mansfield JM, Herrmann P, Jesionowski AM, Vickerman MM. *Streptococcus gordonii* Pheromone s.g.cAM373 May Influence the Reservoir of Antibiotic Resistance Determinants of *Enterococcus faecalis* Origin in the Oral Metagenome. Journal of Medical Microbiology. 2017 Nov 1;66(11):1635-9.

128. Dunny GM, Berntsson RP. Enterococcal Sex Pheromones: Evolutionary Pathways to Complex, Two-Signal Systems. J Bacteriol. 2016 Jun;198(11):1556-62.

129. Shanker E, Federle M. Quorum Sensing Regulation of Competence and Bacteriocins in *Streptococcus pneumoniae* and *mutans*. Genes. 2017 Jan 5;8(1):15.

130. Gardan R, Besset C, Guillot A, Gitton C, Monnet V. The Oligopeptide Transport System Is Essential for the Development of Natural Competence in *Streptococcus thermophilus* Strain LMD-9. J Bacteriol. 2009 Jul 15;191(14):4647-55.

131. Kurushima J, Campo N, van Raaphorst R, Cerckel G, Polard P, Veening J. Unbiased Homeologous Recombination during Pneumococcal Transformation Allows for Multiple Chromosomal Integration Events. eLife. 2020 Sep 23;9

132. Piotrowski A, Luo P, Morrison DA. Competence for Genetic Transformation in *Streptococcus pneumoniae* : Termination of Activity of the Alternative Sigma Factor ComX Is Independent of Proteolysis of ComX and ComW. J Bacteriol. 2009 May 15;191(10):3359-66.

133. Brophy JAN, Triassi AJ, Adams BL, Renberg RL, Stratis-Cullum DN, Grossman AD, et al. Engineered Integrative and Conjugative Elements for Efficient and Inducible DNA Transfer to Undomesticated Bacteria. Nat Microbiol. 2018 Aug 20;3(9):1043-53.

134. Waldron DE, Lindsay JA. Sau1: a Novel Lineage-Specific Type I Restriction-Modification System That Blocks Horizontal Gene Transfer into *Staphylococcus aureus* and between *S. aureus* Isolates of Different Lineages. J Bacteriol. 2006 Aug;188(15):5578-85.

135. Johnson CM, Grossman AD. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. Annu Rev Genet. 2015 Nov 23;49(1):577-601.

136. Varaldo PE, Montanari MP, Giovanetti E. Genetic Elements Responsible for Erythromycin Resistance in Streptococci. Antimicrob Agents Chemother. 2009 Feb;53(2):343-53.

137. Franke AE, Clewell DB. Evidence for a Chromosome-Borne Resistance Transposon (Tn916) in *Streptococcus faecalis* that is Capable of "Conjugal" Transfer in the Absence of a Conjugative Plasmid. Journal of Bacteriology. 1981;145(1):494-502.

138. Lamichhane J, Choi BI, Stegman N, Fontes Noronha M, Wolfe AJ. Macrolide Resistance in the *Aerococcus urinae* Complex: Implications for Integrative and Conjugative Elements. Antibiotics. 2024; 13(5):433.

139. Khan UB, Portal EAR, Sands K, et al. Genomic Analysis Reveals New Integrative Conjugal Elements and Transposons in GBS Conferring Antimicrobial Resistance. Antibiotics. 2023;12(3):544.

140. Rosca AS, Castro J, Cerca N. Evaluation of Different Culture Media to Support in vitro Growth and Biofilm Formation of Bacterial Vaginosis-Associated Anaerobes. PeerJ. 2020 Sep 10;8:e9917.

141. Bhat AH, Nguyen MT, Das A, Ton-That H. Anchoring Surface Proteins to the Bacterial Cell Wall by Sortase Enzymes: How It Started and What We Know Now. Current Opinion in Microbiology. 2021 Apr;60:73-9.

142. Bandara M, Skehel JM, Kadioglu A, Collinson I, Nobbs AH, Blocker AJ, et al. The Accessory Sec System (SecY2A2) in *Streptococcus pneumoniae* is Involved in Export of Pneumolysin Toxin, Adhesion and Biofilm Formation. Microbes and Infection. 2017 Jul;19(7-8):402-12.

143. Chan JM, Gori A, Nobbs AH, Heyderman RS. Streptococcal Serine-Rich Repeat Proteins in Colonization and Disease. Front Microbiol. 2020 Oct 30;11

144. Whelan F, Lafita A, Griffiths SC, Cooper REM, Whittingham JL, Turkenburg JP, et al. Defining the Remarkable Structural Malleability of a Bacterial Surface Protein Rib Domain Implicated in Infection. Proc Natl Acad Sci USA. 2019 Dec 26;116(52):26540-8.

145. Veljović K, Popović N, Miljković M, Tolinački M, Terzić-Vidojević A, Kojić M. Novel Aggregation Promoting Factor AggE Contributes to the Probiotic Properties of *Enterococcus faecium* BGGO9-28. Front Microbiol. 2017 Sep 26;8

146. Lukić J, Strahinić I, Jovčić B, Filipić B, Topisirović L, Kojić M, et al. Different Roles for Lactococcal Aggregation Factor and Mucin Binding Protein in Adhesion to Gastrointestinal Mucosa. Appl Environ Microbiol. 2012 Nov 15;78(22):7993-8000.

147. van Asseldonk M, de Vos WM, Simons G. Functional Analysis of the *Lactococcus lactis* usp45 Secretion Signal in the Secretion of a Homologous Proteinase and a Heterologous α-amylase. Molec Gen Genet. 1993 Sep;240(3):428-34.

148. Schuur PMH, Kasteren MEEv, Sabbe L, Vos MC, Janssens MMPC, Buiting AGM. Urinary tract infections with *Aerococcus urinae* in the South of the Netherlands. Eur J Clin Microbiol Infect Dis. 1997 Dec;16(12):871-5.

149. Chen YB, Hochstedler B, Pham TT, Alvarez MA, Mueller ER, Wolfe AJ. The Urethral Microbiota: A Missing Link in the Female Urinary Microbiota. Journal of Urology. 2020 Aug;204(2):303-9.

150. Azevedo AS, Almeida C, Melo LF, Azevedo NF. Impact of Polymicrobial Biofilms in Catheter-Associated Urinary Tract Infections. Critical Reviews in Microbiology. 2017 Jul 4;43(4):423-39.

151. Daubin V, Szöllősi GJ. Horizontal Gene Transfer and the History of Life. Cold Spring Harb Perspect Biol. 2016 Apr;8(4):a018036.

152. Koeppel AF, Wertheim JO, Barone L, Gentile N, Krizanc D, Cohan FM. Speedy Speciation in a Bacterial Microcosm: New Species can Arise as Frequently as Adaptations within a Species. The ISME Journal. 2013 Jun 1;7(6):1080-91.

153. Suzuki TA, Martins FM, Nachman MW. Altitudinal Variation of the Gut Microbiota in Wild House Mice. Molecular Ecology. 2019 May;28(9):2378-90.

154. Goertz S, de Menezes AB, Birtles RJ, Fenn J, Lowe AE, MacColl ADC, et al. Geographical Location Influences the Composition of the Gut Microbiota in Wild House Mice (Mus musculus domesticus) at a fine spatial scale. PLoS ONE. 2019 Sep 26;14(9):e0222501.

155. Sudakaran S, Salem H, Kost C, Kaltenpoth M. Geographical and Ecological Stability of the Symbiotic Mid‐Gut Microbiota in European Firebugs, *Pyrrhocoris apterus* (Hemiptera, Pyrrhocoridae). Molecular Ecology. 2012 Dec;21(24):6134-51.

156. Hasan N, Yang H. Factors Affecting the Composition of the Gut Microbiota, and Its Modulation. PeerJ. 2019 Aug 16;7:e7502.

157. Ahn J, Hayes RB. Environmental Influences on the Human Microbiome and Implications for Noncommunicable Disease. Annu Rev Public Health. 2021 Apr 1;42(1):277-92.

158. Fine J. Glucose Content of Normal Urine. BMJ. 1965 May 8;1(5444):1209-14.

159. Gupta RC, Goyal A, Ghosh R, Punjabi M, Singh PP. Normal Range for Glucose in Urine: Age-Related Changes. Clin Chem. 1982 Nov;28(11):2335.

160. Mann R, Mediati DG, Duggin IG, Harry EJ, Bottomley AL. Metabolic Adaptations of Uropathogenic *E. coli* in the Urinary Tract. Front Cell Infect Microbiol. 2017 Jun 8;7

161. Ala-Jaakkola R, Laitila A, Ouwehand AC, Lehtoranta L. Role of D-mannose in Urinary Tract Infections – a Narrative Review. Nutr J. 2022 Mar 22;21(1)

162. Neugent ML, Hulyalkar NV, Kumar A, Xing C, Zimmern PE, Shulaev V, et al. Urinary Glycosaminoglycans Are Associated with Recurrent UTI and Urobiome Ecology in Postmenopausal Women. ACS Infect Dis. 2023 Apr 14;9(4):1022-32.

163. Nguyen VH, Khan F, Shipman BM, Neugent ML, Hulyalkar NV, Cha NY, et al. A Semi-Quantitative Assay to Measure Glycosaminoglycan Degradation by the Urinary Microbiota. Front Cell Infect Microbiol. 2022 Jan 3;11

164. Fukuwatari T, Yoshida E, Takahashi K, Shibata K. Effect of Fasting on the Urinary Excretion of Water-Soluble Vitamins in Humans and Rats. J Nutr Sci Vitaminol. 2010;56(1):19-26.

165. Tsuji T, Fukuwatari T, Sasaki S, Shibata K. Urinary Excretion of Vitamin B1, B2, B6, Niacin, Pantothenic Acid, Folate, and Vitamin C Correlates with Dietary Intakes of Free-Living Elderly, Female Japanese. Nutrition Research. 2010 Mar;30(3):171-8.

166. Arseculeratne SN. Surface Properties of *Aerococcus viridans* in Relation to Virulence. Microbial Ecology in Health and Disease. 1989 Jan;2(2):91-7.

167. Cvitkovitch DG, Li Y, Ellen RP. Quorum sensing and biofilm formation in Streptococcal infections. J Clin Invest. 2003 Dec 1;112(11):1626-32.

168. Șchiopu P, Toc DA, Colosi IA, Costache C, Ruospo G, Berar G, et al. An Overview of the Factors Involved in Biofilm Production by the *Enterococcus* Genus. IJMS. 2023 Jul 18;24(14):11577.

169. Goessweiner-Mohr N, Arends K, Keller W, Grohmann E. Conjugation in Gram-Positive Bacteria. Microbiol Spectr. 2014;2(4)

170. Veljović K, Popović N, Miljković M, Tolinački M, Terzić-Vidojević A, Kojić M. Novel Aggregation Promoting Factor AggE Contributes to the Probiotic Properties of *Enterococcus faecium* BGGO9-28. Front Microbiol. 2017 Sep 26;8

171. Lukić J, Strahinić I, Jovčić B, Filipić B, Topisirović L, Kojić M, et al. Different Roles for Lactococcal Aggregation Factor and Mucin Binding Protein in Adhesion to Gastrointestinal Mucosa. Appl Environ Microbiol. 2012 Nov 15;78(22):7993-8000.

172. Jankovic I, Ventura M, Meylan V, Rouvet M, Elli M, Zink R. Contribution of Aggregation-Promoting Factor to Maintenance of Cell Shape in *Lactobacillus gasseri* 4B2. J Bacteriol. 2003 Jun;185(11):3288-96.

173. Speziale P, Pietrocola G. The Multivalent Role of Fibronectin-Binding Proteins A and B (FnBPA and FnBPB) of *Staphylococcus aureus* in Host Infections. Front Microbiol. 2020 Aug 26;11

174. Mulder J, Wels M, Kuipers OP, Kleerebezem M, Bron PA. Unleashing Natural Competence in *Lactococcus lactis* by Induction of the Competence Regulator ComX. Appl Environ Microbiol. 2017 Oct 15;83(20)

175. Smith SD, Colgan P, Yang F, Rieke EL, Soupir ML, Moorman TB, et al. Investigating the Dispersal of Antibiotic Resistance Associated Genes from Manure Application to Soil and Drainage Waters in Simulated Agricultural Farmland Systems. PLoS ONE. 2019 Sep 17;14(9):e0222470.

176. Johnson G, Bataclan S, So M, Banerjee S, Wolfe AJ, Putonti C. Plasmids of the Urinary Microbiota. Access Microbiology. 2022 Nov 1;4(11)

177. Thomas-White KJ, Hilt EE, Fok C, Pearce MM, Mueller ER, Kliethermes S, et al. Incontinence Medication Response Relates to the Female Urinary Microbiota. Int Urogynecol J. 2016 May;27(5):723-33.

178. Price T, Hilt E, Thomas-White K, Mueller E, Wolfe A, Brubaker L. The Urobiome of Continent Adult Women: a Cross‐Sectional Study. BJOG. 2020 Jan;127(2):193-201.

179. Jacobs KM, Price TK, Thomas-White K, Halverson T, Davies A, Myers DL, et al. Cultivable Bacteria in Urine of Women With Interstitial Cystitis: (Not) What We Expected. Female Pelvic Med Reconstr Surg. 2021 May;27(5):322-7.

180. Price TK, Lin H, Gao X, Thomas-White KJ, Hilt EE, Mueller ER, et al. Bladder Bacterial Diversity Differs in Continent and Incontinent Women: a Cross-Sectional Study. American Journal of Obstetrics and Gynecology. 2020 Nov;223(5):729.e1-729.e10.

181. Thomas-White K, Taege S, Limeira R, Brincat C, Joyce C, Hilt EE, et al. Vaginal Estrogen Therapy is Associated with Increased *Lactobacillus* in the Urine of Postmenopausal Women with Overactive Bladder Symptoms. American Journal of Obstetrics and Gynecology. 2020 Nov;223(5):727.e1-727.e11.

182. Barnes HC, Wolff B, Abdul-Rahim O, Harrington A, Hilt EE, Price TK, et al. A Randomized Clinical Trial of Standard versus Expanded Cultures to Diagnose Urinary Tract Infections in Women. Journal of Urology. 2021 Nov;206(5):1212-21.

183. Joyce C, Halverson T, Gonzalez C, Brubaker L, Wolfe AJ. The Urobiomes of Adult Women With Various Lower Urinary Tract Symptoms Status Differ: A Re-Analysis. Front Cell Infect Microbiol. 2022 Jun 9;12

184. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013 Apr 1;30(4):772-80.

185. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010 Mar 10;5(3):e9490.

186. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an Online Tool for Phylogenetic Tree Display and Annotation. Nucleic Acids Research. 2021 Jul 2;49(W1):W293-W296.

187. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an Advanced Analysis and Visualization Platform for 'Omics Data. PeerJ. 2015 Oct 8;3:e1319.

188. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and Taxonomy in Diagnostics for Food Security: Soft-Rotting Enterobacterial Plant Pathogens. Anal Methods. 2016;8(1):12-24.

189. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. hybridSPAdes: an Algorithm for Hybrid Assembly of Short and Long Reads. Bioinformatics. 2016 Apr 1;32(7):1009-15.

190. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. Assembly of Long Error-Prone Reads Using de Bruijn Graphs. Proc Natl Acad Sci USA. 2016 Dec 27;113(52)

191. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-mer Weighting and Repeat Separation. Genome Res. 2017 May;27(5):722-36.

192. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE. 2014 Nov 19;9(11):e112963.

193. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality Assessment Tool for Genome Assemblies. Bioinformatics. 2013 Apr 15;29(8):1072-5.

194. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: Automated Circularization of Genome Assemblies Using Long Sequencing Reads. Genome Biol. 2015 Dec;16(1)

195. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res. 2016 Aug 19;44(14):6614-24.

196. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob Agents Chemother. 2014 Jul;58(7):3895-903.

197. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a Genome Comparison Visualizer. Bioinformatics. 2011 Apr 1;27(7):1009-10.

198. Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. Two Group A Streptococcal Peptide Pheromones Act through Opposing Rgg Regulators to Control Biofilm Development. PLoS Pathog. 2011 Aug 4;7(8):e1002190.

199. Bartilson M, Marra A, Christine J, Asundi JS, Schneider WP, Hromockyj AE. Differential Fluorescence Induction Reveals *Streptococcus pneumoniae* Loci Regulated by Competence Stimulatory Peptide. Molecular Microbiology. 2001 Jan;39(1):126-35.

200. Achen MG, Davidson BE, Hillier AJ. Construction of Plasmid Vectors for the Detection of Streptococcal Promoters. Gene. 1986 Jan;45(1):45-9.

201. DeLoney CR, Bartley TM, Visick KL. Role for Phosphoglucomutase in *Vibrio fischeri-Euprymna scolopes* Symbiosis. J Bacteriol. 2002 Sep 15;184(18):5121-9.

202. Peters JM, Koo B, Patino R, Heussler GE, Hearne CC, Qu J, et al. Enabling Genetic Analysis of Diverse Bacteria with Mobile-CRISPRi. Nat Microbiol. 2019 Jan 7;4(2):244-50.

203. Lao J, Guédon G, Lacroix T, Charron-Bourgoin F, Libante V, Loux V, et al. Abundance, Diversity and Role of ICEs and IMEs in the Adaptation of *Streptococcus salivarius* to the Environment. Genes. 2020 Aug 26;11(9):999.

204. Ambroset C, Coluzzi C, Guédon G, Devignes M, Loux V, Lacroix T, et al. New Insights into the Classification and Integration Specificity of *Streptococcus Integrative* Conjugative Elements through Extensive Genome Exploration. Front Microbiol. 2016 Jan 6;6

205. Liu M, Li X, Xie Y, et al. ICEberg 2.0: an Updated Database of Bacterial Integrative and Conjugative Elements. *Nucleic Acids Res*. 2019;47(D1):D660-D665.

206. Wang M, Goh Y, Tai C, Wang H, Deng Z, Ou H. VRprofile2: Detection of Antibiotic Resistance-Associated Mobilome in Bacterial Pathogens. Nucleic Acids Research. 2022 Jul 5;50(W1):W768-W773.

207. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The Comprehensive Antibiotic Resistance Database. Antimicrob Agents Chemother. 2013 Jul;57(7):3348-57.

208. Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2023). The Conserved Domain Database in 2023. *Nucleic acids research*, *51*(D1), D384–D388.

209. Geer, L. Y., Domrachev, M., Lipman, D. J., & Bryant, S. H. (2002). CDART: Protein Homology by Domain Architecture. *Genome research*, *12*(10), 1619–1623.

VITA

Brian Choi was born in Carmel, IN on November 1, 1994 to David and Min Choi. He attended Indiana University in Bloomington, Indiana, where he earned a Bachelor's of Science in Microbiology with minors in Philosophy and Chemistry in May 2017. As an undergraduate, he worked in the laboratory of Dr. Michael Lynch characterizing evolutionary dynamics in *Escherichia coli* under the mentoring of Dr. Megan Behringer. Brian went on to earn his Masters of Science in Physiology at Indiana University Purdue University Indianapolis in May 2018 where he realized his passion for medical research. Later that year, Brian matriculated into the Loyola Stritch School of Medicine MD/PhD program joining Dr. Alan Wolfe's laboratory within the Department of Microbiology & Immunology.

Brian's research pursues the microbial etiologies of lower urinary tract symptoms primarily in women with urinary tract infections and urgency urinary incontinence. His work included the announcement of three new microbial species including *Aerococcus loyolae,* named after Loyola University. Brian's dissertation work on characterizing these new species was supported by the Arthur J. Schmitt Fellowship, awarded in 2023. After completion of his doctorate, Brian will return to medical school to complete his medical degree in May 2026. His ambitions include maintaining an academic medical career as a physician scientist in Infectious Disease.