U.PORTO

PHD IN PHARMACEUTICAL SCIENCES
MICROBIOLOGY SPECIALITY

Comprehensive urogenital microbiome profiling: towards better understanding of female urinary tract in health and disease

Magdalena Dorota Księżarek

D

2022

Comprehensive urogenital microbiome profiling: towards better understanding of female urinary tract in health and disease

Magdalena Dorota Księżarek

Faculdade de Farmácia

D
2022

U.PORTO

PHD IN PHARMACEUTICAL SCIENCES
MICROBIOLOGY SPECIALITY

Comprehensive urogenital microbiome profiling: towards better understanding of female urinary tract in health and disease

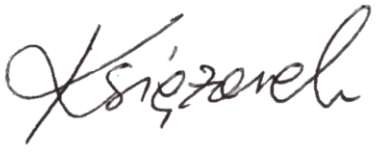
Magdalena Dorota Księżarek

This work was supervised by Professor Dr. Luísa Peixe and co-supervised by Dr. Filipa Grosso and Dr. Svetlana Ugarcina Perovic (Faculty of Pharmacy, University of Porto, Portugal).

January 2022

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Magdalena Dorota Księżarek



_____

This work was performed at the Microbiology Group at the Faculty of Pharmacy, University of Porto (FFUP, Porto, Portugal). The author received financial support through an individual pre-doctoral fellowship funded by the Foundation for Science and Technology (FCT) of Portugal with the reference SFRH/BD/132497/2017.

Acknowledgements

I would like to express my sincere appreciation to my PhD advisor, Professor Luísa Peixe for supporting me in this PhD project and for an infinite source of motivation. All the advice and guidance on good research enabled the successful outcome of this work.

I'm also very grateful to my co-supervisors Dr. Filipa Grosso and Dr. Svetlana Ugarcina Perovic for all input and effort to help me improve my weaknesses, but also for the patience and enthusiasm during everyday work.

In addition, I would like to thank to Dr. Joana Mourão, Dr. Joana Rocha, Dr. João Botelho, Andreia Rebelo, Dr. Teresa Gonçalves Ribeiro, Dr. Ângela Novais, Barbara Duarte, Dr. Ana Freitas, Dr. Liliana Silva, Dr. Joana Campos, Dr. Carla Rodrigues, Professor Carla Novais, Professor Patrícia Antunes and Professor Anabela Cordeiro for professional advice, personal support, and words of encouragements and to all my colleagues from the microbiology lab. Moreover, I'm grateful for all support received from Cristina Pinto da Costa and Nuno Oliveira from Laboratory of Microbiology, Faculty of Pharmacy and from Andrea Gouveia from Academic Service at Faculty of Pharmay, University of Porto.

I want also to acknowledge to Professor Maria de São José Nascimento for allowing to perform my PhD work at the Laboratory of Microbiology of the Faculty of Pharmacy of University of Porto and to UCIBIO@REQUIMTE Investigation Unit for providing the conditions necessary to perform this work.

Finally, I would like to acknowledge Foundation for Science and Technology (FCT) of Portugal for the financial support (with the reference SFRH/BD/132497/2017) and by that, for giving me possibility to continue this PhD project.

Abstract

The microbiome of urinary tract (UT) is among the more recently recognized in humans. Although potentially playing an important role in UT health, fundamental knowledge namely in compositional pattern, longitudinal stability, and species and strain level diversity is lacking for future research studies leading to diagnostic and interventions in UT diseases.

The main goals of this thesis were to comprehensively characterize the bacterial composition of the UT microbiome of reproductive-age asymptomatic women, its long-term stability and interconnection with vaginal microbiome, using high-resolution and accurate culturomic analysis, supported by DNA-based approaches. We evaluated the potential of female urinary microbiome (FUM) to act as a reservoir of putative uropathogens and unveiled features of microbiome members favoring their ability to colonize and persist in the asymptomatic UT. Complementing the studies in the asymptomatic population, a preliminary snapshot of FUM from diseased female cohorts (overactive bladder syndrome and recurrent UTI) was also conducted.

The overall findings of this thesis demonstrate that FUM is highly diverse (297 species), with several community structure types that might be associated with healthy host. Application of both methodologies (culturomics and amplicon sequencing) was highly beneficial to depict more diverse bacterial communities than previously anticipated (median = 53 species/sample; only 22% of species could be identified by both methodologies). Although at genus level *Lactobacillus* was present in all samples, at species level FUM does not have a specific bacterial core. This work also contributed to unveil the outstanding diversity of *Lactobacillus, Gardnerella* and *Corynebacterium* species inhabiting healthy UT. Moreover, although many potentially beneficial bacteria are usually present, healthy individuals are also often colonized by putative pathogens.

Through fine-grained analysis we contributed to description of 3 novel species (*Lactobacillus mulieris, Limosilactobacillus urinaermulieris, Limosilactobacillus portuensis*) and detection of 7 putative novel *Corynebacterium* species in healthy FUM. Moreover, we evaluated the MALDI-TOF MS effectiveness for FUM characterization, since it is the most commonly applied identification technology in UT microbiome studies. This demonstrated that MALDI-TOF MS (using the current databases oriented for clinically relevant species) may often misidentify bacteria belonging to the most prevalent groups inhabiting UT (e.g., *Lactobacillaceae, Gardnerella*).

Longitudinal study on our cohort revealed that FUM may shift overtime, although certain bacterial communities appear to be more stable (e.g., those with abundant *L. crispatus*), which could be a determinant for their high fitness abilities and lower susceptibility to dysbiosis. Thus,

there might be different microbiome structures during the individual lifespan with similar functions. Further characterization and comparison with paired vaginal microbiome demonstrated that although both niches are composed of highly similar taxa, the specific differences might be observed at species level (differential abundance of *L. urinaemulieris,* or species co-occurrence patterns e.g., *Lactobacillus jensenii* and *Lactobacillus iners* positively correlated in vaginal microbiome, while negatively correlated in FUM). Urinary and vaginal samples from the same woman were substantially different and shared an average of 38% species which also supports usage of mid-stream voided urine as representative sample for description of FUM.

Our sneak peek data on UT disease cohort demonstrated that this group, even in the absence of symptoms, is more likely to be colonized with higher number of putative pathogens, including difficult to culture species (e.g., *Ureaplasma parvum*, *Ureaplasma urealyticum*).

Additional whole genome sequencing (WGS) of selected isolates unveiled particular characteristics of FUM members at strain level. The work on *Escherichia coli* demonstrated that most strains inhabiting healthy UT belong to extraintestinal pathogenic *E. coli* lineages and are highly related to the strains causing UTI worldwide. Moreover, they might be resistant to antibiotics commonly used in the clinic which is of particular concern. Additional characterization of *Prevotella corporis* and *Prevotella brunnea* urogenital strains, unveiled genomic features that could contribute to their pathogenic potential (e.g., sialic acid synthesis). WGS of *L. urinaemulieris* and *L. portuensis* added information regarding their potential contribution for urogenital health (e.g., predicted ability to produce lactic acid). These whole genome-based works provide the basis for further studies exploring these and other species' roles in urinary tract homeostasis.

In summary, this thesis contributed to an in-depth understanding of healthy FUM, demonstrating that it is composed of complex bacterial communities, and it is highly variable between and within individuals over time. Lack of a single bacterial species common to all asymptomatic individuals and, especially, the presence of putative uropathogenic species and without clinical signs of infection, suggest that microbial interactions within microbiome are likely to play crucial role in UT homeostasis.

Keywords: Culturomics, amplicon-sequencing, midstream voided urine, novel species, uropathogens

Resumo

O microbioma do trato urinário (TU) foi dos últimos a ser reconhecido no corpo humano. Embora desempenhe potencialmente um papel importante na saúde do TU, continua a não existir um conhecimento de base, nomeadamente na composição da comunidade bacteriana, diversidade de espécies e estirpes, estabilidade ao longo do tempo, essencial a estudos futuros que conduzam a diagnósticos e abordagens terapêuticas em doenças do TU.

Os principais objetivos desta tese foram caracterizar de forma abrangente a composição bacteriana do microbioma do TU de mulheres assintomáticas em idade reprodutiva, a sua estabilidade a longo prazo e interligação ao microbioma vaginal, através de análise culturómica, apoiada por metodologias baseadas em DNA. Avaliou-se o potencial do microbioma urinário feminino (FUM) em atuar como um reservatório de uropatogénicos revelando características importantes de membros do microbioma que favorecem a sua capacidade de colonizar e persistir no TU assintomático. De forma a complementar os estudos na população assintomática, foi também realizado um estudo preliminar de FUM em coortes de mulheres doentes (síndrome da bexiga hiperativa e infeções recorrentes do TU).

Os resultados gerais desta tese demonstram que o FUM apresenta uma grande diversidade (297 espécies), com vários tipos de estrutura associados ao hospedeiro saudável. A aplicação das duas metodologias (culturómica e sequenciação) foi imprescindível para descrever comunidades bacterianas mais diversas do que o esperado (mediana = 53 espécies / amostra; apenas 22% das espécies puderam ser identificadas por ambas as metodologias). Embora o género *Lactobacillus* estivesse presente em todas as amostras, FUM não parece apresentar um núcleo bacteriano específico ao nível das espécies. Este trabalho também contribuiu para desvendar a notável diversidade de espécies de *Lactobacillus, Gardnerella* e *Corynebacterium* que residem no TU saudável. Além disso, apesar de muitas bactérias potencialmente benéficas estarem geralmente presentes, os indivíduos saudáveis também são frequentemente colonizados por potenciais patogénicos.

Através de análises muito detalhadas foi possível descrever 3 novas espécies (*Lactobacillus mulieris, Limosilactobacillus urinaermulieris, Limosilactobacillus portuensis*) assim como 7 potenciais novas espécies de *Corynebacterium* em FUM saudável. A avaliação da capacidade do MALDI-TOF MS na caracterização de FUM, uma vez que é a tecnologia de identificação mais comumente aplicada em estudos de microbioma UT demonstrou que MALDI-TOF MS (usando as bases de dados atuais contendo sobretudo espécies clinicamente relevantes) pode muitas vezes identificar erroneamente bactérias pertencentes aos grupos mais prevalentes que habitam o TU

(por exemplo, *Lactobacillaceae, Gardnerella*). Através de um estudo longitudinal verificou-se que o FUM pode mudar com o tempo, embora certas comunidades bacterianas pareçam mais estáveis (por exemplo, aquelas com maior abundância de *L. crispatus*), o que pode ser um determinante para menor suscetibilidade à disbiose. Assim, pode haver diferentes estruturas do FUM durante a vida do indivíduo com funções semelhantes. A caracterização e a comparação com o microbioma vaginal demonstraram que, embora os dois nichos sejam compostos por grupos bacterianos semelhantes, as diferenças surgem sobretudo ao nível de espécie (abundância relativa de *L. urinaemulieris* ou padrões de co-ocorrência de espécies, por exemplo, *Lactobacillus jensenii* e *Lactobacillus iners* positivamente correlacionados no microbioma vaginal, enquanto negativamente correlacionados no FUM). As amostras urinárias e vaginais da mesma mulher foram substancialmente diferentes compartilhando uma média de 38% das espécies, o que também suporta o uso de urina de jato médio como amostra representativa para a descrição de FUM.

Os dados obtidos com um pequeno coorte de doença do TU demonstraram que este grupo, mesmo na ausência de sintomas, é mais frequentemente colonizado com maior número de potenciais patogénicos, incluindo espécies difíceis de detetar em cultura (por exemplo, *Ureaplasma parvum, Ureaplasma urealyticum*).

Adicionalmente, a sequenciação do genoma completo (WGS) de isolados selecionados revelou características particulares de estirpe. Por exemplo, verificou-se que a maioria das estirpes de *Escherichia coli* que habitam o TU saudável pertencem a linhagens patogénicas extraintestinais de *E. coli* e estão altamente relacionadas com estirpes que causam infeções do TU em todo o mundo. Além disso, podem apresentar resistência aos antibióticos comumente usados na clínica, o que é particularmente preocupante. Com a caracterização genómica adicional de estirpes de *Prevotella corporis* e *Prevotella brunnea* identificaram-se possíveis características que poderão contribuir para o seu potencial patogénico (por exemplo, síntese de ácido siálico). WGS de L. urinaemulieris e L. portuensis adicionaram informações sobre sua potencial contribuição para a saúde urogenital (por exemplo, potencial para produzir ácido láctico). Todos estes trabalhos baseados no genoma fornecem a base para estudos adicionais que permitirão explorar os papéis dessas e de outras espécies na homeostase do trato urinário.

Em suma, esta tese contribuiu para uma compreensão aprofundada do FUM saudável, demonstrando que a sua composição inclui comunidades bacterianas complexas e que é altamente variável intra- e inter-individualmente ao longo do tempo. A falta de uma única espécie bacteriana comum a todos os indivíduos assintomáticos e, especialmente, a presença de espécies

potencialmente uropatogénicas na ausência de sinais clínicos de infeção, sugere que as interações microbianas no microbioma provavelmente desempenham um papel crucial na homeostase do TU.

Palavras-chave: Culturómica, sequenciação de amplicões, urina de jato médio, novas espécies, uropatogénicos

Table of contents

List of Figures

List of Boxes

List of Tables

List of Abbreviations

AMP, antimicrobial peptide
ARG, antimicrobial resistance gene
AV, aerobic vaginitis
BAP, blood agar plate
BV, bacterial vaginosis
CFU, colony forming unit
CNS, central nervous system
COG, cluster of orthologous group
CoNS, coagulase-negative *Staphylococcus*
ECM, extracellular matrix
EQUC, enhanced quantitative urine culture
FMT, fecal microbiota transplantation
FUM, female urinary microbiome
G+C content, guanidine + cytosine content
GABA, gamma-aminobutyric acid
GBS, group B *Streptococcus*
GRAS, Generally Recognized As Safe
HMP, Human Microbiome Project
IBC, intracellular bacterial community
IC, interstitial cystitis
iHMP, integrative Human Microbiome Project
LAB, lactic acid bacteria
LEA, *Lactobacillus* epithelium adhesins
LUT, lower urinary tract
LUTS, lower urinary tract symptoms
MAG, metagenome-assembled genome
MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MMP-8, matrix metalloproteinase-8
MRAS, methicillin-resistant *Staphylococcus aureus*
MSU, midstream urine
MUI, mixed urinary incontinence
OAB, overactive bladder

OTU, operational taxonomic unit
PBS, painful bladder syndrome
QIR, quiescent intracellular reservoir
RA, relative abundance
rRNA, ribosomal RNA
rUTI, recurrent urinary tract infection
SDS, sodium dodecyl sulfate
SLP, surface layer proteins
SPA, suprapubic aspiration
ST, sequence type
SUC, standard urine culture
SUI, stress urinary incontinence
TUC, transurethral catheter
UGT, urogenital tract
UI, urinary incontinence
UPEC, uropathogenic *Escherichia* coli
UT, urinary tract
UTI, urinary tract infection
UUI, urgency urinary incontinence
VIC, vaginal intracellular communities

Introduction

1.1. Urinary tract microbiome

The first description of bacteria inhabiting healthy humans (then called "animalcules") can be found in Antonie van Leeuwenhoek's letters, reaching early 1680s[1]. Over the course of time, several more reports on healthy 'microflora' appeared in the literature[2–4], however this research topic never gained much attention and, at that time, was not extensively explored. With technologies moving forward, the interdisciplinary research towards characterization of normal 'flora' fully emerged in 2007 with the Human Microbiome Project (HMP), launched by the National Institutes of Health[5]. Over the past 15 years, with data provided by HMP and other human microbiome projects, researchers observed a tremendous growth in the knowledge of microbial communities inhabiting different body sites (microbiota), their genomes (metagenome) and the entire habitat, including microorganisms, genomes and environmental conditions (microbiome)[6]. Nevertheless, currently the terms "microbiome" and "microbiota" have interchangeable use[7]. The main research focuses on composition and functional characteristics of healthy microbiome and the impact of microbiome balance (eubiosis) and imbalance (dysbiosis) on human health[8,9].

To date, several large-scale studies have already provided essential information into composition, diversity and function of the microbiome associated with multiple habitats across the human body, namely oral and nasal cavity, skin, vagina and gastrointestinal tract (e.g., HMP, HMP2, iHMP, MetaHIT)[5,10–12]. Their research efforts to investigate human microbiome diversity and host-microbiome interactions will likely contribute to a better understanding of microbiome involvement in health and disease states and, hence, to an earlier diagnosis of potentially related diseases and development of therapeutic strategies[5,8,11].

Currently, the most explored is the microbiome of human gastrointestinal tract - gut microbiome. For instance, it was demonstrated that dysbiosis in gut contributes to inflammatory bowel disease development, including Crohn's disease and ulcerative colitis[13,14]. Additionally, changes in gut microbiome composition have been associated with diabetes type 2 or obesity[15,16]. The available research also suggests that dysbiosis of intestinal microbial community might be involved in autism spectrum disorder development, as it is already known that several bacterial groups can influence brain activity, e.g., specific strains of *Lactobacillus* sp. or *Bifidobacterium* sp. producing gamma-aminobutyric acid (GABA) which act as a neurotransmitter in the central nervous system (CNS)[16–19]. The knowledge

improvements regarding gut microbiome triggered development of novel therapeutic strategy, i.e., fecal microbiota transplantation (FMT), that has been implemented as a treatment of certain microbiome-associated pathologies[20–22]. Insights from other human body niches are also available, such as dysbiosis of skin microbiome as an important risk factor for atopic dermatitis development, or bacterial vaginosis (BV) where the shift of dominant genus is observed, from protective *Lactobacillus* spp. to facultative or strict anaerobes[23,24]. Thus, it is expected that other disorders with not fully understood etiology, such as urinary tract (UT) disorders, may also be associated with the microbiome compositional shifts. However, despite the large range of the tested body niches, the UT was not initially included in the large-scale human microbiome projects, due to the long-standing belief that healthy UT is sterile.

Breakthrough studies using DNA-based approaches to analyze voided urine samples from healthy male (2010) and female (2011) donors detected diverse microbial community in the urogenital tract (UGT)[25,26]. These studies were of great relevance for reassessing the existence of a healthy urinary tract microbiome, since healthy UT was considered sterile with exception of distal urethra. Further studies confirmed the presence of microbiome in the bladder, analyzing samples obtained by suprapubic aspiration and removing any doubts that the microorganisms observed were truly from the UT[27]. Similar to what occurs with other body sites, the urinary tract microbiome is thought to influence both health and disease susceptibilities through its collective metabolic activities, and microbe-microbe and microbe-host interactions. Therefore, unveiling healthy urinary tract microbiome might contribute to a better understanding of disease conditions, including urinary disorders which were not previously associated with microorganisms (e.g., overactive bladder, urinary incontinence), as well as urinary tract infection (UTI).

Importantly, the bacteria inhabiting urinary tract will encounter certain environmental challenges. Bladder is an environment with relatively low oxygen concentration, which could promote the growth of facultative or strict anaerobes. However, bladder urine oxygen concentration may vary among individuals (range of 0.47–51.5 mmHg), which depends on host's associated environmental factors (e.g., level of hydration, inhaled oxygen)[28]. Urine is also limited in nutrients and in healthy human, besides water, it contains mainly urea, creatinine, uric acid, chloride, sodium, potassium, sulphate, ammonium, and phosphate[29]. Urine also contains some amino acids (e.g., D-serine) and low amounts of glucose and hormones (e.g., oxytocin). Thus, urinary tract environment will likely favor the bacteria that

are able to utilize these components (e.g., urease-producers) and with highly versatile metabolism (e.g., *Escherichia coli*)[30,31]. Furthermore, due to low levels of iron, the expression of siderophores for iron acquisition will also facilitate bacterial colonization and survival[32]. Concerning urinary tract cellular structure, the urothelium covered with a layer of glycosaminoglycans (e.g., hyaluronic acid, heparin, chondroitin sulfate) act as a barrier for urine and residing bacteria, limiting their interactions with extracellular matrix (ECM)[33–35]. However, when bacteria express certain features e.g., mucosal attachment, biofilm formation, they may directly interact with ECM components[32,34]. For instance, bladder ECM contains proteins, proteoglycans, proteases (e.g., collagens, fibronectins, laminins) that can be used as an additional carbon source by bacteria producing e.g., hyaluronidases or collagenases, allowing them to penetrate mucosal surfaces[34]. Importantly, urinary tract also produces certain molecules involved in host defense for instance antimicrobial peptides (e.g., uromodulin), secretory immunoglobulin A or may have immune cells residing in the bladder (macrophages, T cells, NK cells, dendritic cells)[33]. The defense mechanism against these protective cells and molecules is relatively known just for some bacteria (e.g., production of EsiB protein by *E. coli* that binds to secretory immunoglobulin A)[36]. Overall, factors that control immune system education to recognize commensal and pathogenic bacteria in urinary tract, as well as bacterial features allowing adaptation to urinary tract despite physiological barriers are not understood yet.

Most studies that have been performed focused on the UT pathology, lacking information on bacteria colonizing UT and their features, thus UT host-microbiome interaction are still on a preliminary research stage.

The following sections summarize recent developments in culture-independent and culture-dependent methodologies for characterization of microbial communities and the current knowledge on urinary tract microbiome, its compositional fluctuations and potential impact on UT integrity and host health.

1.2. Characterization of urinary tract microbiome

To date, there are multiple strategies for characterization of urinary tract microbiome. This may apply to various sample collection methods, different methodological approaches that

may favorize detection of specific bacterial groups or processing largely heterogeneous cohorts.
Since there are no yet clear methodological standards for this research field, a variety of protocols is being used.

1.2.1. Sample collection

The sample collection method is still one of the most controversial aspects of urinary tract microbiome studies. Collection of a voided midstream urine (MSU) sample, as an easy and not invasive method, increases the risk of vulvovaginal contribution[37]. Therefore, analysis of MSU will most likely represent microbiome of UGT. Additionally, MSU also captures urethral bacteria which may be important to understand UT conditions. Unquestionably, suprapubic bladder aspiration (SPA) seems to be the best method to avoid vulvovaginal contamination, however, due to its invasive character and demanding procedure cannot be easily applied, especially to the healthy population that voluntarily collaborate in those studies. Alternatively, transurethral catheterization (TUC) is being used by many groups, although this method presumably introduces urethral bacteria into the bladder while inserting catheter[37,38]. Nevertheless, comparison of the results obtained from urine of the same healthy women, collected by those three methods showed that indeed using TUC microbiome profile is more similar to SPA than MSU[27]. Thus, urine collected by TUC and SPA can be used to represent microbiome of the UT, while studying specifically bladder microbiome only urine collected by SPA should be used.
Nevertheless, investigating urine samples collected by any of these methods is always performed for the purpose of interpretation of UT health, thus in the following chapters we will refer to the findings as urinary tract microbiome, independently of sample type used for the analysis.

While these aspects are undoubtedly crucial for accurate reporting of scientific findings, from practical point of view, MSU is the sample used for diagnosis of UT conditions. Therefore, diagnostic interpretation of UT health is currently based on bacteria inhabiting UGT. In fact, it may be accurate considering the study published by Thomas-White *et al.*, which demonstrated interlink between vaginal and urinary tract microbiome, for both, pathogenic and health-associated species, suggesting that microbiome of UGT is interconnected[39]. Moreover, most prevalent taxa (e.g., *Lactobacillus*,

Gardnerella) in UT have been detected by both MSU and TUC with the same efficiency[40,41], and eventual differences observed would relate to relative abundance, rather than taxonomic profiling[42].

Since it is still a very controversial topic, sample collection method alternatives are being proposed e.g., Peezy® midstream device which is intended to reduce vulvo-vaginal and urethral contamination[43]. This innovative method has been evaluated and, so far, its efficacy is still controversial[44,45].

Noteworthy, there are also additional factors besides sample collection methods that can affect results. For instance, delay and storage before sample processing can significantly influence anaerobes recovery[46,47]. Additionally, temperature at which urine will be transported to the laboratory is crucial to avoid the overgrowth/death of bacteria in the sample or interference with sample stability[48–50]. It was recently demonstrated that urine should be stored in colder temperature for the shortest time possible, and addition of specific preservatives might improve preservation of urinary tract microbiome composition[51,52].

1.2.2. Methodologies

The technological advances in the last decade enabled the application of next-generation sequencing technologies to DNA extracted directly from a sample, which led to identification of a greater number of microorganisms that previously recognized[5,12]. On the other hand, culturomics – improved microbial culture by using a set of different media, growth conditions and incubation times, combined with a comprehensive high-throughput identification – has also emerged as a successful tool to isolate a high number of living bacteria and to identify new species[53]. Both culture-dependent and culture-independent methods are quickly and continuously evolving in order to capture the nature of human microbiome and truly understand its role in human health[46,47,54,55].

Nevertheless, both have some limitations such as the inability to discriminate between live and dead cells for DNA-based methods, or inability to detect non-culturable bacteria for culture-based methods. For those reasons, the complementation of data obtained by DNA-based methods and culturomics is vital for a comprehensive knowledge on urinary tract microbiome composition and function. In the next section the main advances that occurred

in these experimental approaches will be presented, focusing mainly on methodologies principles and their current contribution to urinary tract microbiome characterization.

1.2.2.1. DNA-based approaches

Popular among microbiome researchers, DNA-based approaches are time efficient and have the potential to deliver high-throughput data. Direct sequencing of targeted gene (amplicon sequencing) or whole bacterial genomes (shotgun metagenomic sequencing) of complex microbial community resulted in identification of a broad range of bacteria inhabiting human body[55–58]. These methods enabled the initial profiling of complex urinary bacterial communities and revealed existence of inter-individual differences, in contrast to the findings obtained by routine urine culture[26,27,59,60]. In general, these methodologies are based on following steps: DNA extraction, library preparation and sequencing (**Box 1**), sequencing data pre-processing (e.g., quality control, filtering) and data analysis steps. Each step of the workflow is of critical importance to the quality and reproducibility of the sequencing data[61].

Box 1 - Next-generation sequencing platforms

In the past two decades, the development of next-generation sequencing (NGS), characterized by massive parallel sequencing process, together with developments on data analysis, allowed time and cost reductions for human microbiome studies. This high-throughput sequencing is currently divided in two significantly different approaches, namely 2nd generation sequencing also called short-read sequencing (usually fragments from 100 to 400 bp, with exception of 454 pyrosequencing that results in fragments up to 1000 bp) that requires amplification previous to sequencing, and 3rd generation sequencing where long length sequenced fragments (average read length > 15,000 bp) without previous template amplification are obtained with higher rate of sequencing errors. There are several short-read sequencing platforms available, however, the most used are MiSeq or HiSeq (Illumina), 454 (Roche) and Ion Torrent (Thermo Fisher). They are based on sequencing by synthesis method, which can be divided in two types: cyclic reversible termination characterized by usage of terminator molecule to prevent elongation such as Illumina, or single-nucleotide addition described for Ion Torrent or used in the past 454 Roche for which elongation will be inhibited by absence of next nucleotide. Nowadays, availability of long-read sequencing, such as PacBio (Pacific Biosciences) or MinION, GridION or the most recent PromethION (Oxford Nanopore Technologies - ONT's), allows sequencing of bigger length molecules and provide higher resolution of genomic structures[62–65].

Although all steps can lead to biased results, DNA extraction is an essential and challenging step, namely due to the differences in cell wall composition and structure among the microbiome members (**Figure 1**). Optimal cell lysis is required to obtain sufficient quantity of high-quality DNA and thus detect all members of the bacterial community present in the sample[48,66,67].

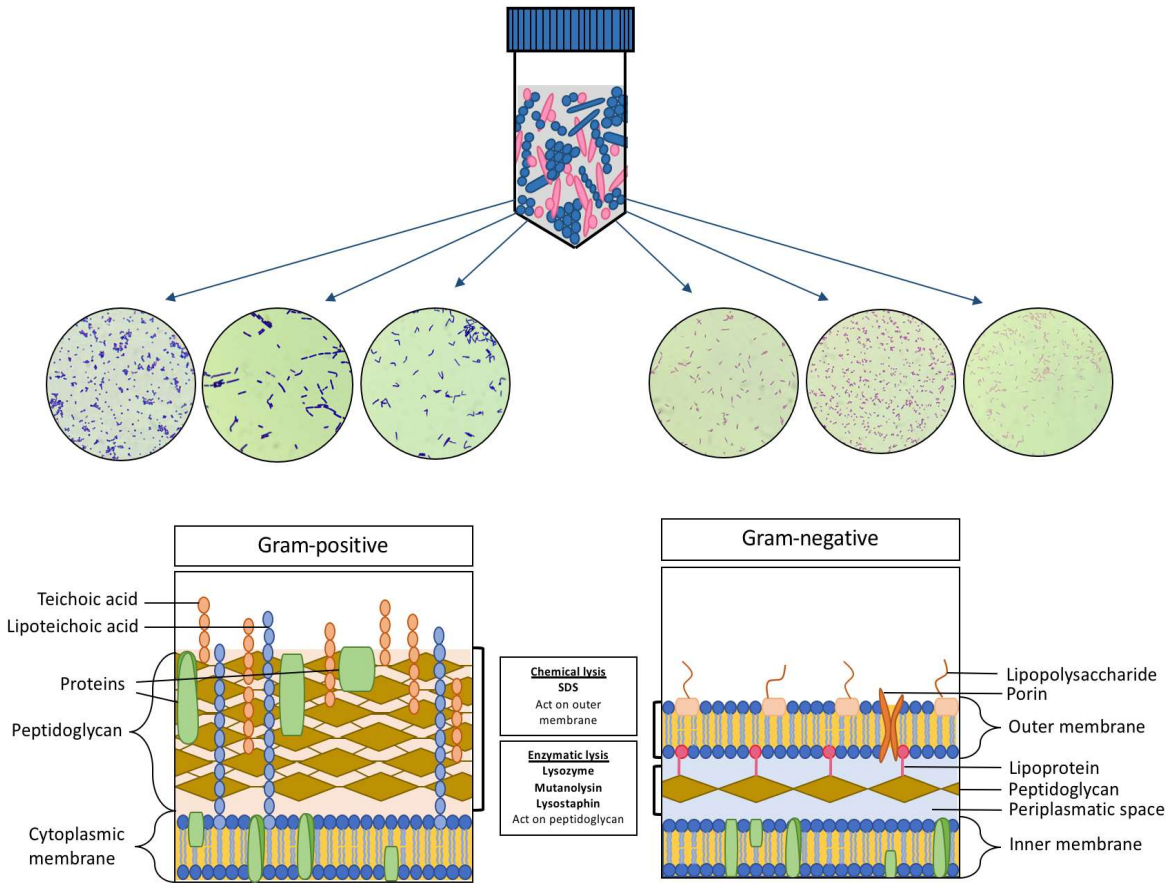


Figure 1 – Differences in cell wall composition of the bacterial community members.

The DNA extraction for microbiome studies is mostly processed based on enzymatic approach that may be supported with mechanical disruption i.e., adding bead-beating step to extraction protocols which may have beneficial effect for recovery of microbiome diversity[68,69]. Nevertheless, major approach is based on enzymatic cells' disruption. The protocol used for Gram-positive bacteria usually contains enzymes that improve cell lysis (thickness of Gram-positive cell wall may vary from 20 to 80 nm)[70] such as lysozyme and mutanolysin acting on

N-acetylglucosamine linkages and/or lysostaphin acting on cross-linked peptide bridges in peptidoglycan structure. This enzyme mixture is commonly associated with a buffer containing soft non-ionic detergent Triton X-100 acting gently on the cell membrane[71]. Anionic sodium dodecyl sulfate (SDS) is another type of detergent added, which denature proteins and strongly causes disruption of the cell membrane. SDS is the first-choice lysis buffer for DNA extraction of Gram-negative bacteria, and it is often combined with chelator EDTA which increases permeability of the outer membrane[72,73]. Of note, urinary tract microbiome studies conducted in previous years applied mostly enzymatic extraction protocols targeting Gram-positive bacteria[25,27,59,74–76] or occasionally used a mechanical extraction[60].

In fact, urine is a challenging type of sample and the DNA extraction method and kits used have to be carefully chosen as it may interfere with extraction of high-quality total DNA[77,78]. Moreover, as it is a relatively low-biomass sample, increased quantity of urine used for the DNA extraction will enhance chances to obtain sufficient DNA quantity to be suitable for sequencing[48].

1.2.2.1.1. Amplicon sequencing

Targeted amplicon sequencing is the most applied method in urinary tract microbiome research over the past 10 years. It allows to target a specific marker gene in the entire microbial community, and it is relatively low-cost. The 16S ribosomal RNA (rRNA) gene is usually the marker of choice for this method, due to its ubiquitous presence in bacteria and specific structure that makes it useful for taxonomic profiling (**Box 2**). The complete gene with its hypervariable regions allows relatively wide bacterial taxonomic profiling, however its different regions separately (often used strategy due to availability of short-read sequencing) have much reduced taxonomic resolution (**Box 2**). Thus, lack of standardization for methodologies used in urinary tract microbiome studies caused development of multiple diverse protocols, focused on different 16S rRNA gene regions. For instance, the most common choices in urinary tract microbiome studies are V1-V3, only V4, or V1-V2 and V6[26,27,51,59,60,74,79–83] which resulted in biased results and more importantly, will challenge future data integration[84,85]. For example, it has been already shown that *Prevotella* may be underrepresented by V6-V9 region, *Corynebacterium* can be overrepresented by V1-V3 and V3-V5, while *Gardnerella* may not be detected at all by V1-V3 16S rRNA gene regions[86].

Of note, for taxonomic profiling of eukaryotic microorganisms other genetic markers are being used, e.g., internal transcribed spacer 1 (ITS1) located between 18S and 5.8S rRNA genes[87,88].

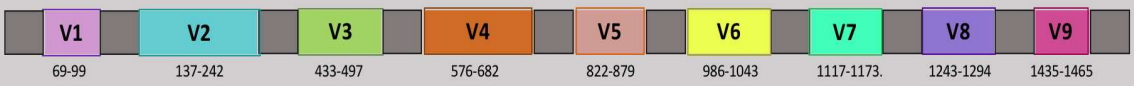
Amplicon sequencing data analysis is performed after quality and chimera checking by clustering closely related sequences into operational taxonomic units (OTUs) according to 16S rRNA gene similarity[55]. Alternatively, a more recent approach may be used, i.e., amplicon sequence variant (ASV) identification that distinguishes sequences with the accuracy of one nucleotide, thus presenting higher taxonomic resolution than based on similarity threshold OTUs clustering[89].
Independently of the chosen approach, taxonomic assignment is only achieved using different databases[56,90] (Ribosomal Database Project-RDP[91], SILVA rRNA database project[92], Greengenes[93]) in combination with bioinformatic tools (e.g., QIIME[94], mothur[95], DADA2[96]). Statistical analysis for diversity and relative abundance can be performed using different statistical programs (e.g., SAS, SPSS, PRIMER-E with PERMANOVA) and free R packages[97,98]. Each of the above-mentioned data analysis steps could introduce the bias in amplicon sequencing data interpretation.

The biggest flaw associated with this specific methodology is that it may overestimate or underestimate bacterial abundance by assuming that bacteria have the same number of copies of 16S rRNA gene per cell[55,66,99]. Furthermore, 16S rRNA gene might be highly similar for closely related species, thus it often allows for reliable identification only up to genus level[90].

Box 2 - The 16S ribosomal RNA gene

The 16S ribosomal RNA gene (*rrs*), present in all Bacteria and Archaea, is widely used for amplicon sequencing analysis, due to its taxonomic classification potential. This gene is usually present in the bacterial genome in several copies and includes highly conserved regions interspersed by nine hypervariable regions. The 16S rRNA gene among bacteria displays variable discriminatory potential which is the main limitation of this marker gene. High similarity level of the sequences between certain species belonging to the same genus is frequent, preventing an accurate species identification. Each region separately has its own potential which may be used for selection purposes, e.g., V1 allows to distinguish *Staphylococcus aureus* from coagulase negative Staphylococci or common pathogenic *Streptococcus* species. Individually, V2, V3 and V6 are the hypervariable regions with the highest discriminatory potential, enabling identification of most of the bacterial groups, though are recognized to be insufficient to properly distinguish members of *Enterobacteriaceae* family.

Nonetheless, even if partial sequencing may result in identification, incomplete amplicon will significantly reduce discriminatory potential of the gene marker. Using short-read sequencing technologies, selected amplified regions have to be interpreted individually, as the assembly would generate random errors, however this limitation can be solved with recently available single-molecule long-read sequencing that enables to sequence longer fragments. For an improved taxonomic identification (at genus or species level), it is essential the amplification of the entire variable region (V1-V9)[66,100–104].



Schematic representation of 16S rRNA gene (~1500 bp) comprising hypervariable (V1-V9) and conserved regions (marked in grey).

1.2.2.1.2. Shotgun metagenomic sequencing

While amplicon sequencing can provide initial taxonomic profiling of the microbiome, the shotgun metagenomic sequencing (or simply metagenomics) enables additional insights. Metagenomics allows for more accurate taxonomic profiling and provides information on whole genetic material i.e., genes that can be used to assess microbial functionality and characterization of e.g., virulence, metabolic pathways[105,106]. It does not limit detection only to bacteria allowing to obtain genetic information about all community members, including viruses and fungi which may also have an important role in the microbiome. Furthermore, it is also more efficient in detecting low-abundant and/or rare microbiome members[105,107,108], however there are also studies reporting the opposite[109].

In this methodology the total genomic DNA extracted from the sample is randomly fragmented and sequenced[61]. Currently, the short-read sequencing platforms are more commonly used in metagenomic studies[61,109], however long-reads will likely be more available for this purpose in the close future.

Analysis of metagenomic data may be challenging, but there are many bioinformatic pipelines that allow to do it in a fast and accurate way. The first steps of analysis are not much different

from any genomics workflow: obtained reads must be trimmed, quality-checked and filtered (e.g., to remove human-derived sequences)[58]. Finally, reads might be used directly for taxonomic profiling using reference genomes databases with appropriate post-processing strategies e.g., lowest common ancestor (LCA) by MEGAN program[110] or may be processed using discriminative markers e.g., unique clade-specific marker genes by MetaPhlAn (Metagenomic Phylogenetic Analysis)[61,111].
However, reads might also be assembled into contigs which will be further grouped to create draft genomes i.e., metagenome-assembled genomes (MAGs)[112]. There are many tools that can be used to perform these steps (e.g., MetaVelvet[113], metaSPAdes[114], MetaBAT[115], GroopM[116], Anvi'o[117]). Further genome exploration is also based on approaches widely used in genomics e.g., genes prediction, pangenome analysis and identification of core and accessory genome, predicting functionality with clusters of orthologous groups (COGs) functional categories, secondary metabolites search[106,118,119]. Noteworthy, tools chosen for data analysis may lead to differences in final results[108,120].

Although metagenomics allows to extract robust genomic information, it is still associated with relatively high cost and, although some human microbiome metagenomics studies are available, regarding UT, the majority was conducted based on targeted amplicon sequencing. To date, only few metagenomics studies have been performed on urine samples using Illumina HiSeq or Ion Torrent sequencing platforms[39,76,121,122]. They allowed initial observations that Bacteria are most abundant in urinary tract microbiome, followed by Eukarya (e.g., *Candida albicans*, *Candida glabrata*, *Malassezia globosa*) and Viruses (phages and human viruses e.g., herpesvirus, papillomavirus)[76]. There are also metagenomic studies focused only on viral population in urinary tract microbiome[123,124] confirming phages (including many putative novel prophages) as the type of viruses most often identified from urine. Some studies also performed metagenomics on UTI cohorts[125,126] demonstrating several benefits associated with using this approach for diagnostic purposes over culture-based methods.

1.2.2.2. Culturomic approaches

With the new insights from DNA-based methodologies, culture techniques have been improved with culturomic approach being more effective at isolation and identification of a large set of cultivable bacteria. Culturomics, high-throughput enhanced culturing, was initiated with fecal samples where a set of culture conditions (including nutrients,

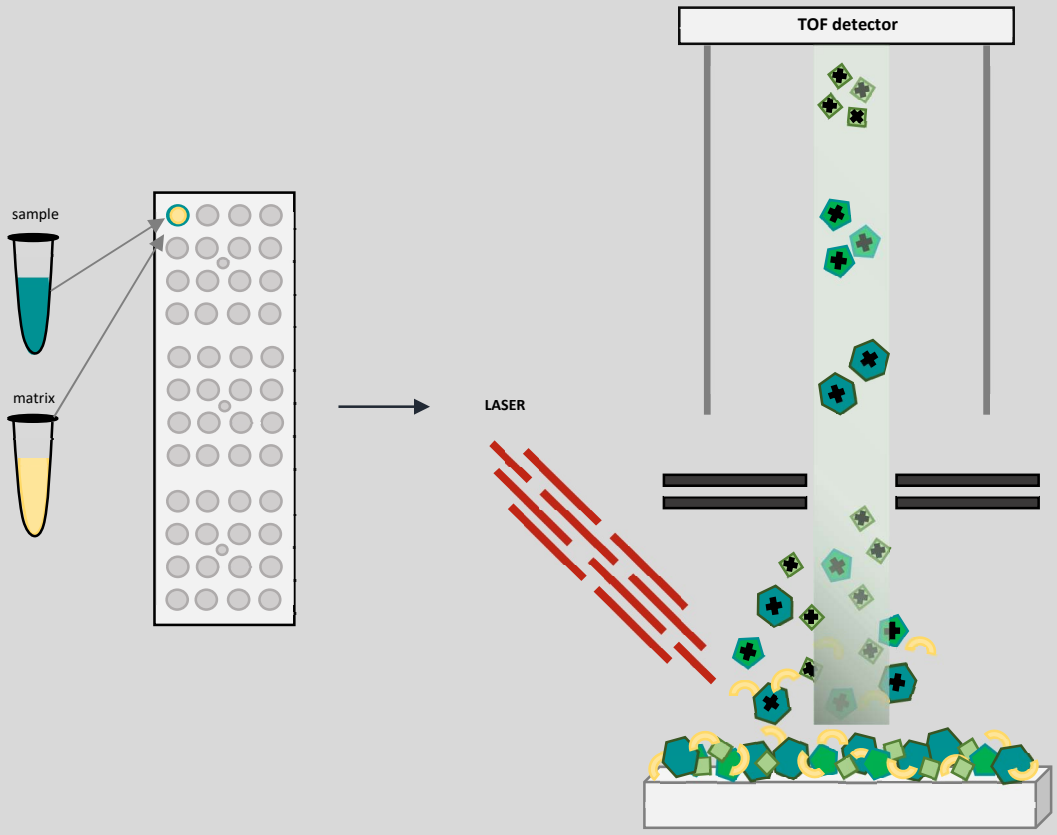
atmospheric conditions, incubation time and temperature) coupled with high-throughput isolate identification such as MALDI-TOF MS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) (**Box 3**) and/or 16S rRNA gene sequencing were used[127,128]. Culturomic approach allows in-vitro growth of various bacterial species for further phenotypical and/or genotypical characterization[57]. This opens multiple possibilities e.g., in-depth characterization of isolates concerning microbe-microbe and host-microbe interactions, in-vitro virulence testing, identification of possible antimicrobial substances or other therapeutic compounds produced by bacterial species that might further contribute to live biotherapeutics development[127,129,130].

The urine culturing protocol and interpretation of bacterial growth from urine established by Kass in 1956, combined with clinical symptoms, have not been challenged for years and remained as the golden standard for a UTI diagnosis[131]. Estimation of 100.000 bacteria per mL still constitutes a border between contamination and infection. Later, it was added to the clinical guidelines that, under special circumstances, and depending on the detected pathogen, growth of 10^3 CFU/mL should also be considered as significant[132]. Standard urine culture (SUC) designed for most common uropathogens establishes the culture of 1 µL of urine in selective, but nutritionally poor media, under aerobic conditions, and overnight incubation. Increasing evidence of the presence of bacterial species detected from healthy UT by DNA-based studies questioned this methodology. Comparison of SUC and improved urine culturing protocols performed on the same set of urine samples showed that most of tested urine samples had bacteria in urine, with SUC showing that more than 90% were negative[75,133]. Thus, the conventional microbiological methods, still used as routine diagnostic technique, have a limited scope to capture the full spectrum of bacterial species present in the urine samples.

Box 3 – MALDI-TOF mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry is an ultra-fast, relatively low-cost technique widely applied for identification of microorganisms. Besides identification purpose, this high-throughput analysis currently presents a variety of applications, for instance, reliable detection of common food-borne pathogens or bacteria responsible for UTI or detection of certain antibiotic resistant bacteria for instance methicillin-resistant *S. aureus* or

carbapenem-resistant Gram-negative rods. The rationale of this methodology is based on soft ionization of the sample previously incorporated in matrix and under the laser stimulation charged molecules are being generated. While the single ions are released from the sample, due to their different molecular mass they will require different times to arrive to the detector. Collected information about mass-to-charge ratio (m/z) and time-of-flight (TOF) is processed by the specific software and converted to spectra called peptide mass fingerprint (PMF) that will be compared with PMFs available in the database for the final identification. Even if available databases are being constantly enlarged, some of bacterial species might have similar or no reference spectra and require further identification by other more specific methodologies[134–137].



MALDI-TOF mass spectrometry preparation and detection process.

During the past decade, culturomic protocols had been properly adapted and applied to the urine samples. Nutrient-rich media included in the protocols, such as Blood agar plates (BAP),

constitute a valuable source of proteins and provide haemin (X factor) which contains iron required for some bacterial metabolic pathways, e.g., for *Streptococcus* sp., or Chocolate agar that provides additional growth factor, nicotinamide adenine dinucleotide (NAD or V factor) required to cultivate *Haemophilus* sp.. Additionally, Enterobacterales-inhibiting Colistin-Nalidixic Acid agar is also being used to facilitate the growth of Gram-positive bacteria, or Schaedler agar, often supplemented with vitamin K1 to support the difficult to growth *Lactobacillus* sp. or anaerobic bacteria such as *Prevotella* sp.. These fastidious bacteria additionally require prolonged incubation time and atmospheric conditions adapted to their oxygen sensitivity. Probably, the most relevant adjustment concerns quantity of sampled urine which also differs among the studies, from 10 μL to 100 μL[75,133], or even sampling re-suspended urine sediment[138]. Bacterial pure culture obtained with those protocols are subjected to identification by MALDI-TOF MS and/or by 16S rRNA gene amplification[40,75,133,138,139].

So far, streamlined EQUC (enhanced quantitative urine culture) has been proposed for supplementary diagnostic procedures regarding urine culture, especially for the samples reported with "no growth" by SUC or for patients experiencing recurrent UTI[139]. Importantly, the variability among currently used protocols often results in data at different taxonomic level, difficult to integrate and compare. For an accurate diagnosis, a new standard protocol should be established, as was already settled for fecal samples[66] (e.g., The International Human Microbiome Standards). Different culturomic protocols applied to the urinary tract microbiome studies are summarized in **Table 1**.

Table 1. Culturomic protocols applied to urinary tract microbiome studies

Protocol	Collection method	Urine volume	Culture media	Atmospheric condition	Temperature	Incubation time	Isolate identification	Reference
Culture of urinary sediment	MSU, TUC	NA	BAP	5% CO_2	37°C	48 h	Standard biochemical tests and 16S rRNA gene	Khasriya *et al*, 2013[40]
		NA	Fastidious anaerobic agar	anaerobic	37°C	7 days		
Enhanced quantitative urine culture EQUC	TUC	100 µl	BAP, Chocolate agar, CNA agar	5% CO_2	35°C	48 h	MALDI-TOF MS	Hilt *et al*, 2014[75]
		100 µl	BAP	aerobic	30°C	48 h		
		100 µl	BAP	aerobic	35°C	48 h		
		100 µl	CDC anaerobe 5% sheep blood agar	Campy gas mixture (5% O_2, 10% CO_2, 85% N)	35°C	48 h		
		100 µl	CDC anaerobe 5% sheep blood agar	anaerobic	35°C	48 h		
		1 ml	thioglycolate medium*	aerobic	35°C	5 days		
Streamlined EQUC	TUC	100 µl	BAP, CNA agar	5% CO_2	35°C	48 h	MALDI-TOF MS	Price *et al*, 2016[139]
		100 µl	MAC	aerobic	35°C	48 h		
Expanded urine culture EUC	MSU	10 µl	TSA with 5% sheep blood, Schaedler agar	aerobic	35°C	7 days	MALDI-TOF MS	Coorevits *et al*, 2017[133]
		10 µl	TSA with 5% sheep blood, Schaedler agar	anaerobic	35°C	7 days		
Culturomics of resuspended sediment of urine	MSU	resuspended pellet from 5 ml of urine	Chocolate agar	5% CO_2	37°C	48 h	V9 region of 16S rRNA gene	Curtiss *et al*, 2017[138]
		resuspended pellet from 5 ml of urine	Chocolate agar	anaerobic	37°C	7 days		

NA - not available, TUC - transurethral catheter, MSU - midstream urine, BAP - Blood agar plate, MAC - MacConkey agar, CNA - Colistin and Nalidixic Acid agar, TSA - Tryptic Soy agar, *only in case of growth; MALDI-TOF MS - matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

1.3. Urinary tract microbiome compositional fluctuations

Over the past decade, many studies using DNA-dependent and -independent approaches confirmed that, undoubtedly, UT and specifically bladder contains its own resident microorganisms independently of presence of UT symptoms[27,60,87,133,140]. Available results point to a relatively low bacterial biomass in UT environment, comparing to other body sites like the gut[48,80]. Additionally, microbiome is often dominated by one or two bacterial genera (>50% of community), which allowed for more useful communities categorization in more general 'urotypes', similarly to 'community state types' in vagina or 'enterotypes' in the gut[67,74,80,82,141–147]. The name of 'urotype' represents the name of dominant genus e.g., *Lactobacillus* urotype is the most reported so far. Also, samples that did not present dominant genus are categorized as 'diverse urotype'[74,80]. This microbiome classification however is based on most abundant genus, thus it may not reflect properly the microbiome diversity within the genus and especially, it may understate less abundant taxa. Moreover, recent comprehensive whole-genome analysis of bacteria isolated from female urinary, genital and gastrointestinal tract highlighted the importance of detailed analysis at the strain level[39]. Precisely, high similarity between some strains obtained from the UT and from vagina suggests that there is a continuum between those two body sites, thus the genital and urinary tract microbiome are interconnected[39]. The available data show that microbial community profiles determined by different methodologies complement each other and that urinary tract microbiome of individuals with LUT disorders differs from healthy controls[40,74,75], and might be also influenced by host-associated factors e.g., hormonal status, menstruation, smoking status, BMI[41,121,148–150]. Importantly, all current findings demonstrate correlations only and, if there is causational relationship between urinary tract microbiome and LUT disorders, it remains to be determined.

1.3.1 Urinary tract microbiome in health

Over past decade, several studies focused exclusively on urinary tract microbiome of asymptomatic population, and they are summarized in **Table 2**. They investigated microbiome of males and females, in relatively wide age range. Some additional data on 'healthy' population might be also extracted from control cohorts in the studies directing research focus on particular disease.

To date, the distribution of bacterial phyla among healthy population is relatively constant among the studies highlighting Firmicutes as the most dominant urinary tract microbiome members, followed by Actinobacteria, Bacteroidetes and Proteobacteria[26,75,81]. For instance, *Lactobacillus* sp. (Firmicutes) have been detected in almost all samples, being a dominant genus in more than half of investigated urine from healthy population[26,74,80–82]. Some of the *Lactobacillus* strains are known to be bacteriocin-producers, presenting inhibitory activity towards potential pathogens, thus it is possible that proper levels of lactobacilli inhibit overgrowth of Gram-negative putative pathogens also in the UT. Regarding Firmicutes, *Streptococcus* sp. and *Staphylococcus* sp. have been also often described. Actinobacteria phylum, which includes genera like *Gardnerella* sp., *Corynebacterium* sp., *Cutibacterium* sp., or *Actinomyces* sp., also seems to constitute an important part of urinary tract microbiome[41,75,79,145]. Finally, *Prevotella* sp., (from Bacteroidetes) and *Escherichia coli* (from Proteobacteria phylum) are also detected from healthy UT[26,42,81,82,133]. Apparently, urinary tract microbiome of asymptomatic individuals contains mostly commensal bacteria, however recognized uropathogens are also isolated[133].

Interestingly, although different sample collection may have an impact on the obtained results, it is suggested that the differences reflect mostly relative abundance of specific taxa rather than their prevalence[42]. Thus, the detection of above-mentioned bacteria is relatively consistent among studies, however different studies/protocols will likely report quantitative differences.

Table 2. A summary of studies and their compositional findings on urinary tract microbiome in health over past decade.

Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
Female	NA	8	MSU	Sequencing V1-V2 and V6 regions of 16S rRNA gene	- combination of more variable 16S rRNA gene regions allowed to capture extended bacterial diversity - urinary tract microbiome of asymptomatic women was characterized by complex microbial composition with intra-individual variations - 45 bacterial genera were identified with *Lactobacillus* sp., *Prevotella* sp. and *Gardnerella* sp. being most dominant - most individual microbiome profiles were dominated by one taxon (>75% RA) and microbiome with different genera distributed more even were observed less often	Siddiqui *et al*, 2011[26]
Male	14-17	18	MSU, CS swab *Monthly sample collection over three-month period	16S rRNA gene sequencing	- urinary tract microbiome was rich in obligate and/or facultative anaerobes - in urinary tract microbiome *Streptococcus* sp. and *Lactobacillus* sp. were most abundant, followed by *Gardnerella* sp. and *Veillonella* sp. - *Lactobacillus* sp. and/or *Streptococcus* sp. were absent in only two urine samples - urinary tract microbiome was less stable than microbiome of the coronal sulcus - comparison of different 16S rRNA gene regions revealed that although most taxa were detected in comparable RA, some might be influenced by region selection (e.g., *Prevotella* is underrepresented by V6-V9; *Gardnerella* is not detected by V1-V3; *Corynebacterium* is overrepresented by V1-V3 and V3-V5)	Nelson *et al*, 2012[86]
Female	NA	12	MSU, VS, TUC, SPA	SUC, sequencing V1-V3 regions of 16S rRNA gene	- microbiome composition obtained from urine collected by TUC was more similar to SPA than MSU - MSU samples contained mixture of urinary and genital tract microbiome - samples positive for *E. coli* by SUC demonstrated other fastidious bacteria in higher abundance than *E. coli* by culture-independent method, which could question the current UTI diagnostic approaches and accuracy	Wolfe *et al*, 2012[27]

Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
					- the findings were additionally supported with analysis of urine collected by TUC and SPA from symptomatic women (n = 11 POP/IU patients)	
Female	26-90	10	MSU	Sequencing V1-V3 regions of 16S rRNA gene	- female microbiome had higher diversity and was significantly different than male - phylum Firmicutes constituted 50% of bacteria in both genders - in females, 23 genera were detected in all age groups (e.g., *Lactobacillus, Streptococcus, Corynebacterium, Enterococcus, Finegoldia, Prevotella*), while in males only genus present across all age groups was *Staphylococcus* - certain genera were detected only in population above 70 years old i.e., *Jonquetella* sp., *Parvimonas* sp., *Proteiniphilum* sp., *Saccharofermentans* sp.	Lewis *et al*, 2013[60]
Male	39-83	6				
Female	23-65	86	MSU	SUC, EUC	- nearly all samples (91.1%) contained ≥10^4 CFU/mL of live bacteria - higher diversity and bacteria load was observed in female samples comparing to male - most dominant genera in both genders were *Lactobacillus* sp., *Streptococcus* sp., *Staphylococcus* sp., *Propionibacterium* sp. and *Corynebacterium* sp. - from total of 98 species identified, the most prevalent were *Staphylococcus epidermidis*, *Streptococcus anginosus*, *Lactobacillus jensenii*, *Lactobacillus crispatus* and *Gardnerella vaginalis* - uropathogens constituted less than 5% of all detected bacteria	Coorevits *et al*, 2017[133]
Male		15				
Female	NA	12	MSU	Sequencing with the ITS1 fungal ribosomal primer set	- first preliminary report on additional fungi in healthy urinary tract microbiome to previously reported *Candida* sp. and *Saccharomyces* sp. - fungi were detected in every urine sample from asymptomatic women - there was a high interindividual variability in fungal population	Ackerman *et al*, 2017[87]
Female	in ranges of 10-19, 20-29, 30-39, 40-49, 50-59, 60-	79	MSU	Culturomics of resuspended sediment of urine and	- there was no correlation observed between age and number of genera identified - overall, 60 bacterial genera were identified with *Streptococcus* sp., *Staphylococcus* sp., *Corynebacterium* sp. and *Lactobacillus* sp. being most dominant, and present in both, pre- and post-menopausal women	Curtiss *et al*, 2018[140]

Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
	69, 70-79, 80-89			sequencing of 16S rRNA gene	- *Lactobacillus* sp. was found to be more common in pre-menopausal women, while *Mobiluncus* sp. was more common genus in post-menopausal women	
Female	Mean 29.1 (SD=5.2)	8	MSU, PUS, RVS *Daily sample collection over three-month period	EQUC	- participants were randomized for *Lactobacillus rhamnosus* GR-1 and *Lactobacillus reuteri* RC-14 oral probiotic use (~40 days) - the use of oral probiotic did not change the ratio of *Lactobacillus*/uropathogens in urinary tract microbiome - at least one species of *Lactobacillus* dominated in most samples (n=7), while remaining sample was dominated by *Streptococcus agalactiae, Staphylococcus epidermidis* and *Corynebacterium tuberculostearicum* - the probiotic species were never recovered from none of the sample types	Wolff *et al*, 2019[151]
Female	Mean 40.4 (SD=3.9)	10	MSU	Sequencing V3-V4 regions of 16S rRNA gene	- significant differences were observed regarding microbiome composition after storage with and without AssayAssure® preservative - storage of urine without AssayAssure® led to RA variations in certain donors, especially regarding *Escherichia, Bacteroides, Faecalibacterium, Lactococcus, Gardnerella, Streptococcaceae, Stenotrophomonas* - overall, storage conditions also affected RA of *Lactobacillus, Klebsiella, Delftia, Gardnerella, Enterococcus* - addition of AssayAssure®, storage in colder temperatures and for shortest time possible was shown to be beneficial for preservation of urinary tract microbiome composition	Jung *et al*, 2019[52]
Female	Mean 48 (SD=14)	224 including 52 previously published	TUC	SUC, EQUC, sequencing V4 region of 16S rRNA gene	- Most common urotypes were *Lactobacillus*, *Streptococcus*, other (urotypes representing less than 5 samples), mixed, *Gardnerella* and *Escherichia* - *Gardnerella* urotype was common in young (mean age 36 years old) women - *Escherichia* urotype was more common in older (mean age 60 years old potentially postmenopausal) women - no correlation was found for *Lactobacillus* and age, menopausal status, parity or vaginal intercourse	Price *et al*, 2020[145]

Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
Female	Mean 29 (SD=5)	8 (previously published in Wolff et al., 2019)	MSU, PUS *Daily sample collection over three-month period	EQUC, sequencing V4 region of 16S rRNA gene (n=2)	- daily assessment for approximately 3 months showed that microbiome of urinary tract is dynamic and resilient - only MSU samples that demonstrated highly different microbiome from PUS were chosen for analysis - short-term changes in microbiome were related to personal factors e.g., menstruation, vaginal intercourse (e.g., increase in *Streptococcus* and *Staphylococcus* species after vaginal intercourse) - the most common patterns observed were *Lactobacillus* predominance (single species or changing species); changing *Lactobacillus* and *Gardnerella* predominance; *Streptococcus, Staphylococcus* and *Corynebacterium* in different ratios	Price *et al*, 2020[148]
Female	24-40 (mean 21.2)	6	MSU, TUC	SUC, sequencing V4 region of 16S rRNA gene	- MSU and TUC shared bacterial genera but in different RA, thus quantitative differences were observed rather than taxonomical - MSU had higher abundance of *Veillonella, Staphylococcus* and *Neisseria* while TUC had higher abundance of *Lactobacillus, Streptococcus* and *Gardnerella* - *Lactobacillus* and *Prevotella* were more abundant in female microbiome than in male, while *Streptococcus, Veillonella, Staphylococcus, Gardnerella, Enterobacter, Neisseria, Haemophilus* were more abundant in male microbiome	Pohl *et al*, 2020[42]
Male	20-61 (mean 31.7)	14				
Female	Mean 66.4; in ranges: >50-54 55-59 60-64 65-69 70-74 75-79 80-84	1600 twins	MSU	Sequencing V4 of 16S rRNA gene, metagenomics (n=178)	- urinary tract microbiome composition was distinct from proximal body part and was not related to stool microbiome - age was the main contributor to microbiome variance, followed by host genetics, menopausal status, history of prior UTI - urinary alpha diversity increased with increasing age - beta-diversity differed according to age - bacteria were the main microorganisms of urinary tract microbiome (99.64% of reads) - taxonomic profiling for the most dominant bacteria was highly similar from both DNA-based approaches used but metagenomics allowed to	Adebayo *et al*, 2020[121]

Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
	85-				identify much larger bacterial diversity within urinary tract microbiome, comparing to 16S rRNA gene profiling	
Female	18-50	4+5	MSU *Daily samples collection on weekday and weekend, morning and evening	Sequencing V4 region of 16S rRNA gene	- urine storage in -20°C or -80°C temperatures up to 72 hours did not have a major impact on microbiome diversity, while storage at 4°C could influence low abundant species - in the short-term assessment, there was minor alpha-diversity difference between samples collected in the morning and evening and samples collected on weekday and weekends - beta-diversity was maintained in most samples independently of collection time point and overall, urinary tract microbiome appeared to be stable over-short time	Bundgaard-Nielsen *et al*, 2020[51]
	5-10	5				
Male	18-50	5				
	5-10	5				
Female	Mean 37 (SD=9.5)	41	TUC	Sequencing V4 region of 16S rRNA gene (n=54)	- microbiome of pre- and postmenopausal women differed; postmenopausal women were characterized with higher alpha diversity than premenopausal women - based on microbiome composition, samples clustered into three urotypes - the most dominant genus in both group was *Lactobacillus*, and its RA decreased with advancing age - *Gardnerella* and *Prevotella* were more abundant in postmenopausal women	Ammitzbøll *et al*, 2021[82]
	Mean 63 (SD=6.9)	42				

NA - not available, RA - relative abundance, MSU - midstream urine, TUC - transurethral catheter, SPA - suprapubic bladder aspiration, PUS - peri-urethral swab, VS - vaginal swab, CS - coronal sulcus, RVS - rectovaginal swab, UI - urinary incontinence, POP - pelvic organ prolapse, UTI - urinary tract infection, EQUC - enhanced quantitative urine culture, EUC - expanded urine culture, SUC - standard urine culture, SD - standard deviation, CFU - colony forming units, ITS - internal transcribed spacer

1.3.1.1. Compositional fluctuations in gender

Male and female seem to have similar set of bacterial genera in urinary tract microbiome but in different relative abundance (**Table 2**). Microbiome of healthy women seems to be more diverse comparing to men, presenting higher quantity of microorganisms[60,133], however there are also studies reporting otherwise[124,143]. Nevertheless, currently there is more evidence supporting different microbiome according to gender, including microbial composition, bacterial abundance and alpha diversity measures[42,59,60,76].

Initially assumed, one of the main differences according to gender was that men are more prone to be colonized by species belonging to genus *Corynebacterium*, while in women *Lactobacillus* species seemed to dominate[25,59,152]. More recently, *Lactobacillus* domination in female urinary tract microbiome has been confirmed by additional reports, and also *Prevotella* sp. are being highlighted as more abundant in women, while observations regarding males are not so consistent. For instance, besides *Corynebacterium*, *Staphylococcus* and *Streptococcus* enriched male urinary tract microbiome[59], some studies highlight *Enterococcus* and *Pseudomonas*[76], or *Streptococcus*, *Veillonella*, *Staphylococcus*, *Gardnerella*[42]. This lack of congruence in the results is likely related to quantity of research done on male population. To date, female populations are more often studied (**Table 2**) probably due to the highest occurrence of lower urinary tract symptoms (LUTS) in women than men.

1.3.1.2. Compositional fluctuations in age

A few studies also suggested that changes in the urinary tract microbiome seem to occur with ageing (**Table 2**). To date, there is only scarce knowledge available on urinary tract microbiome in children which demonstrated that very young children (even neonates), have urinary tract microbiome[153,154]. Nevertheless, the research regarding urinary tract microbiome in babies and young children is in its infancy. In young girls, studies involving vaginal microbiome composition revealed the increasing abundance of *Lactobacillus* sp. with age, which contribute to lower vaginal pH and seems to confer a protective environment against UTI[155–157]. In fact, the recently demonstrated interlink of genital and urinary tract microbiome supports the fundamental role of microbiome in UGT health maintenance[39]. Till recently, it was thought that with ageing, the number of bacterial genera inhabiting female UT decreases e.g., lower incidence of *Lactobacillus* spp. in urinary tract microbiome in post-menopausal women comparing to pre-menopausal women[140]. However, the urinary tract microbiome study with the largest so far cohort (n=1600) published in middle 2020 demonstrated that alpha diversity increases in older

women and the age was recognized as a main contributor to microbiome variations[121]. Similar was concluded more recently, where post-menopausal women were characterized by higher alpha diversity, comparing to pre-menopausal women[82].

The different prevalence of UT disorders at different ages might be related to alterations in microbial community observed with ageing, however, the currently available correlation-based compositional findings are not sufficient to evaluate this hypothesis.

1.3.1.3. Compositional fluctuations in pregnancy

Noteworthy, within the healthy female population, pregnant women should be evaluated as individual cohort, due to physiological changes associated with ongoing pregnancy. To date, there are few studies available on urinary tract microbiome of pregnant women (**Table 3**), suggesting relatively similar compositions to a healthy state[142]. For instance, *Lactobacillus* and *Gardnerella* were the most common observed urotypes, but certain women were also dominated with *Staphylococcus*, *Enterococcus* or *Escherichia coli*[142]. Moreover, there are compositional differences observed in urinary tract microbiome of pregnant women with preterm delivery in comparison with women with term delivery or compositional fluctuations associated with cesarean delivery[158–160].

Table 3. A summary of studies and their compositional findings on urinary tract microbiome in pregnant women.

Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
Mean 32.2	73 non-pregnant	MSU	Sequencing V1-V3 regions of 16S rRNA gene	- compositional differences of bacteria-derived extracellular vesicles (EVs) were detected between non-pregnant and pregnant women - *Bacillus* sp. Evs were enriched in pregnant women, while e.g., *Pseudomonas* sp. and *Lactobacillus* sp. Evs were enriched in non-pregnant women - there were also Evs differences observed in pregnant women with normal and preterm delivery e.g., *Ureaplasma* sp. Evs and Evs from family *Veillonellaceae* (e.g., *Megasphaera* sp.) were more enriched in samples of women with preterm delivery comparing to woman with normal delivery	Yoo *et al*, 2016[159]
Mean 32.5	74 pregnant				
Mean 24.5 (SD=4.8)	48 with term delivery	MSU	Sequencing V4 region of 16S rRNA gene	- *Lactobacillus, Serratia, Prevotella, Atopobium* and *Gordonia* were the most abundant genera - alpha and beta diversity measures did not differ between cohorts (term/preterm delivery) - sequences representing *Prevotella, Sutterella, L. iners, Blautia, Kocuria*, Lachnospiraceae, and *Serratia marcescens* were enriched in preterm delivery cohort - *Serratia marcescens* was also associated with preterm delivery cohort by unsupervised clustering approach - *Lactobacillus, Shuttleworthia*, and *Atopobium vaginae* were associated with term delivery cohort	Ollberding *et al*, 2016[158]
Mean 26.8 (SD=5.4)	49 with preterm delivery				
19-41 (mean 30)	51 pregnant	TUC	SUC, EQUC, sequencing V4 region of 16S rRNA gene	- most pregnant women had detectable live bacteria in urinary tract - the most common urotypes were *Lactobacillus* and *Gardnerella* - *Lactobacillus urotypes* were less diverse than *Gardnerella*, while *Gardnerella* urotypes were characterized by relatively even distribution of other genera e.g., *Lactobacillus, Ureaplasma, Aerococcus, Atopobium* - microbiome of some women was dominated with e.g., *Staphylococcus*, *Enterococcus*, *E. coli* - *Lactobacillus* urotypes were composed of following species: *L. gasseri, L. jensenii, L. iners, L. johnsonii, L. crispatus* - SUC missed many potentially pathogenic species	Jacobs *et al*, 2017[142]

Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
19-43 (mean 31.43; SD=4.37)	30 undergoing cesarean section	TUC *Collected in pre-delivery and post-delivery period	SUC, Sequencing V3-V4 regions of 16S rRNA gene	- cesarean delivery influenced maternal urinary tract microbiome composition - bacterial richness and diversity increased in post-delivery samples (209 bacterial genera) comparing to pre-delivery (181 genera identified) - *Lactobacillus* was the most dominant genus in pre-delivery period, while *Prevotella* was most dominant genus in post-delivery period - comparing to pre-delivery, the post-delivery samples had decreased *Lactobacillus* and increased *Pseudomonas, Bacteroides* and *Ruminococcus* - there was a correlation observed in certain metabolic pathways and cesarean delivery	Liu *et al*, 2018[160]

Evs - extracellular vesicles, MSU - midstream urine, TUC - transurethral catheter, EQUC - enhanced quantitative urine culture, SUC - standard urine culture, SD - standard deviation

1.3.2. Urinary tract microbiome in disease

Compositional fluctuations of urinary tract microbiome in specific disease conditions have been the subject of huge interest since the beginning of urinary tract microbiome studies. The summary of available studies investigating human urinary tract microbiome with the focus on LUT disorders and selected urinary conditions is provided in **Table 4**.
One of the most investigated are LUT disorders, like urinary incontinence (UI), overactive bladder (OAB) or interstitial cystitis (IC) that have severe impact on the quality of life including health-related, social, psychological, and working functions. Lack of efficient therapeutic approaches, and the fact that patients are selectively responding to the available treatment, strengthen the need to better understand the etiology of these chronic disorders. Diagnosis of these persistent UT conditions is still primarily symptom-based, and treatment include antimuscarinic drugs which reduce contractions of the involuntary detrusor, decreasing the symptoms severity in some cases only[161,162].

The UI is a urinary disorder characterized by uncontrolled urine leakage, which affects patients' psychological condition and significantly reduces quality of life[163]. This condition is divided into specific types which have additional characteristics e.g., urgency urinary incontinence (UUI) when is associated with excessive need to urinate or stress urinary incontinence (SUI) associated with the pressure on the bladder[163]. There might be also mixed urinary incontinence (MUI) which is characterized by all above mentioned symptoms[150,163]. Current findings suggest that microbiome detected from UUI and SUI is more diverse comparing to control group and enriched with bacterial species belonging to Firmicutes and Proteobacteria phyla. *Stenotrophomonas* sp., *Methylobacterium* sp., and other Gram-negative bacteria previously reported as UT pathogens have been observed in higher quantity among UI patients than in non-affected controls[79,80,149]. Emerging uropathogens e.g., *Actinotignum schaalii* have been detected more likely from UI patients[145]. Additionally, decrease of relative abundance of *Lactobacillus* sp. in MUI cohorts[150], or in parallel with increase of *Gardnerella* sp. among UUI samples was also observed[74]. Interestingly, comparing diversity only among UUI group, less diverse microbial profiles were found to be associated with more severe symptoms[79]. Culture-dependent approach allowed identification of isolates at species level, which revealed intra-genus differences in both cohorts, e.g., higher prevalence of *Lactobacillus gasseri* in UUI, and *Lactobacillus crispatus* in 'healthy' group[74,80]. Several evaluations regarding microbiome composition and the risk of developing UTI after surgery for SUI[164,165] or response to incontinence treatment[80] are also available, suggesting existence

of microbiome compositional patterns associated with different cohorts (e.g., UUI vs non-UUI, responders vs. nonresponders). For instance, *Atopobium vaginae* and *Finegoldia magna* in higher abundance were associated with symptoms severity[165] or presence of *Lactobacillus iners* in urinary tract microbiome was associated with protective role against developing post-operative UTI[166].

The OAB syndrome is also manifested by urinary urgency without UTI, usually associated with urinary frequency and nocturia[161,162,167]. In the past, OAB syndrome was not considered of microbial origin, however, bacterial community detected so far in urine from OAB patients seems to be different in comparison to healthy controls. Apparently, one of the main changes associated with this condition is the decrease in *Lactobacillus* sp., accompanied by the increase in Gram-negative bacteria, e.g., *Proteus* sp. or other recognized pathogens[138,168]. Some other bacterial genera might be overrepresented in OAB group like *Sneathia* sp., or *Mycoplasma* sp.[75,81]. However, not all these findings have been confirmed by other research groups, which emphasizes the need for further studies with enlarged cohorts and standardized methodologies. Interestingly, the research into possible new therapeutic strategies has already started e.g., vaginal estrogen therapy, which influenced increased level of *Lactobacillus* in UT and slightly contributed to symptoms improvements[169].

Another often investigated condition is IC or painful bladder syndrome (PBS) that could be associated with urinary tract microbiome as reported by a few studies[170–172], however there are also reports saying that IC/PBS might not be associated with bacterial microbiome members[83,144,173]. Focusing on those that reported compositional changes, IC/PBS cohorts seem to have lower diversity and richness comparing to non-affected cohorts[170,171]. However, contrary to other LUT conditions, it was reported that IC/PBS cohort has microbiome enriched in *Lactobacillus* sp. and often contain bacteria belonging to *Enterococcus, Atopobium, Proteus or Cronobacter* genera[170] and may also be associated with fungal composition and abundance[174]. Of note, other study reported low prevalence of *Lactobacillus acidophilus* in IC cohort and higher levels of proinflammatory cytokines[171]. There are also insights into symptoms flare event (sudden worsening of the symptoms, typical in chronic and inflammatory diseases[175,176]), which also seems to be associated with microbiome enriched in certain fungi i.e., *Candida* sp. and *Saccharomyces* sp.[172].

It is important to mention that compositional research on urinary tract microbiome has been expanded to other disease conditions such as different clinical manifestations of

UTI[122,124,126,139,153,177–182], sexually transmitted infections[25,152], type 2 diabetes[146,183–185], bacterial vaginosis[143] and neuropathic bladder[59,186,187]. Furthermore, there are also studies available on pathologies related to cancer e.g., bladder cancer[188–200], prostate cancer[201,202], breast cancer[203]; diseases related to upper UT such as kidney stones[204–209], chronic kidney disease[210], kidney transplant[211–215] or other conditions e.g., gout[216], prostatitis/chronic prostatitis/chronic pelvic pain syndrome[88,217–219] and obstructive urinary retention[220].

Microbial colonization of the bladder and reported variation of bacterial communities among different cohorts strongly support the possibility that some of these dysfunctional conditions of UT have a bacterial etiology. Additionally, bacteria possess the ability to affect nervous system and to modulate immune system or even directly cause pain by nociceptors activation as it was already demonstrated for *Staphylococcus aureus*[221]. Presence of microbes with these properties in the UT may possibly modulate or stimulate nervous responses, since LUT functions are controlled by the CNS.

Table 4. A summary of studies and compositional findings on urinary tract microbiome in lower urinary tract disorders and additional selected urinary conditions.

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
URINARY INCONTINENCE							
non-UUI	Female	Mean 49 (SD=14)	25	TUC	EQUC, sequencing V4 region of 16S rRNA gene	- there were differences in median numbers of bacteria cultured from both cohorts (higher number for UUI) - increase in *Gardnerella* sp. and decrease in *Lactobacillus* sp. was observed in UUI group - *Actinobaculum schaalii, Actinomyces neuii, Aerococcus urinae, Arthrobacter cumminsii, Corynebacterium coyleae, Gardnerella vaginalis, Oligella urethralis* and *Streptococcus anginosus* were more frequently identified in UUI cohort - *Lactobacillus* species-level differences were observed according with cohorts i.e., in healthy: *L. crispatus*, in UUI: *L. gasseri*	Pearce *et al*, 2014[74]
UUI		Mean 63 (SD=12)	23				
UUI	Female	Mean 57.7 (SD=11.3)	155	TUC	quantitative PCR	- randomized double-blind trial for UUI treatment (onabotulinum toxin A and anticholinergic drug) - nearly 40% of samples had detectable bacterial DNA - the bacterial detection was more likely in patients with greater daily UUI episodes - presence of bacterial DNA was associated with risk to develop post-treatment UTI	Brubaker *et al*, 2014[222]
POP/UI	Female	35-89	54	TUC	SUC, Sequencing V1-V3 regions of 16S rRNA gene	- microbiome diversity and abundance of some bacteria (e.g., *Fulvimonas, Dyella, Klebsiella, Escherichia/Shigella, Pseudomonas, Actinobaculum*) were correlated to susceptibility to post-operative UTI - the risk of developing post-operative UTI was correlated with the levels of b-defensin-1, psoriasin and lactoferrin	Nienhouse *et al*, 2014[164]
UUI	Female	31-85 (mean 58.5)	90	TUC	Sequencing V4 region of 16S rRNA gene	- *Lactobacillus* was the most dominating genus, followed by *Gardnerella*, *Gardnerella/Prevotella*, *Enterobacteriaceae*, *Staphylococcus*, *Aerococcus* and *Bifidobacterium*	Pearce *et al*, 2015[141]

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- the most common urotypes were "*Lactobacillus*", "*Gardnerella*" and "diverse" - decrease of *Lactobacillus* was observed in women that developed UTI after UUI treatment	
Asymptomatic	Female	42-68	9	TUC	Sequencing V4 region of 16S rRNA gene	- there was a significant difference in RA of 14 bacteria between UUI and non-UUI cohort - decrease in microbiome diversity in UUI cohort was associated with increase in UUI symptoms severity - *Alteromonadaceae* sp, *Stenotrophomonas* sp., *Brevundimonas* sp, *Elizabethkingia* sp., *Methylobacterium* sp. were observed in higher RA in UUI cohort	Karstens *et al*, 2016[79]
UUI			10				
non-UUI	Female	Mean 49 (SD=14.7)	26-33	TUC	EQUC, Sequencing V4 region of 16S rRNA gene	- Most prevalent urotypes detected in both cohorts were *Lactobacillus* sp. (non-UUI: 45.5%, UUI: 22%) and diverse (non-UUI: 6%, UUI: 32%) - UUI cohort had more bacteria and higher diversity - effectiveness of treatment in UUI cohort (responders) was corelated to less diverse microbiome community, while nonreponders had higher microbiome diversity - since microbiome composition seemed to influence clinical response to the treatment, microbiome diversity could be used to predict treatment responses	Thomas-White *et al*, 2016[80]
UUI		Mean 61.5 (SD=11.5)	37-50				
SUI	Female	Mean 51 (SD=9.7)	197	MSU (n=174), TUC (n=23)	Sequencing V4 region of 16S rRNA gene	- microbiome diversity was associated with UI symptoms - urinary tract microbiome of women with urinary incontinence symptoms seemed to not be dominated by single taxa - *Lactobacillus* and *Gardnerella* urotypes were more frequent in premenopausal women and postmenopausal women under hormonal treatment - nonpredominant urotypes were more frequent in postmenopausal women without hormonal therapy	Thomas-White *et al*, 2017[149]
POP/SUI	Female	30-85 (mean 57)	55	TUC, VS, PS	Sequencing V4 region of 16S rRNA gene	- *Lactobacillus, Corynebacterium, Gardnerella, Staphylococcus* and *Enterobacter* were most abundant in urinary tract microbiome	Fok *et al*, 2018[165]

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- *Lactobacillus, Corynebacterium, Anaerococcus, Peptoniphilus* and *Gardnerella* were the most abundant in vagina - *Lactobacillus, Gardnerella, Prevotella, Anaerococcus*, and *Corynebacterium* were most abundant in perineum - higher abundance of *Atopobium vaginae* and *Finegoldia magna* in urinary tract microbiome was associated with symptoms severity - *Atopobium vaginae* was detected from urine, vagina and perineum, suggesting they may act as reservoirs for this bacterium	
POP/SUI	Female	25-85 (mean 57)	104	TUC, VS, PS *twice - 6 weeks post-operative period (TUC)	SUC, Sequencing V4 region of 16S rRNA gene	- *Lactobacillus* was the most common genus in all sample types (TUC, VS, PS), followed by *Gardnerella* - decrease of *L. iners* (prevalence and abundance) and increase in certain pathogens (e.g., *Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa*) was observed in women with post-operative UTI comparing to from before surgery - urinary tract microbiome composition at the day of surgery was associated with risk for developing post-operative UTI - *L. iners* could have protective activity against post-operative UTI	Thomas-White *et al*, 2018[166]
Asymptomatic	Female	Mean 53 (SD=10.8)	84	TUC	Sequencing V4-V6 regions of 16S rRNA gene	- the cohorts did not differ in *Lactobacillus* predominance - overall microbiome composition differed according to presence of MUI - it was possible to observe significant differences in microbiome composition between cohorts only for younger women (<51 years old) - microbiome with highly abundant *Lactobacillus* was characteristic for asymptomatic women, while decrease in *Lactobacillus* was more common for MUI cohort - *Lactobacillus* was also more abundant in women <51 - BMI and menopausal status seem to be associated with MUI	Komesu *et al*, 2018[150]
MUI		Mean 53 (SD=11.7)	123				
Asymptomatic	Female	Mean 53 (SD=11)	84	TUC, VS	Sequencing V1-V3 regions of 16S rRNA gene	- 60% of OTUs detected in urinary and vaginal samples overlapped - alpha diversity was higher in urine samples than vaginal	Komesu *et al*, 2020[223]
MUI			128				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- in both, urine and vagina, *Lactobacillus* was the most abundant genus, followed by *Gardnerella*, *Prevotella*, *Ureaplasma* - *Lactobacillus* sp. (mainly *L. iners, L. crispatus, L. gasseri, L. jensenii*) were detected in similar proportion from urine and vaginal microbiome and they showed relatively strong correlation in paired vaginal and urine samples - *Tepidimonas* and *Flavobacterium* were detected only from urine samples	
non-UI	Female	Mean 47 (SD=14)	150	TUC	EQUC	- bacteria were detected in 70% of samples, and they were more often in UI cohorts, comparing to non-UI - UI cohorts were characterized by higher species richness comparing to non-UI women - *Lactobacillus iners, Streptococcus anginosus, L. crispatus*, and *L. gasseri* were detected most often from non-UI cohort - *S. anginosus*, *L. iners*, *Staphylococcus epidermidis*, and *L. jensenii* were detected most often in SUI cohort, while *S. anginosus*, *L. gasseri*, *Aerococcus urinae* and *Gardnerella vaginalis* in UUI cohort - *Actinotignum schaalii, Aerococcus urinae, A. sanguinicola, and Corynebacterium lipophile* group were unique species found in significantly higher RA in at least one UI cohort, comparing to non-UI women	Price *et al*, 2020[224]
SUI		Mean 54 (SD=14)	50				
UUI		Mean 61 (SD=13)	109				
OVERACTIVE BLADDER							
non-OAB	Female	NA	18	TUC	SUC, EQUC, sequencing V4 region of 16S rRNA gene	- *Lactobacillus* sp., *Streptococcus* sp., *Corynebacterium* sp., *Actinomyces* sp. and *Bifidobacterium* sp. were identified in both cohorts - *Lactobacillus* sp., *Streptococcus* sp., *Bifidobacterium* sp., *Staphylococcus* sp., *Enterococcus* sp., *Micrococcus* sp. were more abundant in asymptomatic women - *Corynebacterium* sp., *Actinomyces* sp. were more abundant in OAB cohort	Hilt *et al*, 2014[75]
OAB			34				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- *Aerococcus* sp., *Actinobaculum* sp. were identified only in OAB cohort	
OAB	Female	61	1	NA *twice, repeated after one year	SUC, Sequencing V1-V2 and V6 regions of 16S rRNA gene	- at the first sampling the patient showed negative analysis of urine dipstick, positive SUC with one single bacterium (*Streptococcus viridians* group), and diverse microbiome by 16S rRNA gene sequencing (additional detection of *Atopobium, Ureaplasma, Prevotella, Bacteroides urealyticum*) - at the second sampling, still with persistent symptoms, urine dipstick and SUC were negative but sequencing still showed diverse microbiome (*Streptococcus, Ureaplasma, Prevotella*, trace of *Bacteroides urealyticum*) - use of sequencing methods should be considered for patients with chronic urinary symptoms	Siddiqui *et al*, 2014[225]
Asymptomatic	Female	41-83	35	MSU	Culturomics of resuspended sediment of urine and sequencing V9 region of 16S rRNA gene	- overall, *Staphylococcus* sp., *Streptococcus* sp., *Corynebacterium* sp. and *Lactobacillus* sp. were most frequently identified - significant differences in prevalence of certain bacteria was observed between asymptomatic women and women with OAB - *Proteus* sp. was identified more often in OAB cohort - *Lactobacillus* sp. was identified more often in asymptomatic controls - recognized pathogens have been more likely identified from OAB group	Curtiss *et al*, 2017[138]
OAB		14-87	60				
Asymptomatic	Female	Mean 26	25	TUC	Sequencing V4 region of 16S rRNA gene	- microbiome composition was significantly different in asymptomatic comparing to OAB women - higher diversity of urinary tract microbiome was observed in asymptomatic cohort - *Lactobacillus* sp. was dominant genus in half of the asymptomatic women - in OAB cohort RA of 7 genera increased e.g., *Bifidobacteriaceae*, *Proteus* sp. and *Aerococcus* sp. and RA of 13 genera decreased e.g., *Lactobacillus* sp. and *Prevotella* sp.	Wu *et al*, 2017[81]
OAB		Mean 27,5	30				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
Asymptomatic	Female	Mean 59	22	MSU *4 times, each 4 weeks	SUC, urothelial-cell sediment culture	- significant differences in microbiome, pyuria counts, and infected urothelial cells were observed in both groups - *Staphylococcus* sp., *Streptococcus* sp., *Citrobacter* sp. and *Lactobacillus* sp. were most dominant genera in asymptomatic cohort - recognized uropathogens were more often detected in OAB i.e., *E. coli*, *E. faecalis*, *Proteus*, *Klebsiella-Enterobacter-Serratia* group - SUC did not allow to distinguish samples from asymptomatic and OAB cohort	Gill *et al*, 2018[168]
OAB		Mean 63	23				
Asymptomatic	Female	NA	10	TUC	metagenomics	- twelve recovered metagenomes contained sequences of phages and/or eukaryotic viruses - contigs homologous to putative prophage of *Gardnerella vaginalis*, *Streptococcus agalactiae* or some *Lactobacillus* sp. were identified but also novel sequences with no homology in GenBank database - human polyomavirus JC was identified only in OAB cohort	Garretto *et al*, 2018[123]
OAB			20				
OAB	Female	62-73	41	TUC, MSU	EQUC	- 12 weeks on estrogen therapy did not cause decrease in diversity of urinary tract microbiome - after estrogen treatment significant increase in *Lactobacillus* was observed in urine collected by TUC, but not in MSU - increase in *Lactobacillus* was also correlated with modest improvement of urgency incontinence symptoms severity - hormonal therapy also contributed to increase in urinary antimicrobial peptides which correlated with higher abundance of *Corynebacterium*	Thomas-White *et al*, 2020[169]
INTERSTITIAL CYSTITIS/PAINFUL BLADDER SYNDROME							
IC/PBS	Female	27-67	8	MSU	Sequencing V1-V2 and V6 regions of 16S rRNA gene	- urinary tract microbiome composition of IC patients differs from asymptomatic controls - 93% of detected microbes belonged to Firmicutes phylum	Siddiqui *et al*, 2012[170]

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- *Lactobacillus* sp. was significantly enriched in IC patients, comparing to asymptomatic women - *Enterococcus* sp., *Atopobium* sp., *Proteus* sp. and *Cronobacter* sp. were found only in IC patients - microbiome of IC patients showed lower diversity and richness	
IC/BPS	Female	Mean 41	213	ISU, MSU	SUC, culture-independent using 16 primer pair BAC/Fungal detection system	- more bacterial species were detected in ISU (81) than in MSU (73) samples - evaluation of patients with and without symptom flare showed the higher prevalence of fungi (*Candida* sp. and *Saccharomyces* sp.) in patients who reported a flare - no significant difference of bacterial species/genus occurrence and uropathogens detection was observed between flare and non-flare cohorts	Nickel *et al*, 2016[172]
non-IC/PBS	Female	Median 34 (SD=10)	18	MSU, VS	Sequencing V3-V4 region of 16S rRNA gene	- *Lactobacillus* was the most dominant taxa in both cohorts - no significant compositional differences of urogenital microbiome between non-IC/PBS and IC/PBS cohorts	Meriwether *et al*, 2019[173]
IC/PBS		Median 34 (SD=8)	23				
non-IC/PBS	Female	in groups: <35, 35-50, 50+	182	MSU	culture-independent - using 16 primer pair BAC/Fungal detection system and Plex-ID molecular diagnostic platform	- *Lactobacillus gasseri* and *Corynebacterium* were the species responsible for most compositional variations (e.g., *L. gasseri* was more prevalent in IC/PBS cohort) - 29 species were identified only in IC/PBS cohort (e.g., *Proteus mirabilis, Pseudomonas aeruginosa, Mycoplasma hyorhinis, Helicobacter hepaticus*) - there was no statistically significant difference between microbiome composition (species, nor genus level) of non-IC/PBS and IC/PBS cohort	Nickel *et al*, 2019[226]
IC/PBS			181				
Asymptomatic	Female	Mean 48 (SD=12)	20	MSU	EQUC and sequencing V4 region of 16S rRNA gene	- urotypes from both cohorts did not differ - alpha diversity, principal component analysis and hierarchical clustering found no statistical significance between cohorts	Bresler *et al*, 2019[144]
IC/PBS		Mean 50 (SD=13)	21				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- putative uropathogenic species were detected in both cohorts, with no statistical significance - IC/PBS may not be related to bacterial members of urinary tract microbiome	
IC/BPS	Female	Mean 41.25	202	MSU *Sampling repeated after 6 and 12 months	culture-independent using 16 primer pair BAC/Fungal detection system	- the fungi were detected in 92.4% samples tested - fungi-positive urinary tract microbiome demonstrated 8 or 45 different fungal genera (depending on analysis) - there was no association between fungal taxa detected and flare status or pain severity, however fungi were more likely detected in patients with severe symptoms - there might be an association of increasing RA of *Candida* and *Malassezia* in patients with higher symptoms severity	Nickel *et al*, 2020[174]
non-IC/PBS	Female	Mean 51 (SD=15)	40	TUC, MSU (n=10)	EQUC, sequencing V4 region of 16S rRNA gene (n=66)	- more than half of the IC/PBS samples did not grow live bacteria by EQUC which could question microbiome involvement in this pathology - distribution of the urotypes was different between non-IC/PBS and IC/PBS cohort (e.g., *Streptococcus* urotype was observed only in IC/PBS group) - *Lactobacillus* was the most common urotype (and *L. iners* was the most common species) in both cohorts	Jacobs *et al*, 2021[83]
IC/PBS		Mean 45 (SD=13)	49				
LOWER URINARY TRACT SYMPTOMS							
Asymptomatic	Female	Mean 43 (SD=17)	26	MSU, TUC	SUC, Culture of urinary sediment, sequencing of 16S rRNA gene	- LUTS patients demonstrated polymicrobial colonization of the urothelium (quantitatively different from controls) with many uropathogens being likely intracellular or close to urothelium (e.g., only in LUTS cohort *E. coli* was closely associated with cells) - *E. coli*, *Enterococcus* sp., *Staphylococcus* sp., *Streptococcus* sp. were found in both asymptomatic and LUTS cohort - *Proteus* sp., *Micrococcus* sp. were specific to LUTS cohort - *Actinomyces* sp., *Prevotella* sp. were specific to asymptomatic cohort	Khasriya *et al*, 2013[40]
LUTS		Mean 48 (SD=16)	55				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- *Lactobacillus* sp. were more often detected from TUC than from MSU, which supports previous findings on these bacteria residing urinary tract - the initial cohort was further enlarged with MSU samples of 47 asymptomatic controls and 165 patients, including women and men, which confirmed the previous findings	
Asymptomatic LUTS	NA	NA	previously published	MSU or TUC	whole genome sequencing	- examining 181 genomes from previous female urinary tract microbiome studies (asymptomatic and LUTS cohort) - at least one phage was found in 86% of examined genomes - 457 phage sequences were found, with 226 predicted with high confidence and many of which were novel with no homology with known sequences deposited in public databases - *Lactobacillus* strains had frequently more than one phage - novel prophages were found in species of e.g., *Actinomyces, Varibaculum, Bifidobacterium, Gardnerella, Streptococcus* - there was a difference in phages abundance observed in microbiome of asymptomatic and OAB cohorts	Miller-Ensminger *et al*, 2018[227]
Asymptomatic Symptomatic/ LUTS	Female	NA	77, all previously published	TUC	whole genome sequencing, metagenomics (n=12)	- examining 149 genomes from previous female urinary tract microbiome studies - highly similar strains were isolated from urinary tract and vagina - urinary and vaginal microbiome appeared to be interconnected - interlink applied to putative pathogenic species (e.g., *Escherichia coli, Streptococcus anginosus*) as well as health-associated commensals (e.g., *Lactobacillus crispatus, Lactobacillus iners*)	Thomas-White *et al*, 2018[39]
Asymptomatic and LUTS	Female	21-85 (median 55)	49	TUC, MSU, PUS, US	EQUC	- microbiome obtained by TUC was different from PUS, US and MSU - urethral microbiome was different from MSU - the same genera were grown from all sample types, however in different distributions - prevalence of *Lactobacillus* and *Gardnerella* was similar in all sample types - *Staphylococcus* and *Escherichia* were more prevalent in MSU	Chen *et al*, 2020[41]

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- *Escherichia* was more abundant and *Lactobacillus* was less abundant in TUC, while the *Corynebacterium* and *Streptococcus* was more abundant in other types of samples - younger women were more prone to have *Lactobacillus* and *Gardnerella* in urethral samples - *Corynebacterium* and *Streptococcus* were more abundant in all samples but TUC - no association was found for LUTS and abundance and/or prevalence of urethral bacteria - the microbial load in urogenital samples increases with more distal location	
LUTS/BPE	Male	40-85	49	MSU, TUC	EQUC, sequencing of 16S rRNA gene	- microbiome was detected in 98% of MSU, and 39% of urine collected by TUC by both methodologies (EQUC and sequencing) - microbiome was detected most often in men with severe LUTS, and less often from men with mild LUTS - microbiome of MSU and urine collected by TUC differed significantly	Bajic *et al*, 2020[228]
SYMPTOMATIC POPULATION AND OTHER CONDITIONS							
Asymptomatic and symptomatic	Female	NA	92	MSU	SUC, Sequencing V1-V3 regions of 16S rRNA gene (n=116), metagenomics (n=49)	- Bacteria were dominating microbes in urinary tract microbiome (94.6%), followed by eukarya (0.05%), viruses (0.0027%) and trace of archea (0.0001%) - contamination of reads with human-derived data was in a range 1.3 - 99.9% - although shotgun metagenomics and 16S rRNA gene profiling gave similar results, some bacterial detection (e.g., *Gardnerella*) was improved with metagenomics approach - the median of 2 (range 1-8) for eukaryotic species per sample was identified (e.g., *Candida albicans, Candida glabrata*) - the median of 3 (range 1-9) viruses per sample were identified (phages and human viruses)	Moustafa *et al*, 2018[76]
	Male		29				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- the biggest differences between female and male microbiome were related to high numbers of reads for *Lactobacillus* and *Prevotella* in women, and *Enterococcus* and *Pseudomonas* in men	
Prepubertal boys who required elective urologic procedures	Male	3 months - 8 years (median 15 months)	20	TUC (n=10) and rectal swab (n=20)	Sequencing V6 region of 16S rRNA gene	- urinary tract microbiome in prepubertal boys contained several genera reported in microbiome of adults - the cohort was divided into boys without (n=5) and with (n=5) prior antibiotic treatment - the higher bacterial load was observed in group previously exposed to antibiotics - alpha diversity measures did not differ in both cohorts	Kassiri *et al*, 2019[154]
NA *volunteers had conditions that were not known to involve infections	Female	22-48 (median 31.6)	147	MSU (n=137), TUC (n=10)	Sequencing V4-V5 of 16S rRNA gene, culturmics, quantitative real-time PCR	- Chinese reproductive-age population was studied and *Streptococcus* and *Lactobacillus* was the most dominant bacteria in urinary tract microbiome - additional six locations had been sampled within female reproductive tract - vaginal samples were mostly dominated by one taxa, while urinary tract microbiome showed higher diversity - urinary tract microbiome was shown to be more similar to cervix and uterine cavity, than to vaginal microbiome in the same women - the bacteria were also identified from urine collected by TUC (3/10) and included *Lactobacillus, Staphylococcus, Clostridium, Enterococcus* and *Propionibacterium*	Chen *et al*, 2020[229]
Asymptomatic	Female	19-62 (median 29)	49	MSU, vaginal fluid	Sequencing V1-V2 regions of 16S rRNA gene	- it was not possible to differentiate healthy female and male microbiome by clustering based on bacterial composition - female samples contained bacteria not detected in males, while male samples showed the largest diversity - unique urotype present only in healthy female was characterized by *Lactobacillus crispatus* - there was no urotype characteristic only for male	Gottschick *et al*, 2017[143]
	Male	23-59 (median 29)	31				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
BV	Female	18-51 (median 31)	109			- female urinary tract microbiome differed in healthy women and women with BV	
UTI-RELATED SYMPTOMS							
Symptomatic (suspected cystitis)	Female	18-49 (median 22)	202	MSU, TUC	culture-based	- *Escherichia coli* was the major pathogen identified (65%), with slightly higher detection from MSU (131 episodes) than from TUC (120 episodes) - strong correlation was observed for colony counts from MSU and TUC regarding *E. coli* cases, *Klebsiella pneumoniae* (n=10 cases) and *Staphylococcus saprophyticus* (n=6) - *Enterococcus* was better detected from MSU (20 episodes) than TUC (2 episodes). The same was observed for *Streptococcus* group B with 25 cases detected from MSU and just 2 from TUC - in 61% of episodes where *Enterococcus* and/or *Streptococcus* was detected, *E. coli* was also identified - additionaly some *Lactobacillus* and *Gardnerella vaginalis* were isolated from TUC	Hooton *et al*, 2013[177]
urethritis	Male	NA	38	MSU	Sequencing of 16S rRNA gene, PCR for selected species	- most frequently identified were *Propionibacterium, Staphylococcus, Corynebacterium, Streptococcus, Pseudomonas, Gardnerella vaginalis, Anaerococcus, Peptoniphilus, Finegoldia, Bacteroides* - the detection of *Neisseria gonorrhoeae, Chlamydia trachomatis, Mycoplasma genitalium* was less efficient by 16S rRNA gene sequencing than by traditional PCR - *Neisseria gonorrhoeae, Chlamydia trachomatis, Mycoplasma genitalium* were often identified as predominant bacterial species in the sample	You *et al*, 2016[180]
non-idiopathic urethritis	Male	19-51 (median 28)	46	MSU, US	Sequencing V3-V4 regions of 16S rRNA gene	- *Lactobacillus* (50% of the samples) and unclassified *Alphaproteobacterium* (71%) were highly present in non-idiopathic urethritis cohort - the microbiome composition of both cohorts could not be properly distinguished by hierarchical clustering	Frølund *et al*., 2018[179]
idiopathic urethritis			39				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- several patients had microbiome dominated with single genus of putative relevance e.g., *Gardnerella, Haemophilus, Ureaplasma*	
cystitis	Female	Mean 54.10 (SD=12.69)	42	TUC	SUC, Sequencing V4 region of 16S rRNA gene	- the microbiome composition of acute uncomplicated cystitis (AUC) and recurrent cystitis (RC) cohort was significantly different (e.g., *Pseudomonas, Acinetobacter*, Enterobacteriaceae in AUC cohort; *Sphingomonas, Staphylococcus, Streptococcus, Rothia* in RC cohort) - RC cohort was characterized by higher microbiome diversity - culture-independent methodology significantly improved bacterial detection, comparing to SUC, and may be useful in the clinical diagnostics	Yoo *et al*, 2021[181]
Symptomatic (suspected UTI)	NA	NA	35	MSU	SUC, whole genome sequencing, metagenomics	- metagenomics allows to identify additional bacteria of putative relevance present in the samples and missed by SUC (e.g., *Aerococcus urinae*, in addition to *Proteus* and *E. coli*) - detection of certain species that might be involved in UTI was only possible with metagenomics (e.g., *Lactobacillus iners, Gardnerella vaginalis*, *Prevotella* sp., and *A. urinae*) - metagenomics may be used for clinical purposes as a fast and more accurate approach comparing to culture	Hasman *et al*, 2014[122]
non-UTI	Female	25-54	5	NA	viral metagenomics and sequencing V3 region of 16S rRNA gene	- bacteriophages were the most common viruses identified - both cohorts had herpesviruses, polyomaviruses, and human papillomaviruses (HPV) - HPV seemed to be common in urinary tract microbiome - alpha and beta diversity for viral community did not differ between non-UTI and UTI cohort, however it significantly differed for bacterial communities - no significant differences in viral and bacterial composition were observed according to gender	Santiago-Rodriguez *et al*, 2015[124]
	Male	47-94	5				
UTI	Female	18-69	5				
	Male	52-78	5				
non-UTI	Female	Mean 60.6 (SD=12.3)	75	TUC		- greater microbiome diversity was observed in non-UTI cohort	Price *et al*, 2016[139]

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
UTI		Mean 64 (SD=17.1)	75		SUC, EQUC, streamlined EQUC	- *Streptococcus* (*S. mitis/oralis/pneumoniae, S. parasanguinis, S. salivarius, S. sanguinis*) and *Gardnerella vaginalis* were detected more often in non-UTI cohort, while *Escherichia coli* was more prevalent in UTI cohort - UTI cohort had uropathogens in the higher load than non-UTI - streamlined EQUC (simplified EQUC protocol) can greatly improve detection of clinically relevant bacteria comparing to SUC	
UTI/urosepsis	NA	NA	10	NA	metagenomics	- additional urine samples from asymptomatic controls but heavily inoculated with known *Escherichia coli* strain were tested - metagenomics performed with MinION allowed for the correct identification of pathogens directly from urine and showed high accuracy with detection of acquired resistance genes (51/55) - gene mutations and allelic variants could not be reliably identified by this method - metagenomics would be attractive alternative for diagnostic purposes with few hours to obtain results	Schmidt *et al*, 2017[125]
UTI-related sepsis	NA	NA	40	NA	SUC, metagenomics	- metagenomics was shown to accurately detect microbiome and resistome from urine samples - *Escherichia coli* was the most pathogenic agent identified - detection of *Actinotignum schaalii* in one patient was only possible with metagenomics - metagenomics could be a suitable and time-saving tool in clinical diagnostics	Barraud *et al*, 2019[126]
non-UTI and UTI	Female	< 48 months	59	TUC	Sequencing V3-V4 regions of 16S rRNA gene	- urinary tract microbiome was identified in every children, even in neonates - children with UTI had lower alpha diversity (than non-UTI) - microbiome of children with UTI clustered separately from those without UTI - alpha diversity decreased in group who had been given antibiotics within 2 weeks of the urine sample collection - the significant beta-diversity differences were found between antibiotic use, but not between age, gender, maternal ethnicity, country of origin, delivery mode, probiotic use - 5 most abundant genera were *Prevotella, Peptoniphilus, Escherichia, Veillonella, Finegoldia*	Kinneman *et al*, 2020[153]
	Male		26				

Status of urine donors	Gender	Age (years)	Number of donors	Sample collection method	Methodology used	Main findings	Reference
						- non-UTI samples had heterogenous patterns of the most abundant taxa - the phylum *Actinobacteria* was more commonly found in boys - a strong correlation was detected between age and abundance of *Mobiluncus* (mainly driven by female group)	
rUTI	Female	Mean 67 (SD=14.2)	43	MSU, TUC	SUC, EQUC	- EQUC detected more uropathogens and more unique bacterial species than SUC - *E. coli* and *E. faecalis* were most commonly detected in both sample types - emerging uropathogens were detected only by EQUC e.g., *Actinotignum schaalii, Streptococcus anginosus* - comparing to TUC, voided urine had high false-positive rates	Hochstedler *et al*, 2021[182]

NA - not available, RA - relative abundance, OTU - operational taxonomic unit, MSU - midstream urine, ISU - initial stream urine, TUC - transurethral catheter, SPA - suprapubic bladder aspiration, SPC - suprapubic catheter, US - urethral swab, PUS - peri-urethral swab, VS - vaginal swab, RVS - rectovaginal swab, PS - perineal swab, TUR - transurethral resection, CS - coronal sulcus, IC/PBS - interstitial cystitis/painful bladder syndrome, OAB - overactive bladder, UI - urinary incontinence, UUI - urge urinary incontinence, SUI - stress urinary incontinence, MUI - mixed urinary incontinence, POP - pelvic organ prolapse, LUTS - lower urinary tract symptoms, BPE - benign prostate enlargement, NB - neuropathic bladder, CP/CPPS - chronic prostatitis/chronic pelvic pain syndrome, UTI - urinary tract infection, rUTI - recurrent urinary tract infection, BV - bacterial vaginosis, T2DM - type 2 diabetes mellitus, EQUC - enhanced quantitative urine culture, EUC - expanded urine culture, SUC - standard urine culture , SD - standard deviation.

1.4. Properties of certain bacterial groups of urinary tract microbiome

Despite continuous efforts, the 'core' urinary tract microbiome has not yet been established. As suggested for other niches, the existence of 'core set of genes' reflecting urinary tract microbiome functionality is more likely than simply a set of taxonomically identical profiles[15]. Nevertheless, several bacterial groups have been consistently reported by the urinary tract microbiome studies, suggesting that they are more likely to inhabit UT and may have specific role in UGT health and disease. To date, prevalent and/or abundant members are *Lactobacillaceae*, *Gardnerella* sp., Enterobacterales, *Streptococcus* sp., *Staphylococcus* sp., *Corynebacterium* sp. and *Prevotella* sp.[27,42,75,133,145,148].

1.4.1. *Lactobacillaceae*

Lactobacillaceae family (phylum Firmicutes) comprises ubiquitous Gram-positive, coccoid or rod-shaped bacteria, that are isolated from various sources (e.g., humans, animals, food products)[230,231]. Genera included in *Lactobacillaceae* belong to the heterogenous group of lactic acid bacteria (LAB) which is characterized by production of lactic acid as the main fermentation end-product[232]. LAB present capsular exopolysaccharides that can modulate adhesion, microbe-microbe or host-microbe interaction[233–236].
Its biggest genus was *Lactobacillus*, however, it comprised species with significant genomic diversity, e.g., variable genome size and G+C content, thus, there was a need to re-evaluate its taxonomic classification[237]. Indeed, in early 2020, the *Lactobacillus* taxonomy has been revisited based on a polyphasic approach, which resulted in its reclassification to 25 genera[231].

Among those 'novel' genera, three seem to be of particular interest regarding human microbiome i.e., *Lactobacillus*, *Limosilactobacillus* gen. nov. (former *Lactobacillus reuteri* group) and *Lacticaseibacillus* gen. nov. (former *Lactobacillus casei* group)[231]. They are widely isolated from vagina, gut or mouth cavity as part of human microbiome[238–241]. Moreover, many of their species are Generally Recognized As Safe (GRAS), and are under constant investigation due to their diverse useful properties and have been used for years as probiotics or in the food industry[242–245]. Interestingly, they also represent the most prevalent bacterial group isolated from healthy human urinary tract microbiome[39,74,75,133,145].

Lactobacillus sp., *Limosilactobacillus* sp. and *Lacticaseibacillus* sp. use carbon as an electron acceptor instead of oxygen[232] and are commonly found in nearly anaerobic conditions (e.g., vagina). Apparently, due to relatively low oxygen concentration in the bladder, varying from 4 to 5.5%, also UT is an advantageous environment for these species[28,246]. They produce lactic acid that may be present in two isomeric forms, D- and L-lactic acid and it is fundamental for *Lactobacillaceae* beneficial activity (**Box 5**). The vaginal epithelial cells are only able to produce the L-form of lactic acid under hormonal stimulation[238]. However, it is the D-chiral isomer, only produced by microorganisms, that represents the lactic acid form more active in various metabolic pathways, such as down-regulation of matrix metalloproteinase-8 (MMP-8), histone acetylation, damaged DNA reparation or gene transcription[238]. Some specific strains can also produce bacteriocins – low molecular mass peptides that have antimicrobial activity, thus with protective features[232].

Box 4 – Lactic acid role in vaginal environment

Healthy vaginal microbiome is usually abundant in *Lactobacillus* sp. and, in a smaller extent, in *Limosilactobacillus* sp. and *Lacticaseibacillus* sp.[247]. The end product of their glycolysis - lactic acid, is an organic acid that ensures maintenance of low vaginal pH, which together with high glycogen deposit promotes an environment convenient for *Lactobacillaceae* proliferation and creates difficulties in opportunistic pathogens colonization[238]. Dominance of *Lactobacillus* sp. in this body site presents a protective activity not only by lactic acid production, but also hydrogen peroxide, strain-specific bacteriocin production or competition for the epithelium attachment[23].

Lactic acid also contributes to various health-beneficial actions such as antimicrobial defense or immunomodulatory activities. Besides the differences between both isomeric forms of lactic acid where D-isomer presents higher protective activity and stability for vaginal environment than L-isomer, its efficacy is significantly dependent on dissociated state of this organic compound, as it influences the ability to cross plasma membrane and act inside the targeted cell[23]. Uncharged or protonated form of lactic acid present in acidic pH (pH < 3.8) can easily permeate membrane and by cytosol acidification cause lysis of other bacterial cells, comparing to lactate anion which requires transporters support. Protonated form of both isomers contributes also to immunological action, for instance, by increasing production of anti-inflammatory cytokines or inhibiting production of inflammatory cytokines[23,248].

The most prevalent *Lactobacillus* species detected in healthy vagina and UT so far are *Lactobacillus crispatus, Lactobacillus gasseri, Lactobacillus jensenii*, and *Lactobacillus iners*[39,75,83,133,142,223,238,249]. Among them, only *L. iners* does not possess ability to produce D-

lactic acid[238] but produce inerolysin – a toxin similar to cytolysin produced by *Gardnerella vaginalis* – main pathogen involved in BV[250,251]. Unsurprisingly, although also present in healthy UGT, *L. iners* seems to be associated with vaginal dysbiosis and has been previously reported as opportunistic pathogen associated with UTI development[252–254]. Occasionally, *Lactobacillus delbrueckii*, *Lactobacillus johnsonii* and *Lactobacillus paragasser*i are detected in urinary tract microbiome[39,75,133,255].

L. crispatus is the most prevalent *Lactobacillus* in the microbiome of UGT, and it contributes to healthy state of female genital tract e.g., seems to promote autophagy in vaginal epithelial cells, in the higher level than other *Lactobacillus* species[256]. In *L. crispatus* ST1, expression of LEA-1 – strain-specific *Lactobacillus* epithelium adhesins – provides efficient binding to epithelial cells, for instance, of human vagina[257]. Other proteins involved in epithelium colonization, for instance in gut, are surface layer proteins (S-layer proteins, SLP), relatively small surface molecules present in many *Lactobacillus* species[258–260]. Specific strains of *L. crispatus* possess genes coding for CbsA or SlpB which are S-layer proteins binding to the collagens (type I and IV) in intestinal mucosa which is also the main component of extracellular matrix (ECM) of bladder basement cells[34,259,261,262]. Moreover, certain *L. crispatus* strains compete for the adhesion to the vaginal epithelial cells with the species *G. vaginalis,* frequently associated with BV, by overlapping with the proteins' domains associated with hyaluronate or fibronectin-bindings[245,263]. Furthermore, some *L. crispatus* and *L. jensenii* strains showed inhibitory effect against *Escherichia coli* and various *L. crispatus* and *L. gasseri* demonstrated activity against *Streptococcus agalactiae*[232,264–267]. Additional strain-specific *Lactobacillus* activity is production of GABA, inhibitory neurotransmitter that regulates many CNS functions. Certain gut *Lactobacillus* in the presence of monosodium glutamate can secrete this substance which can potentially benefit brain function[268–271].

Species belonging to *Limosilactobacillus* gen. nov. are also found in urinary tract microbiome, however less often and less abundant than *Lactobacillus* sp.. Most reported are *Limosilactobacillus fermentum*, *Limosilactobacillus reuteri*, *Limosilactobacillus vaginalis* and *Limosilactobacillus mucosae*[39,75,133,247,272,273]. Interestingly, in microbiome, they often seem to be present in combination with *Lactobacillus* sp.[39,274]. Various surface layer proteins have been reported also for those species, allowing adhesion to collagen – component of bladder submucosa ECM e.g., Cnb protein of *L. reuteri* NCIB 11951 or CBP from *L. fermentum* 3872[34,245,261,262]. Furthermore, certain *L. reuteri* strains may have a role in bacterial auto-

aggregation (*L. reuteri* ATCC 53608) or immunomodulatory activity (*L. reuteri* ATCC PTA6475)[261]. Although available insights are from gut, *L. reuteri* 17938 or ATCC 23272 seems to reduce inflammation by e.g., reduction of proinflammatory cytokines levels[275,276]. *L. fermentum* however has demonstrated beneficial activity in genital tract e.g., strain LF15 can inhibit *G. vaginalis* and improve Nugent score – a classification used to estimate vaginal health based on dominating bacterial morphotypes[277].

Finally, *Lacticaseibacillus* gen. nov. is occasionally identified among urinary tract microbiome, especially *Lacticaseibacillus rhamnosus* and *Lacticaseibacillus paracasei*[39,75,133]. Similarly to other *Lactobacillaceae,* species of *Lacticaseibacillus* showed adhesion abilities to ECM proteins e.g., collagen IV and fibronectin, that are likely involved in host UGT colonization[278,279]. Specific *Lacticaseibacillus* strains are used as probiotics and they are promising bacteriocin producers e.g., lactocin 160 produced by *L. rhamnosus* vaginal strain[280]. They also present potent outcompeting skills, for instance, *L. rhamnosus* 160 inhibits *G. vaginalis*[234,280], while *L. rhamnosus* HN001 showed antagonistic activity also against e.g., *E. coli*, *Staphylococcus aureus* and *Streptococcus agalactiae* that may be involved in aerobic vaginitis (AV)[281]. Furthermore, *L. rhamnosus* GG was shown to have inhibitory activity against *Candida albicans*, by production of lactic acid and hydrolase Msp1 (major secreted protein 1)[234,282]. Strains of both species were also found to have antibacterial activity against carbapenem-resistant *E. coli* and *Klebsiella pneumoniae*[283]. It was also demonstrated that strain-specific EPS-SJ produced by *L. paracasei subsp. paracasei* BGSJ2-8 has ability to reduce *E. coli* association with Caco-2 human cell line[233].

Interestingly, a mix of *L. rhamnosus* GR-1 and *L. reuteri* RC-14 or *L. rhamnosus* DSM 14870 and *L. gasseri* DSM 14869 are common probiotic combinations which may improve vaginal health[284–286]. Although potential beneficial effect for vaginal health of these probiotics was demonstrated with oral and vaginal administration, oral capsules might not be the effective way to improve UT health[284]. To date, one double-blinded randomized controlled trial was performed to observe effect of oral administration of GR-1 and RC-14 probiotic on ratio between *Lactobacillus* and uropathogens in UT[151]. This preliminary trail on asymptomatic pre-menopausal women showed no effect on *Lactobacillus*/uropathogens ratio and absence of strains delivered with probiotic in participants' voided urine[151]. Nevertheless, further research is needed to validate the usefulness of oral or vaginal probiotics on UT health improvement.

Considering the *Lactobacillaceae* features in the other body niches mentioned above, these bacteria express several mechanisms that could facilitate colonization of the bladder and provide protective activity on the UT, as interact with components of bladder ECM such as collagens, hyaluronic acid or fibronectin[34]. Possibly, the assumption that the decrease of protective *Lactobacillaceae* in microbiome leads to health complications observed in vagina, can be similarly applied to UT, as several studies reported lower detection of these species from people suffering from UT disorders[74,81]. Referring to the UT, changes in prevalence and/or abundance of *Lactobacillaeae* are not understood yet, however, it is suggested that proper level of specific strains inhibit overgrowth of Gram-negative bacteria and contributes to UT health maintenance.

Noteworthy, most studies investigating human urinary tract microbiome apply identification methods insufficient for proper species level *Lactobacillaceae* taxonomic classification (MALDI-TOF MS, 16S rRNA amplicon sequencing). Thus, it is expected, that presence and diversity of *Lactobacillaceae* in urinary tract microbiome is highly underrated.

1.4.2. *Gardnerella*

Genus *Gardnerella* belongs to Actinobacteria phylum and was previously associated with only one species i.e., *Gardnerella vaginalis*. However, it was reported since several years that within available *G. vaginalis* collections there is significant diversity of genomic features suggesting taxonomical subgroups, likely associated with different pathogenic potential, since it might be present both in healthy and diseased women[287,288]. Recently, the taxonomy of *Gardnerella* has been revisited which resulted in description of 3 novel species in addition to the 'real' *G. vaginalis* (*Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov.) and several putative novel genomospecies[287,289].

This taxonomic update is truly relevant for understanding health of urogenital tract, and raises basic questions related to urogenital *Gardnerella* diversity and its contribution to disease state, particularly BV[287,289,290]. For instance, newly described species differ in ability to produce sialidase, which translates into host-microbe interaction and could be an important feature contributing to colonization capacity. However, initial research on *Gardnerella* sp. did not show association between sialidase production and BV symptoms[287,289].

Importantly, most strains of *Gardnerella* encodes pore-forming toxin called vaginolysin which together with inerolysin produced by *L. iners* belong to cholesterol-dependent cytolysins which are capable of damaging cell membranes and are found in some Gram-positive bacteria causing diseases at mucosal surfaces[250,291,292]
Another potent feature of *Gardnerella* sp. is the ability to form biofilm which is a complex structure of bacterial community surrounded by self-produced extracellular matrix[293]. *Gardnerella* biofilm promotes bacterial persistence, resistance to hydrogen peroxide and lactic acid, and may express toxic activity towards vaginal epithelium[294,295]. It is often suggested that *Gardnerella* promotes early colonization and biofilm formation in BV, which facilitates further attachment of other, secondary BV-associated species[289,293,296,297]. Indeed, *Gardnerella*-associated biofilms are often polymicrobial which unveil important microbe-microbe interactions with bacterial species commonly detected in vagina and UT (e.g., *Enterococcus faecalis*, *Actinomyces neuii*)[296,297]. It was also demonstrated that in dual-species biofilm, *Gardnerella* transcriptome may change and thus, to influence its virulence. For instance, dual-species biofilm with *E. faecalis* or *A. neuii* may stimulate increased expression of genes coding for vaginolysin or sialidase. Some species may also cause contrary effect, thus reducing virulence (e.g., reduced expression of gene coding for sialidase in biofilm with *S. anginosus*)[297].

Noteworthy, several independent studies demonstrated that *Gardnerella* sp. was widely present in the kidneys of hospitalized patient, thus *Gardnerella* is occasionally referred to as 'underestimated' putative uropathogen[298]. Indeed, there is some important research done on *Gardnerella* contribution to UTI pathogenesis[299,300]. For instance, in animal model, exposure to '*G. vaginalis*' caused bladder epithelial exfoliation and kidney injury, which activated intercellular reservoirs of *E. coli* and caused severe rUTI[300]. Interestingly, the initial damage caused by *G. vaginalis* remained even after *G. vaginalis* was cleared from UT which created the paradigm called "covert pathogenesis"[300,301]. Moreover, it was demonstrated that *Gardnerella* strain used had capacity of causing kidney damage itself, even without *E. coli* contribution[299].

According to current understanding of prevalence of different *Gardnerella* species, all novel species were found in vaginal microbiome, however with different frequency and abundance[289]. Curiously, *G. vaginalis* and *G. swidsinskii* were observed in higher relative abundance among women with vaginal discharge and abnormal odor[289].

Although '*G. vaginalis*' had been widely reported from healthy and diseased UT[41,42,82,148,223,224], until now very few studies have explored genomic differences supporting the existence of distinct *Gardnerella* species and their occurrence in the UT[302,303]. The initial whole genome sequencing of urinary *Gardnerella* isolates indeed revealed distinguishable clades based on core genome phylogeny and existence of numerous prophages in these genomes[303].

Unfortunately, current methodological approaches widely used in urinary tract microbiome studies are limiting proper *Gardnerella* identification. Although it was shown to be possible to distinguish novel species by MALDI-TOF MS[287], their reference spectra are not yet included in the databases. Furthermore, high 16S rRNA gene similarity does not allow to identify them at species level[289], thus 16S rRNA amplicon sequencing cannot resolve this limitation. Alternatively, usage of other, more specific marker gene e.g., *cpn60* for targeted amplicon sequencing could be a solution[289]. From the available methods for microbiome profiling, shotgun metagenomic sequencing can provide high taxonomic resolution data for species and strain-level identification and hopefully will be applied more often in near future to study urinary tract microbiome.

1.4.3. Enterobacterales

Proteobacteria is another phylum usually reported from healthy UT, mostly as a minor contributor. It contains many bacteria often classified as pathogens and is the most explored due to the order Enterobacterales. It comprises bacterial families that include Gram-negative species relevant to UT health e.g., *Escherichia coli*, *Klebsiella pneumoniae*, *Citrobacter koseri*, *Enterobacter cloacae* (all belonging to family *Enterobacteriaceae*) and *Proteus mirabilis* (from *Morganellaceae* fam. nov.)[30,304]. They are common uropathogens however most of them are not or are rarely (e.g., *K. pneumoniae*[133]) reported from healthy urinary tract microbiome. *E. coli* however is often present in healthy urinary tract microbiome and can be there as scarce community member or less often, as dominating species[27,40,41,133,139,145]. Interestingly, of all recognized uropathogens, *E. coli* is the most frequent cause of UTI (>80%) worldwide[305,306].
Thus, with discovery of urinary tract microbiome and initial knowledge on *E. coli* prevalence in healthy population, the reservoirs of *E. coli* causing infection need to be reevaluated.

Although the behavior of *E. coli* in healthy UT is not clear yet[307], it is established that *E. coli* can grow well and fast in urine and UT[308–310].
A variety of *E. coli* genetic determinant are known to contribute to colonization and survival in UT. *E. coli* strains with the ability to cause UTI are referred as uropathogenic *E. coli* (UPEC)[310]. Majority of the UPEC strains belong to globally spread lineages causing UTI worldwide. The most predominant lineages include e.g., sequence type (ST) 131, ST95, ST73, ST69, ST127 or ST12, they are often antibiotics resistant and they are enriched in virulence-associated traits[311–314]. For instance, the production of adhesins (e.g., *fimH* – type 1 fimbriae, *papG* – P fimbriae, *csgA* – curli) by many *E. coli* strains promote bacterial survival and attachment to urothelium[30,306,315,316]. Many strains are also able to secrete toxins (e.g., α-hemolysin, HlyA; cytotoxic necrotizing factor 1, CNF1) which were linked to pathogenesis of UTI and were shown to be involved in cell invasion, host cells damage and inflammation induction[310,314,316].

To date, there are established *E. coli* reservoirs within the human body i.e., gut and vagina, however, *E. coli* is also capable to establish intracellular reservoirs that play a role in pathogenesis of rUTI. Several types are proposed by now i.e., intracellular bacterial communities (IBCs) which are replicating biofilm-like structures; quiescent intracellular reservoirs (QIRs) – small structures composed of few bacterial cells which are not replicating; and recently proposed vaginal intracellular communities (VICs) which are reservoirs inside vaginal epithelial cells[305,317]. Moreover, its ability to create biofilms additionally contribute to its persistence within inhabiting niche[246,318].

Noteworthy, whether infection is established or not (and its outcome) may also depend on interactions with other bacteria, for instance, the previously mentioned example of *E. coli* and *G. vaginalis* contribution to infection development and severity[299,300]. Additionally, interactions with the host cells will also have an important impact. It was demonstrated that bladder tissue pathophysiology changes after *E. coli* infection and modulates its susceptibility for the future recurrent UTIs[319]. Nevertheless, whether just colonization with *E. coli* may also stimulate changes in epithelium, gene expression and further host-microbe interactions remains to be evaluated.

1.4.4. Other bacterial groups of urinary tract microbiome

Streptococcus sp., *Staphylococcus* sp., *Corynebacterium* sp. (Gram-positive) and *Prevotella* sp. (Gram-negative) are other bacteria often reported from healthy urinary tract microbiome.

Streptococcus sp. (phylum Firmicutes) also belong to LAB and dominate in oral cavity. These Gram-positive, catalase-negative cocci constitute around 35% of oral bacteria from healthy controls but are also detected from nasal cavity, gut, and vagina, however in lower abundance[5,320]. Due to variable expression of several adhesins or structures facilitating adhesion to the human cells or to ECM components among different *Streptococcus* species, they demonstrate selective site preferences for colonization in the human body[321]. Additionally, ability to create biofilm, as well as presence of the polysaccharide capsule contribute to their persistence and ability to reproduce in different environments[322,323]. All *Streptococcus* species express fibronectin binding proteins (FnBP) which could also facilitate their adhesion to the bladder mucosa[34,323]. Among members of that genus, *Streptococcus anginosus* is the most often isolated species from female UT[39,75,133,224]. However, increased detection of this species has been observed among UUI patients comparing to healthy individuals[74]. Recently reported as an emerging pathogen, *S. anginosus*, similarly to other oral streptococci can co-aggregate with other microorganisms, not necessarily commensals, such as *Candida albicans* and due to this mechanism contribute to yeast colonization[322].

Members of *Staphylococcus* genus, mostly coagulase-negative *Staphylococcus* (CoNS), are the other group of bacteria from the phylum Firmicutes often reported from healthy UT. Coagulase-positive *Staphylococcus aureus* is reported less often. They are catalase-positive cocci and they are widely found commensals in various human body niches and constitute major part of skin and mucosal surfaces flora. They are sometimes reported as causative agents of UTI (mainly *Staphylococcus saprophyticus*, less often *S. aureus* and *Staphylococcus epidermidis*[298]). Some CoNS are known to produce antimicrobial substances, such as antimicrobial peptides (AMPs), that can be efficient against a broad-spectrum of Gram-positive bacteria, including strains of methicillin-resistant *S. aureus* (MRSA). For instance, epidermin from *S. epidermidis* or hominicin produced by *Staphylococcus hominis*[324,325]. Lugdunin, obtained from *S. lugdunensis* is another potent bactericidal peptide with potential activity against some pathogenic bacteria including strains of MRSA or vancomycin-resistant *Enterococcus* sp.[326]. Additionally, extracellular serine proteases Esp secreted by those two

species can provide beneficial effect on skin or nasal cavity by inhibiting colonization with *S. aureus*[327,328]. *Staphylococcus* species possess various genes coding for adhesion and contributing to their colonization potential. For instance, polysaccharide intracellular adhesins (PIA), biofilm-associated protein (Bap) or extracellular matrix-binding proteins (Embp) are molecules facilitating bacterial adhesion to the ECM components (such as collagen, fibrinogen or fibronectin) or to host tissue and mediate biofilm formation[329–332]. Moreover, ability to form biofilm is one of the most relevant particularity of the *Staphylococcus* genus, as biofilm protects bacteria from host immunological defense, increases their antimicrobial resistance and facilitates colonization, persistence and chronic infection development[333,334].

From the phylum Actinobacteria, it is important to refer to the ubiquitous genus *Corynebacterium* sp. which encompasses Gram-positive bacteria presenting genomic high G+C content and a particular cell wall that includes a layer of mycolic acids connected to an additional polysaccharides layer on the top of peptidoglycan – arabinogalactan. This cell wall layer, called mycomembrane, is similar to outer membrane typical for Gram-negative bacteria by acting as permeability barrier and being involved in the host-microbe interactions[335–337]. Moreover, it provides protection from detergents, chemicals and enzymes including lysozyme but also antibiotics or physical factors. Beside structures characteristic to all members of *Corynebacterium* sp., like particular cross-linked peptidoglycan, lacking interpeptide bridges typical of other Gram-positive bacteria, this genus is highly diverse with variety of commensal and pathogenic species being able to adapt to distinct niches. One of the main differences includes lipid requirements by which members of this genus are classified as lipophilic or non-lipophilic *Corynebacterium* species, which also contributes for niche specificity, for instance *Corynebacterium* species from skin are mainly from the lipophilic group[338–340]. Interestingly, one of the most prevalent in urinary tract microbiome is lipophilic *Corynebacterium tuberculostearicum*, a common skin colonizer which can induce inflammation in human skin cell lines and is occasionally reported from diseased humans[341–343].
Other structures or molecules expressed in *Corynebacterium* cell wall could facilitate host colonization including pili or non-pilus adhesins, as *Corynebacterium* cell surface beside proteins is also enriched with lipoglycans (lipomannan, lipoarabinomannan) that may interact with host epithelial cells and favor adhesion[335,336,344,345]. Moreover, co-existence of *Corynebacterium* sp. and *Staphylococcus* sp. in the same niche (e.g., nasal cavity or skin) has unveiled interesting insights into microbe-microbe interactions. For instance,

Corynebacterium striatum can modify expression of genes involved in colonization and virulence of *S. aureus*, increasing adhesion to murine epithelial cells but reducing the virulence, in particular decreasing hemolysin activity[346,347]. Those features strongly support importance of *Corynebacterium* species as a part of microbiome, possibly also in the UT.

Another intriguing bacteria often found in urinary tract microbiome is *Prevotella* (phylum Bacteroidetes). These genus is characterized by high intra and inter species diversity, which reflects specific adaptations to different niches and diverse virulence potential[348]. *Prevotella* species are often found in healthy microbiome e.g., in oral cavity, gut[349]. Since they are inhabiting human body, some are considered commensal organisms, however they are also often associated with human infections. Interestingly, the same species found in healthy microbiome might be associated to disease condition (e.g., *Prevotella intermedia* in oral microbiome associated with periodontal disease, *Prevotella copri* in gut might be associated with immunological conditions)[349,350]. Currently, due to large strains variability, virulence factors conferring *Prevotella* pathogenicity are not clear.

For instance, the ability of *Prevotella* species to attach to the human mucosa could be consider virulence factor, however it is also required for colonization of the human body, even for the good purpose. Many species, including those often found in urinary tract microbiome e.g., *Prevotella bivia, Prevotella disiens*[39,75,133] has been shown to bind well to ECM[351]. Interestingly, in vaginal environment, abundant presence of *Prevotella* species is usually associated with the disease state[352]. Indeed, it seems to be important contributor to BV and there are various reports on possible microbe-microbe interactions that supports pathogen's growth and contribute to infection development[352–356].

However, with available data on urinary tract microbiome, *Prevotella* is reported as often detected bacteria, in asymptomatic and symptomatic populations[42,59,81]. Additional associations have been observed for its abundance and prevalence, which seems to be higher in women, especially in post-menopausal age[42,59,76,82].

Noteworthy, in gut microbiome, host's diet, lifestyle and other factors may have a strong influence on *Prevotella* diversity and prevalence[349], thus it is not excluded that in urinary tract microbiome, particular host specificities will play similar role.

1.5. Microbiome as a source of antimicrobial resistance genes

The extensive research on human microbiome every day improves our understanding of microbial communities inhabiting human body. However, in the context of human microbiome, antimicrobial resistance is also an important subject to be explored. Globally raising rate of microbes resistant to antibiotics constitute a threat for current and future centuries[357,358].

One of the evolutionary advantages of bacteria is horizontal gene transfer, which allows them to share specific gene content among each other[359]. This often relates to antimicrobial resistant genes (ARGs) mobilization, consequently contributing to increasing number of resistant bacteria. Considering that human body is naturally occupied by a large loads of diverse bacterial species, the possibility of ARGs spread between permanent and transient microbiome members (including putative pathogenic bacteria) is huge[360]. Thus, the microbiome is considered a reservoir of ARGs and its 'resistome' should be extensively studied[360,361].

Although some data on microbiome resistome is already available from widely studied human body niches (e.g., wide array of ARGs directed against antibiotics used in the clinic in gut microbiome or β-lactamases found in respiratory microbiome[361]), this matter has not yet been explored in urinary tract microbiome.

1.6. Current challenges in urinary tract microbiome research

Overall, over the past ten years, a significant progress has been made in urinary tract microbiome research. However, as this research field is still in its infancy, many fundamental questions remain unclear, and several limitations may be easily identified.

For instance, major part of data presented in available studies refer to genus level bacterial identification or does not apply identification methods sensitive enough to accurately identify bacteria at species level, which could be crucial to understand specific microbial interactions. Likewise, the understanding of population diversity at strain level is also essential and only now starts to be explored.

Moreover, common trend to report 'urotypes' focusing on most dominant taxa may contribute to undervalue low abundant microbiome members that might be potentially relevant. The urotype-based data representation is indeed most convenient for fast comparison of main findings however, microbiome is a complex community and eubiosis may be achieved due to various microbe-microbe, microbe-host and microbe-environment interactions, thus the complete composition of microbiome should be always provided.

Another important aspect would be understanding on how urinary tract microbiome shifts over time. To date, very few studies investigated temporal variations of urinary tract microbiome, with the focus on short-term shifts in asymptomatic population[51,148], however these initial insights are based on only few heterogenous individuals in each cohort. There are also single reports available on population with LUT syndromes e.g., OAB[168], IC[174] or UI[166]. Overall, the longitudinal studies on asymptomatic cohorts, investigating to which degree urinary tract microbiome may shift over time, especially in the perspective of its long-term stability, are still lacking.

Although several important studies had already been published, the urinary tract microbiome still needs some fundamental research focusing on robust characterization of microbiome members at species and at strain level, reliable identification methods, understanding of novel species inhabiting UT and temporal dynamics. Correcting this knowledge gaps may allow to translate urinary tract microbiome findings into diagnostics improvement and clinical practices.

Objectives

Our hypothesis was that an enhanced characterization of the female urinary tract microbiome (with in-depth species and strains' characterization) in asymptomatic individuals will allow the healthy urinary tract community profiling and/or identify members with pathogenic potential. Additionally, we hypothesized that an extensive interconnection between urinary and vaginal community may exist.

The **main goals** of this work were:

I. To comprehensively characterize the bacterial composition of the urinary tract microbiome of reproductive-age asymptomatic women, its long-term stability and interconnection with vaginal microbiome, using high-resolution culturomic analysis, supported with DNA-based approaches.

II. To unveil features of urinary tract microbiome members favoring their ability to colonize and persist in the asymptomatic urinary tract.

III. To evaluate the relevance of female urinary tract microbiome as a reservoir of putative uropathogens.

To accomplish these goals, reproductive-age European women were enrolled to participate in the study, and they each provided first morning mid-stream voided urine sample and vaginal swab. For the investigation of the temporal stability of urinary tract microbiome, a subset of women provided additional first morning urine and vaginal sample within 2,5 year since the previous sample collection. Additionally, 6 OAB and 3 rUTI patients provided first morning mid-stream voided urine for FUM analysis. Genomes available in public databases were included for comparative genomics study on *Escherichia coli*, *Limosilactobacillus* sp. and *Prevotella* sp. isolates.

The **specific objectives** were the following:

- To improve FUM structure characterization, throughout:

  i) Improvement of the culturomic and amplicon-based protocol, ensuring enhanced recovery of bacteria from urine samples and their reliable identification at species level;

  ii) critical appraisal of MALDI-TOF MS performance for accurate identification of FUM members;

  iii) detailed characterization (e.g., phylogenetic, genomic, biochemical analyses) of putatively new species.

- To explore infra-species diversity through genomic comparisons and to characterize their functional activities that might be related to colonization, resilience, and ability to cause infection.

Results

3.1. Healthy Female Urinary Tract Microbiome

Moving beyond genus to bacterial species clustering: community structure types of the healthy female urinary microbiome

Long-term stability of the urogenital microbiota of asymptomatic European women

Step towards understanding interconnection between vaginal and urinary tract microbiome

Moving beyond genus to bacterial species clustering: community structure types of the healthy female urinary microbiome

Svetlana Ugarcina-Perovic*[1], Magdalena Ksiezarek*[2,3], Joana Rocha[4], Márcia Sousa[2,3], Teresa Gonçalves Ribeiro[2,3], Filipa Grosso[2,3], Luísa Peixe[2,3]#

*SUP and MK contributed equally to this work.

[1]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

[2]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[3]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

[4]UCGenomics/GenomePT, Laboratório de Sequenciação e Genómica Funcional da Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal

Manuscript in preparation

Abstract

The recognition of a microbiota inhabiting the healthy bladder engendered the need for comprehensive characterization of the urinary microbiota in health and disease. Still, insufficient sensitivity of methodological approaches is hindering this knowledge. Moreover, a proper species identification of key players is critical in exploring the urinary tract microbiota under eubiotic and dysbiotic state.

The aim of this study was to uncover bacterial species combinations in midstream urine of healthy reproductive-age European women by extended culturomics and a cutting-edge long-reads third generation sequencing of the 16S rRNA gene V1-V8 regions to comprehensively characterize female urinary tract microbiota (FUM).

A wide bacterial species diversity was identified (297 species) by both methodologies, and there was not a single species present in all samples, although the genus *Lactobacillus* was detected in all. Instead, 14 bacterial species represented mostly by low abundant members were present in > 50% of samples, with at least 1% of abundance in one sample. Also we identified 16 uropathogens among healthy FUM. At genus level, the most prevalent community structure types were characterized by combinations of *Lactobacillus* spp. and other genera, while at species level a higher number of community structure types were identified by both methodologies. By moving beyond MALDI-TOF MS to sequencing of reliable gene markers, we unveiled identification of bacterial species from different genera not previously reported from healthy FUM and putative novel species.

This study revealed important details on the FUM composition at genus and species level, which is critical to unveil the potential relationship between specific community structure members and urogenital diseases/disorders.

Background

Emerging studies in female urinary microbiota (FUM) have suggested the importance of this unique bacterial community in maintaining urinary tract (UT) healthy (1–6). The current progress in investigating FUM through next generation sequencing and culture-based methodologies has allowed identification of FUM members and indication of their involvement in various UT conditions. These breakthrough findings have triggered the reassessment of current diagnosis practice of urinary tract infection (UTI) (7, 8) and investigation of still poorly understood etiologies of UT disorders (e.g., overactive bladder syndrome, urgency urinary incontinence and interstitial cystitis/bladder pain syndrome) (9–11).

Up to date, studies have described healthy FUM as a community dominated by certain genera such as *Lactobacillus*, *Gardnerella* or *Streptococcus*, or a mixed community without a single dominant genus involving, for example the combination of *Staphylococcus, Corynebacterium* and *Prevotella* genera (10, 12–14). *Lactobacillus crispatus, Lactobacillus jensenii*, and *Gardnerella vaginalis* are examples of bacterial species that have been reported as dominant in the lower urinary tract of healthy women of different age and menopausal status (13, 15), and the presence of certain potential pathogens such as *Escherichia coli* and *Enterococcus faecalis* were identified usually in low amounts (7, 13).

Notably, previous studies on healthy FUM were based on age-heterogeneous study cohorts, making difficult data systematization (1, 2, 13, 16–18). Moreover, so far applied DNA sequencing methodologies, targeting individual short hyper-variable regions of the 16S rRNA gene, are often limited to a reliable identification of FUM members only at genus level (19). On the other hand, culture-based methodologies with isolates' identification of insufficient resolution power still do not fully capture bacterial species diversity.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is commonly used identification methods for cultured isolates and is often not sufficient for species level discrimination e.g., *Lactobacillus* (unpublished data). Moreover, the recent taxonomic reassessment of the *Gardnerella* genus unveiled 4 new species and 13 genomic species (20), highlighting the need to revise the previous findings in healthy FUM that considered *Gardnerella* as *Gardnerella vaginalis*. Although some studies demonstrated that MALDI-TOF MS can be used to differentiate *Gardnerella* species (20, 21), commonly used MALDI-TOF MS spectral databases do not include reference spectra for novel species, thus the accurate identification of *Gardnerella* species is still challenging.

Considering the above, there is a critical need for accurate and sensitive characterization of the urinary tract microbiota of asymptomatic, healthy individuals to fully support future action in deciphering the microbial community shifts from a eubiotic to dysbiotic state, in order to guide the development of approaches to maintain or restore healthy FUM composition.

To improve our understanding of FUM, we analyzed midstream urine samples of twenty reproductive-age healthy women using a complementary approach of an extended culturomics supported in a deep bacterial taxonomic resolution and 16S rRNA long-reads sequencing. Our complementary approaches for a reliable bacterial taxonomic profiling, together with a rigorous study design, enabled identification of community structure types represented by species combinations, including well-known uropathogenic bacteria. With

improved identification methods based on amplification of reliable gene markers, we also provided evidence on a high diversity of species within the genera *Lactobacillus*, *Gardnerella*, and *Corynebacterium* (including putative novel species) in the healthy lower urinary tract, indicating their potential essential contribution in shaping urinary microbiota structure.

Methods

Participants and sample collection

This study was approved by the Faculty of Pharmacy (University of Porto, Porto, Portugal) Ethics Committee and written informed consent was obtained from all study participants. Twenty women of reproductive age were recruited between November 2016 and July 2018, following strict criteria: no pregnancy, no symptoms nor diagnosis of current UTI, and no antibiotic exposure in the previous month. A questionnaire was conducted concerning personal and health information that was encrypted, ensuring data confidentiality. Participants were carefully instructed in the collection technique. In the third week of menstrual cycle, each participant provided a first-morning midstream voided urine sample by self-performed non-invasive procedure via 40 ml sterile containers. Sampling procedure included vaginal swabbing, prior to urine collection, for minimizing cross-contamination. Urinary dipstick (Combur-Test, Roche) analysis and microscopic examination of the re-suspended sediment of centrifuged urine (1 ml) were performed. Up to 2 hours after collection, urine samples were subjected to extended culturomic protocol, concurrently pre-treated for amplicon sequencing analysis, and stored at -80 ^{o}C. The FUM culturomic data from ten women published in the context of urinary tract microbiome temporal stability (22), were included in this study. Since this manuscript includes novel data from amplicon sequencing performed on the same samples, previous culturomic data was used for comparison of efficacy of two methodologies and accurate assessment of community structure types.

Extended culturomics

The extended culturomic protocol included inoculation of 0.1 ml of urine onto the large plate surface (140 mm diameter) of Columbia agar with 5% sheep blood (blood agar plates - BAP, Biogerm, Portugal) and chromogenic agar (CAP, HiCrome UTI, HiMedia, India) supplemented as previously described (23, 24). BAPs and CAPs were incubated under aerobic and microaerophilic conditions (GENbox MICROAER, bioMérieux, France) at 37 °C for 48 h. Additionally, BAPs were incubated under anaerobic conditions (GENbox ANAER, bioMérieux, France) at 37 °C for 48 h. In case of a suspected high bacterial load based on

microscopic observation, ten-fold serial dilutions (up to 0.001) were performed using saline solution (0.9% NaCl) to obtain a countable range of colony forming units (CFU/ml). Each morphologically distinct colony type was counted, and 1-5 colonies of each morphology were further identified. The plate presenting the higher CFU count was considered as the representative count of each isolate in a sample. Relative abundance (RA; %) was calculated by generating the percent of total CFU/sample.

Identification of cultured bacteria

MALDI-TOF MS with the *in vitro* diagnostic (IVD) database version 3.0 (VITEK MS automation control and Myla software, bioMérieux, France) was used to identify the bacterial isolates. Isolates with no identification, with discrepant results between MALDI-TOF MS identification and phenotypic characteristics, or with known insufficient resolution power for species identification were further subjected to sequencing of 16S rRNA gene, other genetic markers (*pheS* for *Lactobacillus* and *Limosilactobacillus*, *cpn60* for *Gardnerella, rpoB* for *Acinetobacter*, *Corynebacterium* or *Staphylococcus,* and *recN* for *Citrobacter*) and/or PCR assays for the detection of species-specific genes (*dltS* for Group B *Streptococcus, sodA* for *Enterococcus faecalis*, and *malB* for *Escherichia coli*) (**Additional file 1:** Table S1). Phylogenetic analysis based on individual gene were performed to access putative novel species by using MEGA version 7.0 (25), constructed according to neighbour-joining method, and genetic distances were estimated using Kimura's 2-parameter model. The reliability of internal branches was assessed from bootstrapping based on 1000 resamplings.

DNA extraction and amplicon sequencing

Samples were pretreated prior to DNA extraction, which included centrifugation of 20 ml of urine at 5,500 rpm for 15 min, with resulting pellet suspended in 1 ml of phosphate buffered saline, and stored at -80 °C until further processing. Genomic DNA from urine samples was extracted using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Germany), according to the manufacturer's protocol. DNA was eluted into 50 µl of Tris-HCl [pH 8.0]) and stored at 4°C. DNA quality was analyzed by agarose gel electrophoresis, and quantity was measured on Qubit dsDNA HS Assay Kit (Invitrogen, Life Technologies, UK). Controls consisting of reagent blanks (washing buffer, lysis buffer and kit reagents) were processed as the urine samples. Because extraction controls showed no traceable amounts of DNA they were not included for sequencing. PCR amplification of the hypervariable 16S rRNA gene V1-V8 regions sequenced with universal primers (27F:AGAGTTTGATCCTGGCTCAG, and BS-

R1407:GACGGGCGGTGWGTRC), library construction and sequencing with SMRT® technology on PacBio RS II sequencing system was provided as a custom service of Eurofins GATC Biotech GmbH (Germany).

Sequencing data analysis

After sequencing, primers, sequence adaptors, and low base quality calls were removed by Cutadapt. Chimera sequences were checked, and removed by UCHIME (version 4.2.40) (26). The non-chimera and unique sequences were subjected to BLASTn (27) analysis using non-redundant 16S rRNA reference sequences with an E-value cutoff of 1e-06. Reference 16S rRNA gene sequences were obtained from the Ribosomal Database Project Classifier (28). Only good quality and unique 16S rRNA sequences which have a taxonomic assignment were considered and used as a reference database to assign operational taxonomic unit (OTU) status with a 97% similarity. Taxonomic classification was based on the NCBI Taxonomy (29). All the hits to reference 16S rRNA database are considered and specific filters are applied to the hits to remove false positives. The thresholds applied were: ≥ 97.00% identity, ≥ 95.00% alignment coverage, 1000 minimum query length, 10% bitscore threshold for multiple hits, and 250 maximum hits to consider for multiple hits. If the final number of high-quality reads after all filtering steps was less than 1000, the corresponding sample was excluded. Finally, RA was calculated by generating the percent of total reads for each sample.

Statistical analysis

Community structural analyses were done using relative proportions of CFU/ml and reads for each genus and species within individual urine samples. Based on similarity (or dissimilarity) of community composition between samples and taking into account all members and their proportion in a community, we identified community structure types performing hierarchical clustering of Bray-Curtis dissimilarity distance matrices with a cutoff of 0.8, via the package vegan (version 2.5-2) (30) in R (version 3.4.4) (31). Alpha diversity was estimated using the Shannon index. Principal coordinates analysis (PCoA) and Mantel test between the dissimilarity distance matrices (based on Bray-Curtis index) were performed to compare structure types obtained by both methodologies. To identify species responsible for community structure differences, biplot of the PCoA was created using a weighted average of the species scores, based on their RA in the samples. Data visualisation was carried out using the gplots (version 3.0.1.1) (32), ggplot2 (version 3.2.1) (33) and eulerr (version 5.1.0) R packages.

Results

Overview of the healthy female study cohort

Our study cohort included twenty female participants aged 24-38 years (average = 31; standard deviation = 4). Most women identified themselves as Portuguese nationality (80%) followed by other European nationalities (20%). Average body mass index was 21.9 kg/m^2. Most women had a normal menstrual cycle (90%) with the use of contraceptives (85%), with few having experienced at least one pregnancy (25%). Characteristics of our study cohort comprised of healthy highly educated women, including clinical and behavioral questionnaire data (personal medical history, UT health and infection history, pregnancy history, demographic and lifestyle information), and results of urine dipstick and sediment microscopic analysis are available in **Additional file 1:** Table S2-S3.

Characterization of community structure types by culturomics

Using extended culturomics we observed a high bacterial load in urine samples (10^3-10^8 CFU/ml, ≥ 10^4 CFU/ml in 80% of samples). Two thousand and forty-three isolates were studied (median = 103 isolates/sample) and assigned to 131 species (median = 20 species/sample), and 54 genera, as identified either by MALDI-TOF MS and/or sequencing of most suitable genes (Additional file 1: Table S4). We identified for the first time 4 bacterial species from different genera [*Gardnerella leopoldii*, *Globicatella sulfidifaciens*, *Limosilactobacillus mucosae* (former *Lactobacillus mucosae*), and *Staphylococcus equorum*], and 5 putative novel *Corynebacterium* species that were not reported previously in the urinary tract of asymptomatic non-pregnant women (Additional file 1: Table S4, Additional file 2: Figure S4). The alpha diversity varied from 0.001 to 2.65 (median H′ = 1.5). Bacterial species detected by culturomics and their RA per sample are listed in Additional file 1: Table S4.

Clustering FUM into community structure types (CST) was performed at genus and species level (samples in the same CST shared >80% similarity by Bray-Curtis distance). Hierarchical clustering at genus level identified 3 CST (Additional file 2: Fig. S1). The most common CST was CST3 (n=15/20) largely dominated by *Lactobacillus* in combination with other genera (e.g., *Staphylococcus*, *Corynebacterium*, *Streptococcus* and *Cutibacterium*), followed by CST2 (n=4) characterized mostly by *Gardnerella*, and CST1 dominated by *Citrobacter* (n=1). On the other hand, species-level clustering resulted in 13 CST (Fig. 1, Table 1), most representing individual urine specimens as only 5 CST included more than one sample. With exception of 2 clusters dominated by a single bacterial species [CST1-

Citrobacter koseri, and CST2-*Gardnerella vaginalis*, >90%), the remaining CST were mostly represented by an extraordinarily diverse bacterial community (different combinations and RA of bacterial species), which varied widely from 1.21±0.05 to 2.65 as calculated by the Shannon diversity index (Figure 1, Table 1). For instance, CST5 was characterized by combination of *Lactobacillus iners* with other bacterial species, and CST12 included *Lactobacillus crispatus, Lactobacillus mulieris* and other bacterial species (Figure 1).

Characterization of community structure types by amplicon sequencing

A total of 58,534 reads were generated, with most of them being assigned to the species level (88%; 51,317 reads). One sample (U6a) had <1000 reads and was excluded from the analysis, while for the remaining a median of 2493 reads/sample (interquartile range, IQR 1625 - 3920) was generated. A total of 231 species (IQR 5-115, median = 39 species/sample) belonging to 107 genera and 8 phyla were identified. The alpha diversity varied from 0.135 to 2.79 (median H′ = 0.90). Bacterial species detected by amplicon sequencing and their RA are listed in Additional file 1: Table S5.

The same FUM clustering approach was applied to amplicon sequencing data. Genus-level clustering resulted in 5 CST (Additional file 2: Fig. S2). The *Lactobacillus* genus in combination with other bacterial genera (e.g., *Prevotella, Dialister,* and *Corynebacterium*) represented the most prevalent CST (CST5; 79%, n=15/19). Contrariwise, species-level clustering resulted in 7 CSTs (Fig. 2, Table 2), with the most common 3 CSTs (n=15) being characterized by combination of a highly abundant *Lactobacillus* species (CST3-*L. iners*, CST5-*Lactobacillus gasseri*, CT7-*L. crispatus*) with species from other genera (Fig. 2). Remarkably, the *Lactobacillus iners* enriched CST was characterized by a reduced species diversity (CST3, H′=0.56±0.42) comparing to other *Lactobacillus* CSTs. The remaining CST included highly abundant *C. koseri* (CST1; n=1/19), *Atopobium vaginae* (CST2; n=1/19), or combination of different species (CST4: *Anaerococcus tetradius* and *Prevotella timonensis*; CST6: *Ralstonia mannitolilytica* and *Streptococcus agalactiae*; n=1 each).

Correlation between community structure types assigned by culturomics and amplicon sequencing

A moderate correlation was observed using the Mantel test (r = 0.5, p< 0.05) between the CST assigned by culturomics and amplicon sequencing. Congruence was observed for the types of highly abundant *C. koseri* and combinations of different *Lactobacillus* species (Fig. 1, Fig. 2), while *Lactobacillus* amongst others were responsible for the reduction in correlation

between CST detected by different methodologies (e.g., *Lactobacillus iners* was more frequently detected in a higher RA by amplicon sequencing, while *Cutibacterium acnes* by culturomics) (Fig. 3).

Overview of bacterial species in healthy FUM

In total we captured an extended set of bacteria belonging to 8 phyla, 116 genera and 297 species (median = 53 species/sample) in healthy FUM (Additional file 1: Table S4-S5; Additional file 2: Figure S3). Out of 297 species, we have identified by both methodologies 65 species (22% of total species) belonging to 35 genera and 5 phyla. We could not identify a single species present in all samples, although the genus *Lactobacillus* was detected in all. Instead, we were able to unveil a core of 14 bacterial species by selecting its presence in more than 50% of samples, with at least 1% of abundance in one sample (Fig. 4, Additional file 1: Table S6). *Staphylococcus epidermidis* was the most common species (n=18/20), followed by *Finegoldia magna* (n=16/20), *Corynebacterium tuberculostearicum* (n=15/20), and *Prevotella bivia* (n=15/20) (Additional file 1: Table S6). Remarkably, this common species were mostly low-abundant members (RA < 5%) (Fig. 4).

Additionally, we looked for the presence of opportunistic pathogens associated with the urogenital tract health and found 16 bacterial species largely varying in their RAs (IRQ 0,03-96.62%), among which *Enterococcus faecalis*, *Streptococcus anginosus*, and *Ureaplasma parvum* were the most frequently identified by both methodologies (Table 3). Noteworthy, *C. koseri* was a highly abundant member detected by both methodologies, while *Atopobium vaginae* was by amplicon sequencing. All opportunistic pathogens associated with the urogenital tract detected by culturomics and/or amplicon sequencing in healthy FUM are listed in Table 3.

Discussion

Understanding the microbial composition of the lower urinary tract in healthy individuals is essential so that microbial changes associated with urinary disorders can be recognized and modulated as a therapeutic strategy. In this study, using a complementary approach of two methodologies (extended culturomics and amplicon sequencing), we expanded the knowledge on compositional patterns of the female lower urinary tract microbiome.

Each technique detected similar but not identical microbiome profiles, and only 22% of bacterial species were detected by both methodologies. Predictably, amplicon sequencing

allowed more frequent detection of slow-growing species (e.g., *Campylobacter ureolyticus*), and obligate anaerobes (e.g., *Finegoldia magna*) that require particular culturing conditions (34). On the other hand, the cultured isolates could be accurately identified to the species level, thus providing a higher level of resolution, and allowing further investigation to unveil their symbiotic or pathogenic potential. Some species detected in low-reads count (e.g., *Staphylococcus aureus* and *Actinomyces urogenitalis*, RA < 0.1%) were also identified by extended culturomics. Thus, our findings support previous studies highlighting the possible underestimation of bacterial diversity from exclusively DNA-based studies, and that the complementarity of both methodological approaches ensures a more comprehensive description of the FUM diversity (34).

Clustering FUM at genus level reveal that the most prevalent CST was characterized by the combination of highly abundant *Lactobacillus* and other genera, which was detected by amplicon sequencing and extended culturomics, confirming previously reported high occurrence of this genus in urinary microbiota (10, 13, 22). At species level, the majority of the CST were represented by different *Lactobacillus* or *Gardnerella* species in different RA, and in combination with species from other genera, including low-abundant FUM members (both methodologies or extended culturomic, respectively), as observed in our previous study (22).

Remarkably, we identified for the first time a CST dominated by *Atopobium vaginae* (RA ~ 87%) based on amplicon sequencing in an asymptomatic individual (U15a) (Fig. 2), which was also isolated (RA ~ 33%) in combination with *Gardenerella swidsinskii* (RA ~ 49%) (Fig. 1) (Additional file 1: Table S4, Table S5). *Atopobium vaginae* is associated with bacterial vaginosis and rarely occurs in the absence of *Gardnerella vaginalis* (35, 36). However, *t*he woman did not report any symptom associated with urogenital diseases, suggesting that *Atopobium vaginae* and other identified opportunistic uropathogens (e.g. *E. coli*, *C. koseri*, or *E. faecalis*), or species more frequently isolated from women with specific urinary disorders (e.g., *Aerococcus urinae*, *Lactobacillus gasseri*) (Table 3, Fig. 1) (37, 38), might not be the cause or biomarkers of urogenital infections/disorders, but rather correlated with strains' specific pathogenicity factors. Further elucidation of the function of urinary microbiota, including characterization of virulence factors *sensu stricto* playing a significant role in pathogenesis, and not found in commensal bacteria occupying the same body or niche might help to understand the development of urogenital dysbiosis (39, 40).

Looking into more detail to the *Gardnerella* species identified in this study by culturomics (Additional file 1: Table S4), we detected *Gardnerella vaginalis*, *Gardnerella leopoldii*, *Gardnerella swidsinskii*, and *Gardnerella* genomospecies 3 among the 14

Gardnerella species/groups described to date (20, 21). Interestingly, we did not find *Gardnerella piotii* among our urinary samples, further reinforcing previous evidence that *Gardnerella piotii* seems to be a commensal of vagina (21, 41). Although *Gardnerella* was also identified by the 16S rRNA gene amplicon sequencing, species cannot be delineated on the basis of this gene (20).

Interestingly, we detected a high diversity of *Corynebacterium* species (25 species; 10 by extended culturomics - including 5 putative novel species, 7 by amplicon sequencing, and 8 by both methodologies), and *Lactobacillaceae* members (3 genera; 13 species; 4 by extended culturomics, 2 by amplicon sequencing, and 7 by both methodologies) that was never reported in previous studies characterizing the asymptomatic FUM (Additional file 1: Table S4, Table S5) (1–3, 7, 10, 13, 42, 43), which demonstrates that the reliable identification of isolated strains by specific genotypic markers rather than only MALDI-TOF MS, and by the use of cutting-edge long-reads third generation sequencing of the 16S rRNA gene rather than short-reads increase the knowledge on the composition of bacterial community to the species level in microbiome studies (22, 44).

The strengths of this study include the sample processing up to 2 hours after collection, allowing us to identify anaerobic bacteria that seems to significantly contribute to urinary microbiota repertoire (34), but are rarely or not reported by other healthy FUM culturomics studies (e.g., *Prevotella corporis)* (3, 7, 45). The study was further strengthened through the use of a larger volume sample size (20 ml), compared to previously used urine volume (mostly 1 ml) in DNA extraction protocols, which increased high-quality microbial DNA yield required for high-resolution sequencing, and unveiled detection of species not previously reported in DNA-based studies (e.g., *Alistipes putredinis*) (1, 10, 12, 34). Another important strength of this study was the use of a cutting-edge sequencing technique, including near full-length 16S rRNA gene sequencing using PacBio SMRT cell technology (46–48), and appropriate gene markers to identify cultured isolates at species level, which enable increased taxonomic resolution, as well as validation of several low-read sequencing data (< 0.1% RA) by our extended culturomic protocol.

One limitation of this study was the small cohort size, yet our strictly selected participants (e.g., no antibiotics for any medical reason within the month prior to urine collection and samples collected on 3rd week of menstrual cycle) represented a homogenous healthy female group. Another limitation of this study was the use of voided urine instead of suprapubic aspiration or transurethral catheterization specimen urine may have contributed to an

increased diversity in the urinary microbiome owing to possible genital contribution (49, 50). However, careful vaginal swabbing was employed, and suprapubic aspiration or catheterization of participants who were not at a high risk of bacterial infection or not with any clinical urinary symptoms was not ethically feasible as per our local ethics committee. Additionally, voided urine samples capture urethral bacteria, which can play important role in urinary conditions.

Conclusions

Our study substantially enlarged the knowledge on bacterial species diversity and low abundant members of healthy FUM, highlighting the importance of detection and characterization of low-abundant members that also build FUM community structure types. We provided extensive taxonomic characterization of *Gardnerella, Lactobacillus*, and *Corynebacterium* which are prevalent members in this niche. We also demonstrated that healthy FUM is composed of various combinations of species, thus should be described in detail, rather than just by dominant genus. Finally, our findings provide essential species level information for further studies on microbiota dysbiosis associated with urinary tract infection and lower urinary tract symptoms, required for development of more effective diagnostic and/or therapeutic strategies. As we begin to understand composition and diversity of urinary microbiota, future studies accessing the functionality of resident microbiota in human urinary tract should be considered high priority research.

List of abbreviations

BAP: columbia agar with 5% sheep blood plate
CAP: chromogenic agar plate
CFU: colony forming unit
CST: community structure type
dltS: histidine kinase specific to group B *Streptococcus*
FUM: female urinary microbiota
IVD: *in vitro* diagnostic
malB: maltose operon protein B
MALDI-TOF MS: matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
ND: not detected
OTU: operational taxonomic unit

PCoA: principal coordinates analysis
pheS: phenylalanyl-tRNA synthetase alpha subunit
RA: relative abundance
recN: DNA repair protein
rpoB: RNA polymerase beta subunit
SDS: sodium dodecyl sulfate
sodA: superoxide dismutase
UT: urinary tract
UTI: urinary tract infection

Declarations

Ethics approval and consent to participate

Approval of the study was obtained from the Faculty of Pharmacy (University of Porto, Porto, Portugal) Ethics Committee. Procedures performed in the study were all in accordance with the ethical standards of the institutional and national research committee, with the 1964 Helsinki Declaration, and its later amendments. All individual participants included in the study had given written informed consent.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files, and available in the Sequence Read Archive repository, under BioProject accession number PRJNA548360 (51).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal, and by the UCIBIO – Applied Molecular Biosciences Unit, Laboratory of Microbiology, Department of Biological Sciences, REQUIMTE, Faculty of Pharmacy, University of Porto, 4050-313

Porto, which are financed by national funds from FCT (Fundação para a Ciência e a Tecnologia, I.P.) (UIDB/MULTI/04378/2020 and UIDP/04378/2020). SP was supported as researcher from the project NORTH-01-0145-FEDER-000024, MK and MS by an FCT PhD grant (SFRH/BD/132497/2017 and SFRH/BD/05038/2020, respectively), and JR by a postdoctoral fellowship from ICETA (UID/MULTI/04378/2013). FG is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0034].

Authors' contributions

SUP, MK, FG and LP designed the study and supervised participant recruitment. SUP, MK and JR processed the samples and collected the data. SUP, MK, JR, MS, EC and TGR performed the isolates' identification. TGR supervised MS and EC. SUP conducted the community data analysis and visualization. SUP and MK interpreted the data and wrote the manuscript. TGR, FG and LP revised the article. All authors read and approved the manuscript.

Acknowledgements

We thank the women who participated and contributed time and effort to make this study a success. We thank Helena Ramos and Paulo Pinto (Hospital Geral de Santo António, Porto, Portugal) for their technical support with MALDI-TOF MS-based microbial identification. The bioMérieux (Portugal, Lda.) provided equipment and material for MALDI-TOF MS analysis, and had no role in the study design, data collection and analysis, decision to publish, or writing of the manuscript.

References

1. Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS. 2011. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. BMC Microbiol 11:244.
2. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, Fitzgerald M, Mueller ER, Schreckenberger P, Dong Q, Nelson DE, Brubaker L. 2012. Evidence of uncultivated bacteria in the adult female bladder. J Clin Microbiol 50:1376–1383.
3. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. Journal of Clinical Microbiology 52:871–876.

4. Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, Brubaker L, Wolfe AJ, Mueller ER, Schreckenberger PC. 2016. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. J Clin Microbiol 54:1216–1222.
5. Govender Y, Gabriel I, Minassian V, Fichorova R. 2019. The Current Evidence on the Association Between the Urinary Microbiome and Urinary Incontinence in Women. Front Cell Infect Microbiol 9:133.
6. Dornbier RA, Bajic P, Van Kuiken M, Jardaneh A, Lin H, Gao X, Knudsen B, Dong Q, Wolfe AJ, Schwaderer AL. 2020. The microbiome of calcium-based urinary stones. Urolithiasis 48:191–199.
7. Coorevits L, Heytens S, Boelens J, Claeys G. 2017. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. Eur J Clin Microbiol Infect Dis 36:635–639.
8. Sathiananthamoorthy S, Malone-Lee J, Gill K, Tymon A, Nguyen TK, Gurung S, Collins L, Kupelian AS, Swamy S, Khasriya R, Spratt DA, Rohn JL. Reassessment of Routine Midstream Culture in Diagnosis of Urinary Tract Infection. Journal of Clinical Microbiology 57:e01452-18.
9. Wu P, Chen Y, Zhao J, Zhang G, Chen J, Wang J, Zhang H. 2017. Urinary Microbiome and Psychological Factors in Women with Overactive Bladder. Front Cell Infect Microbiol 7:488.
10. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. mBio 5:e01283-01214.
11. Abernethy MG, Rosenfeld A, White JR, Mueller MG, Lewicky-Gaupp C, Kenton K. 2017. Urinary Microbiome and Cytokine Levels in Women With Interstitial Cystitis. Obstet Gynecol 129:500–506.
12. Thomas-White KJ, Hilt EE, Fok C, Pearce MM, Mueller ER, Kliethermes S, Jacobs K, Zilliox MJ, Brincat C, Price TK, Kuffel G, Schreckenberger P, Gai X, Brubaker L, Wolfe AJ. 2016. Incontinence medication response relates to the female urinary microbiota. Int Urogynecol J 27:723–733.
13. Price TK, Hilt EE, Thomas-White K, Mueller ER, Wolfe AJ, Brubaker L. 2020. The urobiome of continent adult women: a cross-sectional study. BJOG 127:193–201.

14. Curtiss N, Balachandran A, Krska L, Peppiatt-Wildman C, Wildman S, Duckett J. 2017. A case controlled study examining the bladder microbiome in women with Overactive Bladder (OAB) and healthy controls. Eur J Obstet Gynecol Reprod Biol 214:31–35.
15. Gottschick C, Deng Z-L, Vital M, Masur C, Abels C, Pieper DH, Wagner-Döbler I. 2017. The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. Microbiome 14:5(1).
16. Fouts DE, Pieper R, Szpakowski S, Pohl H, Knoblach S, Suh M-J, Huang S-T, Ljungberg I, Sprague BM, Lucas SK, Torralba M, Nelson KE, Groah SL. 2012. Integrated next-generation sequencing of 16S rDNA and metaproteomics differentiate the healthy urine microbiome from asymptomatic bacteriuria in neuropathic bladder associated with spinal cord injury. J Transl Med 10:174.
17. Lewis DA, Brown R, Williams J, White P, Jacobson SK, Marchesi JR, Drake MJ. 2013. The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. Front Cell Infect Microbiol 3.
18. Curtiss N, Balachandran A, Krska L, Peppiatt-Wildman C, Wildman S, Duckett J. 2018. Age, menopausal status and the bladder microbiome. European Journal of Obstetrics & Gynecology and Reproductive Biology 228:126–129.
19. Hoffman C, Siddiqui NY, Fields I, Gregory WT, Simon HM, Mooney MA, Wolfe AJ, Karstens L. Species-Level Resolution of Female Bladder Microbiota from 16S rRNA Amplicon Sequencing. mSystems 0:e00518-21.
20. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P. 2019. Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. Int J Syst Evol Microbiol 69:679–687.
21. Putonti C, Thomas-White K, Crum E, Hilt EE, Price TK, Wolfe AJ. 2021. Genome Investigation of Urinary *Gardnerella* Strains and Their Relationship to Isolates of the Vaginal Microbiota. mSphere 6.
22. Ksiezarek M, Ugarcina-Perovic S, Rocha J, Grosso F, Peixe L. 2021. Long-term stability of the urogenital microbiota of asymptomatic European women. BMC Microbiol 21:64.
23. Alves P, Castro J, Sousa C, Cereija TB, Cerca N. 2014. *Gardnerella vaginalis* outcompetes 29 other bacterial species isolated from patients with bacterial vaginosis, using in an in vitro biofilm formation model. J Infect Dis 210:593–596.

24. Man JCD, Rogosa M, Sharpe ME. 1960. A Medium for the Cultivation of Lactobacilli. Journal of Applied Bacteriology 23:130–135.
25. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874.
26. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200.
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.
28. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141-145.
29. Federhen S. 2012. The NCBI Taxonomy database. Nucleic Acids Res 40:D136-143.
30. Oksanen J, Guillaume B, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevens H, Szoecs E, Wagner H. 2019. Vegan: Community Ecology Package.
31. RCoreTeam. 2019. R: a Language and Environment for Statistical Computing.
32. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B. 2019. gplots: Various R programming tools for plotting data.
33. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
34. Dubourg G, Morand A, Mekhalif F, Godefroy R, Corthier A, Yacouba A, Diakite A, Cornu F, Cresci M, Brahimi S, Caputo A, Lechevallier E, Tsimaratos M, Moal V, Lagier J-C, Raoult D. 2020. Deciphering the Urinary Microbiota Repertoire by Culturomics Reveals Mostly Anaerobic Bacteria From the Gut. Front Microbiol 11.
35. Verhelst R, Verstraelen H, Claeys G, Verschraegen G, Delanghe J, Van Simaey L, De Ganck C, Temmerman M, Vaneechoutte M. 2004. Cloning of 16S rRNA genes amplified from normal and disturbed vaginal microflora suggests a strong association between *Atopobium vaginae*, *Gardnerella vaginalis* and bacterial vaginosis. BMC Microbiol 4:16.
36. Muzny CA, Taylor CM, Swords WE, Tamhane A, Chattopadhyay D, Cerca N, Schwebke JR. 2019. An Updated Conceptual Model on the Pathogenesis of Bacterial Vaginosis. J Infect Dis 220:1399–1405.

37. Pearce MM, Zilliox MJ, Rosenfeld AB, Thomas-White KJ, Richter HE, Nager CW, Visco AG, Nygaard IE, Barber MD, Schaffer J, Moalli P, Sung VW, Smith AL, Rogers R, Nolen TL, Wallace D, Meikle SF, Gai X, Wolfe AJ, Brubaker L, Pelvic Floor Disorders Network. 2015. The female urinary microbiome in urgency urinary incontinence. Am J Obstet Gynecol 213:347.e1–11.
38. Fok CS, Gao X, Lin H, Thomas-White KJ, Mueller ER, Wolfe AJ, Dong Q, Brubaker L. 2018. Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. Int Urogynecol J 29:1765–1771.
39. Hill C. 2012. Virulence or niche factors: what's in a name? J Bacteriol 194:5725–5727.
40. Tauch A, Burkovski A. 2015. Molecular armory or niche factors: virulence determinants of *Corynebacterium* species. FEMS Microbiol Lett 362:fnv185.
41. Hill JE, Albert AYK, Group the VR. 2019. Resolution and Cooccurrence Patterns of *Gardnerella leopoldii*, *G. swidsinskii*, *G. piotii*, and *G. vaginalis* within the Vaginal Microbiome. Infection and Immunity 87.
42. Karstens L, Asquith M, Davin S, Stauffer P, Fair D, Gregory WT, Rosenbaum JT, McWeeney SK, Nardos R. 2016. Does the Urinary Microbiome Play a Role in Urgency Urinary Incontinence and Its Severity? Front Cell Infect Microbiol 6:78.
43. Komesu YM, Richter HE, Carper B, Dinwiddie DL, Lukacz ES, Siddiqui NY, Sung VW, Zyczynski HM, Ridgeway B, Rogers RG, Arya LA, Mazloomdoost D, Gantz MG, Pelvic Floor Disorders Network. 2018. The urinary microbiome in women with mixed urinary incontinence compared to similarly aged controls. Int Urogynecol J 29:1785–1795.
44. Franzén O, Hu J, Bao X, Itzkowitz SH, Peter I, Bashir A. 2015. Improved OTU-picking using long-read 16S rRNA gene amplicon sequencing and generic hierarchical clustering. Microbiome 3:43.
45. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, Hilt EE, Price TK, Wolfe AJ, Lawley TD. 2018. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. Nat Commun 9.
46. Wagner J, Coupland P, Browne HP, Lawley TD, Francis SC, Parkhill J. 2016. Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. BMC Microbiol 16:274.
47. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK. 2016. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. PeerJ 4:e1869.

48. Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, Gies EA, Cheng J-F, Copeland A, Klenk H-P, Hallam SJ, Hugenholtz P, Tringe SG, Woyke T. 2016. High-resolution phylogenetic microbial community profiling. ISME J 10:2020–2032.
49. Nickel JC, Stephens-Shields AJ, Landis JR, Mullins C, van Bokhoven A, Lucia MS, Henderson JP, Sen B, Krol JE, Ehrlich GD, MAPP Research Network. 2019. A Culture-Independent Analysis of the Microbiota of Female Interstitial Cystitis/Bladder Pain Syndrome Participants in the MAPP Research Network. J Clin Med 8.
50. Bresler L, Price TK, Hilt EE, Joyce C, Fitzgerald CM, Wolfe AJ. 2019. Female lower urinary tract microbiota do not associate with IC/PBS symptoms: a case-controlled study. Int Urogynecol J 30:1835–1842.
51. Ugarcina-Perovic S, Ksiezarek M, Rocha J, Grosso F, Peixe L. Female Urinary Microbiota in Health and Disease [Internet]. Sequence Read Archive. 2019. Available from: https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA548360

Table 1. Overview of all healthy community structure types and their characteristic species by culturomics. Shared species within a structure type are presented in order of decreasing RA (RA > 1%, only top 5 shown).

Structure type	Characteristic species	Samples	Shannon index (mean H′ ± Standard Deviation)
1	*Citrobacter koseri*	U26a	0.001
2	*Gardnerella vaginalis* *Lactobacillus gasseri*	U29a	0.33
3	*Gardnerella leopoldii* *Alloscardovia omnicolens* *Bifidobacterium spp.* *Winkia neuii* *Streptococcus anginosus*	U1b	1.61
4	*Streptococcus mitis/oralis* *Staphylococcus haemolyticus* *Micrococcus luteus* *Actinomyces* spp. *Lactobacillus crispatus*	U4b	1.85
5	*Lactobacillus iners* *Corynebacterium tuberculostearicum* *Staphylococcus epidermidis* *Staphylococcus hominis* *Staphylococcus capitis*	U5a, U22a, U25a	1.61±0.20
6	*Streptococcus agalactiae* *Streptococcus salivarius/vestibularis* *Micrococcus luteus* *Staphylococcus haemolyticus*	U3b, U6a	1.40±0.35
7	*Lactobacillus paragasseri* *Lactobacillus delbrueckii* *Brevibacterium* spp. *Pseudoglutamicibacter cumminsii* *Corynebacterium jeikeium*	U7a	1.87
8	*Enterococcus faecalis* *Staphylococcus epidermidis* *Lactobacillus gasseri* *Streptococcus anginosus* *Corynebacterium aurimucosum*	U2a	1.97
9	*Lactobacillus jensenii* *Staphylococcus haemolyticus* *Staphylococcus epidermidis* *Corynebacterium amycolatum* *Corynebacterium coyleae*	U12a	2.65
10	*Atopobium vaginae* *Streptococcus anginosus*	U15a, U23a	1.21±0.05
11	*Corynebacterium striatum* *Dermabacter hominis* *Staphylococcus aureus* *Corynebacterium* sp. nov. 4 *Lactobacillus crispatus*	U9a	1.72
12	*Lactobacillus crispatus* *Lactobacillus mulieris* *Staphylococcus epidermidis* *Cutibacterium avidum*	U8a, U10a, U19a	1.77±0.45
13	*Lactobacillus crispatus* *Corynebacterium tuberculostearicum* *Finegoldia magna*	U11a, U24a	1.52±0.58

Table 2. Overview of all healthy community structure types and their characteristic species by amplicon sequencing. Shared species within a structure type are presented in order of decreasing RA (RA > 1%, only top 5 shown).

Structure type	**Characteristic species**	**Samples**	**Shannon index** (mean H′± Standard Deviation)
1	*Citrobacter koseri* *Citrobacter* spp. *Lactobacillus iners*	U26a	0.21
2	*Atopobium vaginae* *Sneathia sanguinegens*	U15a	0.59
3	*Lactobacillus iners* *Prevotella timonensis*	U1b, U8a, U5a, U11a, U2a, U22a, U12a, U25a	0.56±0.42
4	*Anaerococcus tetradius* *Prevotella timonensis* *Lactobacillus jensenii* *Atopobium vaginae* *Ureaplasma parvum*	U23a	1.78
5	*Lactobacillus gasseri* *Prevotella timonensis* *Dialister propionicifaciens* *Campylobacter ureolyticus*	U7a, U29a	1.70±1.13
6	*Ralstonia mannitolilytica* *Streptococcus agalactiae* *Kocuria* spp.	U3b	2.12
7	*Lactobacillus crispatus* *Corynebacterium* spp. *Corynebacterium tuberculostearicum* *Peptoniphilus* spp.	U4b, U9a, U19a, U10a, U24a	1.86±0.85

Table 3. Opportunistic pathogens associated with the urogenital tract. Species are listed in order of decreasing detection frequency in FUM.

Species	Frequency in FUM* (%)	Culturomics		Amplicon sequencing	
		Frequency in 20 samples (%)	RA (%)	Frequency in 19 samples (%)	RA (%)
Enterococcus faecalis	12/20 (60%)	11/20 (55%)	0.01-20	3/19 (16%)	0.04-0.22
Streptococcus anginosus	11/20 (55%)	10/20 (50%)	0.06-13.33	7/19 (37%)	0.03-1.31
Ureaplasma parvum	8/20 (40%)	ND	ND	8/19 (42%)	0.10-15.10
Escherichia coli	6/20 (30%)	4/20 (20%)	0.02-0.28	4/19 (21%)	0.04-1.54
Streptococcus agalactiae	6/20 (30%)	6/20 (30%)	0.03-55.96	2/19 (10%)	0.85-23.20
Ureaplasma urealyticum	5/20 (25%)	ND	ND	5/19 (26%)	0.08-1.38
Atopobium vaginae	4/20 (20%)	2/20 (10%)	21.66-36.44	4/19 (21%)	0.09-86.63
Staphylococcus aureus	3/20 (15%)	3/20 (15%)	0.33-16.68	1/19 (5%)	0.04
Staphylococcus saprophyticus	3/20 (15%)	3/20 (15%)	0.79-6.67	ND	ND
Corynebacterium coyleae	3/20 (15%)	3/20 (15%)	0.08-12.70	3/19 (16%)	0.09-1.54
Citrobacter koseri	3/20 (15%)	1/20 (5%)	99.98	3/19 (16%)	0.03-96.62
Actinotignum schaalii	2/20 (10%)	1/20 (5%)	0.28	2/19 (10%)	0.08-0.55
Aerococcus urinae	2/20 (10%)	2/20 (10%)	0.12-3.17	ND	ND
Alloscardovia omnicolens	1/20 (5%)	1/20 (5%)	24.55	ND	ND
Pseudomonas putida	1/20 (5%)	1/20 (5%)	14.72	ND	ND
Stenotrophomonas maltophilia	1/20 (5%)	1/20 (5%)	0.13	ND	ND

*Total detection in FUM of 20 participants by both methodologies.
ND - not detected; RA - relative abundance

Figure 1. Species-level community structure types of healthy FUM by culturomics.
(i) Hierarchical clustering of Bray-Curtis dissimilarity distance matrices on the relative proportions of CFU/ml within individual urine samples. (ii) Bars below dendrogram denote community structure types. (iii) Heatmap of RA of bacterial species within each urinary microbiota. Only species that are at least 1% abundant in at least one sample are shown in order of decreasing prevalence (from top to bottom). Asterisk denotes detection only by culturomics and not by amplicon sequencing.

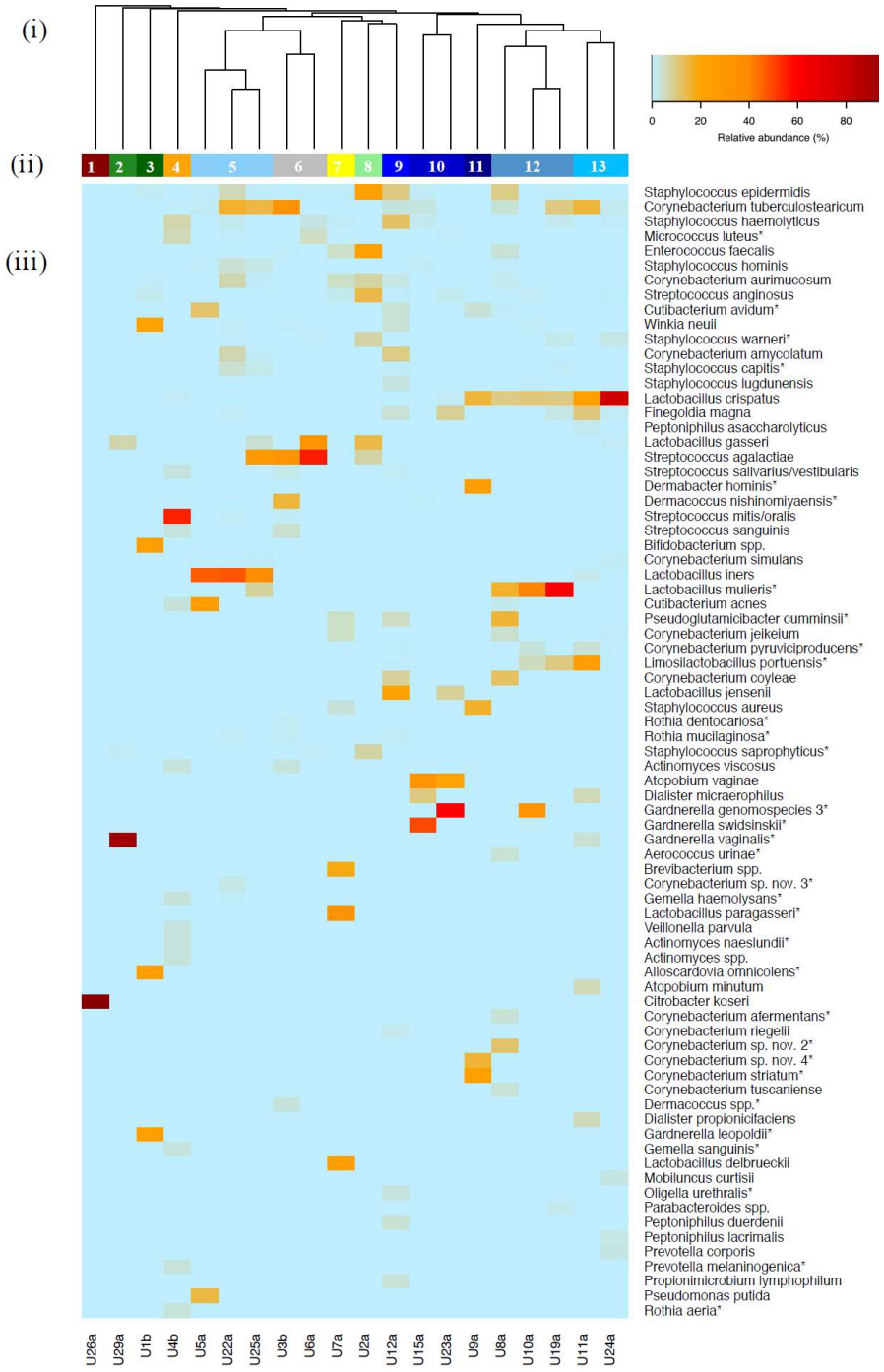


Figure 2. Species-level community structure types of healthy FUM by amplicon sequencing. (i) Hierarchical clustering of Bray-Curtis dissimilarity distance matrices on the relative proportions of reads for each OTU within individual urine samples. (ii) Bars below dendrogram denote community structure types. (iii) Heatmap of RA of bacterial species within each urinary microbiota. Only species that are at least 1% abundant in at least one sample are shown in order of decreasing prevalence (from top to bottom). Asterisk denotes detection only by amplicon sequencing and not by culturomics.

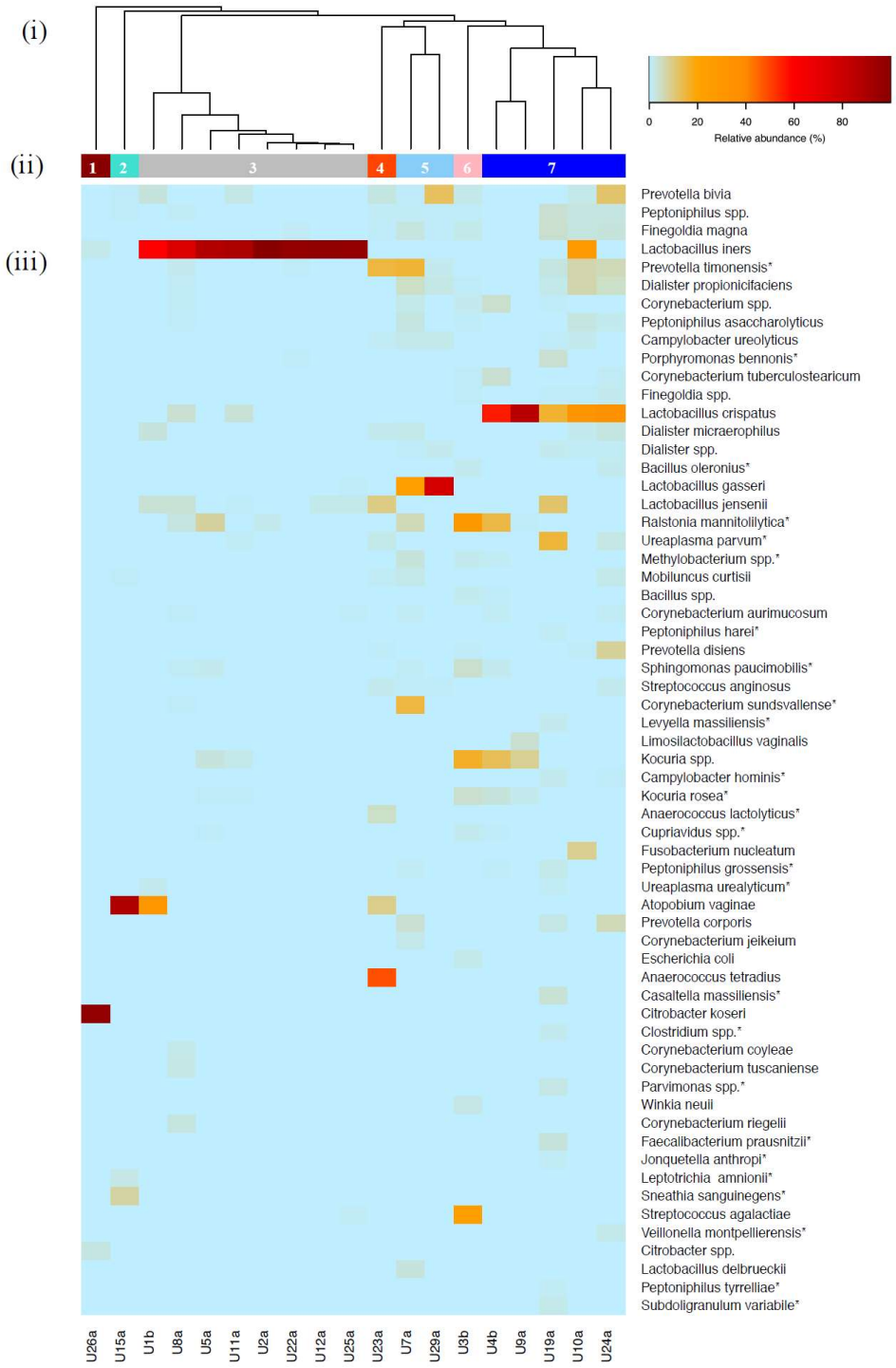


Figure 3. Biplot of the principal coordinate analysis (PCoA) based on the species-level Bray-Curtis dissimilarity matrices. Two-dimensional distances identify dissimilarities between bacterial community structures detected by culturomics and amplicon sequencing. The biplot, based on weighted average of the species scores, shows the top 10 species with the largest contributions to dissimilarities. Same colour indicates the same sample.

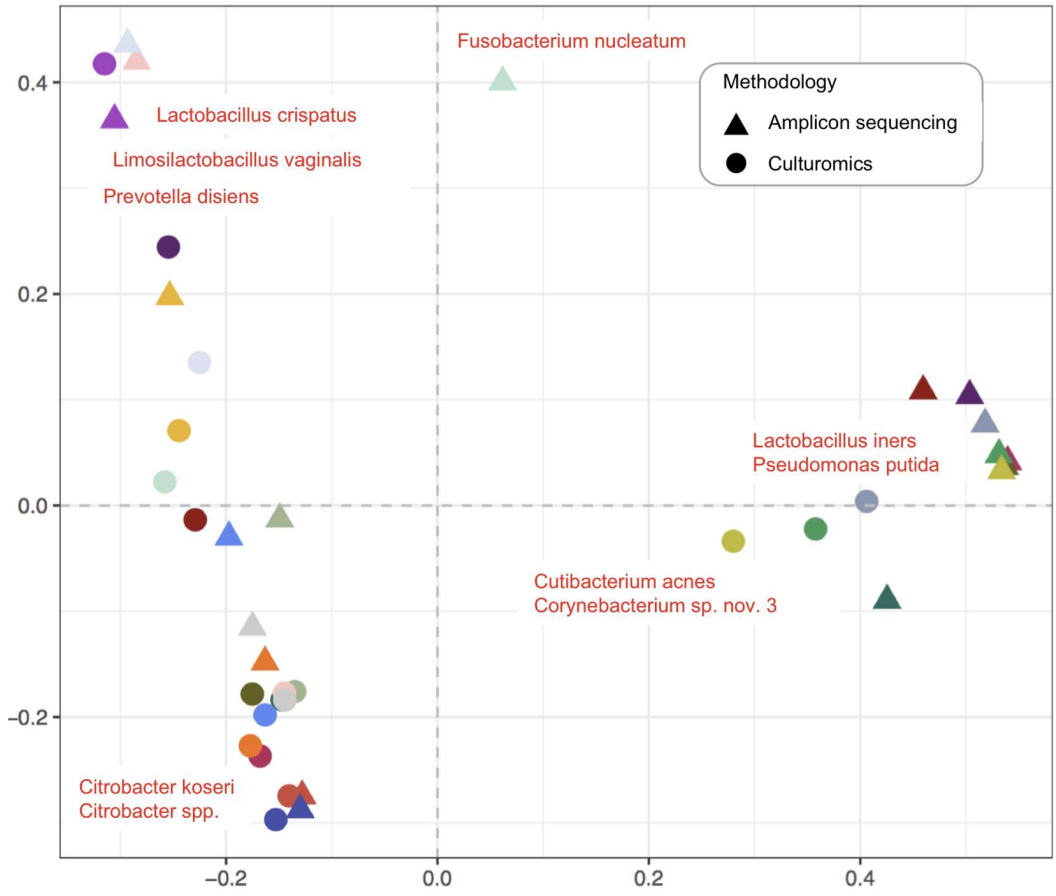
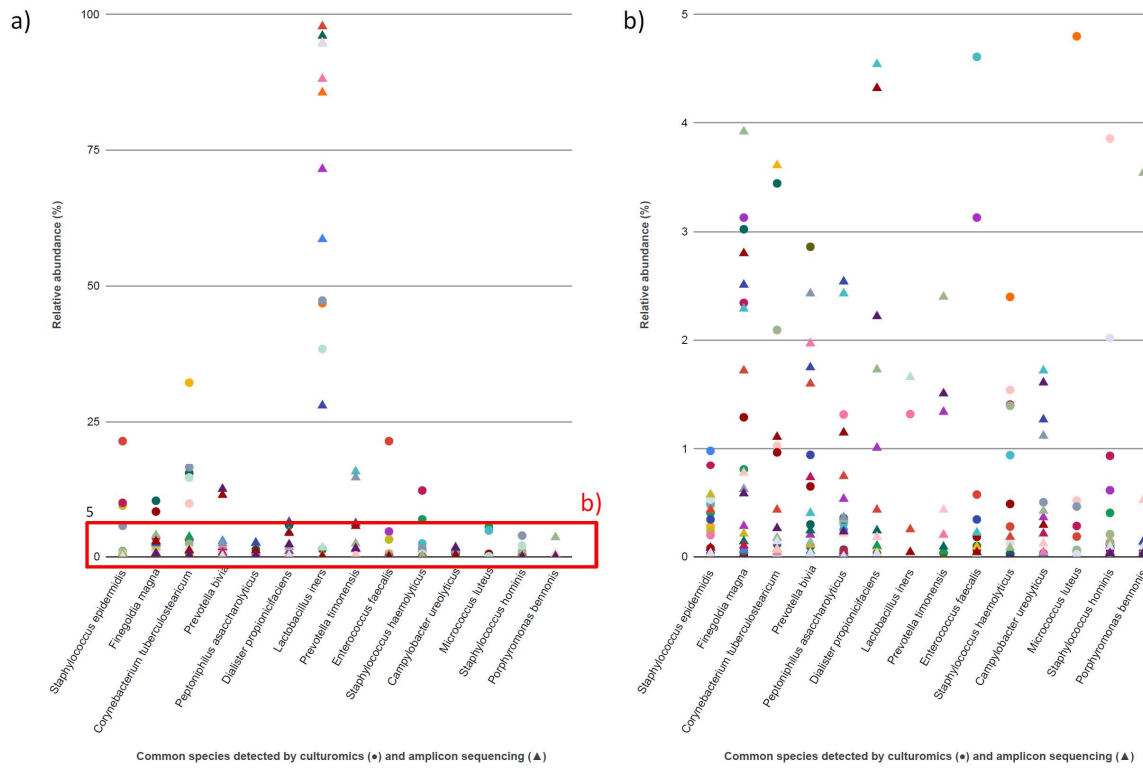


Figure 4. Common bacterial species of healthy FUM detected by culturomics and amplicon sequencing.
(a) RA per sample of species present in more than 50% of samples by culturomics and amplicon sequencing. Only species that are detected by culturomics or amplicon sequencing with at least 1% abundance in at least one sample are presented in order of decreasing prevalence (from left to right). Same colour indicates the same sample. (b) Close-up of section of Fig. 4(a) showing the RA range 0.01-5%.



Additional material associated with the manuscript entitled “Moving beyond genus to bacterial species clustering: community structure types of the healthy female urinary microbiome” is available through the following links:

Additional file 1:
https://docs.google.com/spreadsheets/d/1sj0dHIxoyqxNFiuD6cLVpRaTLSvv2E8E/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Additional file 2:
https://docs.google.com/document/d/1WuB7Z3SG8RUiLViTA05GMmATSio3sWnR/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Long-term stability of the urogenital microbiota of asymptomatic European women

Magdalena Ksiezarek, Svetlana Ugarcina-Perovic, Joana Rocha, Filipa Grosso, Luísa Peixe

[1]UCIBIO-REQUIMTE. Laboratory of Microbiology, Faculty of Pharmacy, University of Porto, Porto, Portugal

BMC Microbiology 2021 Feb 25;21(1):64.

Ksiezarek et al. BMC Microbiology (2021) 21:64
https://doi.org/10.1186/s12866-021-02123-3

BMC Microbiology

RESEARCH ARTICLE Open Access

Long-term stability of the urogenital microbiota of asymptomatic European women

Magdalena Ksiezarek, Svetlana Ugarcina-Perovic, Joana Rocha, Filipa Grosso and Luísa Peixe*

Abstract

Background: To date, information on healthy female urinary microbiota is available mostly at genus level and at one time point. However, profound species-level characterization of healthy urinary microbiome and its stability over time are essential for further correct interpretation of its role in healthy urogenital tract. In this study, we investigated female urogenital microbiome (FUM) at two timepoints (within 2.5-year interval) in young asymptomatic European women. We used culturomics with accurate isolates' identification (MALDI-TOF MS and gene markers sequencing) to understand species stability within healthy FUM.

Results: Extended culturomics of voided midstream urine sample pairs revealed a mean Shannon diversity index of 1.25 and mean of 19 species/sample (range 5–39 species; total of 115 species; 1830 isolates). High overall species variability between individuals was captured by beta diversity and a variety of community structure types, with the largest cluster characterized by *Lactobacillus crispatus*, often in combination with *Gardnerella vaginalis* or *Gardnerella* genomospecies 3. Significant FUM composition differences, related to *Finegoldia magna* and *Streptococcus anginosus,* according to smoking status were found.
A high species variability within individuals (Shannon index SD > 0.5 in 7 out of 10 sample pairs) with a mean of 29% of shared species (range 9.1–41.7%) was observed. Moreover, 4 out of 10 sample pairs clustered in the same community structure type. The stable FUM sample pairs presented high abundance of *Lactobacillus crispatus, Streptococcus agalactiae* or *Lactobacillus paragasseri* and *Bifidobacterium* spp.. Moreover, *Gardnerella vaginalis*, *Gardnerella* genomospecies 3 or *Gardnerella swidsinskii* were often maintained within individuals in high abundance.

Conclusions: Shift in species composition at two distant timepoints was frequently observed among urogenital microbiome of European asymptomatic women. This suggests possible interchange of particular species in healthy FUM and the existence of multiple health-associated FUM compositions in certain individuals.
Additionally, we provided additional evidence on resilience of particular bacterial communities and identified certain species more prone to persist in urogenital tract.
This study revealed important details on the FUM composition complexity relevant for studies aiming to understand microbiota role in the urogenital tract health and for identification of eubiotic and dysbiotic FUM.

Keywords: Microbiome, Culturomics, Species diversity, Uropathogens, Voided midstream urine

* Correspondence: lpeixe@ff.up.pt
UCIBIO-REQUIMTE. Laboratory of Microbiology, Faculty of Pharmacy, University of Porto, Porto, Portugal

Background

In the recent years, novel high-throughput culture- and DNA-based studies revealed the existence of a microbial community inhabiting the human lower urinary tract [1–8]. The majority of available observations have been made in female urogenital microbiota (FUM) composition under disease state and, simultaneously, data originated from asymptomatic controls provided a broad overview at high taxonomic levels on healthy FUM [2, 6, 9–11].

To date, a set of microbiota profiles based on dominant taxa, with interpersonal differences in bacterial load, diversity and abundance of specific bacteria, has been reported. *Lactobacillus*, *Gardnerella* and *Streptococcus* genera have been often highlighted as most prevalent healthy FUM members, in combinations with other genera such as *Staphylococcus, Corynebacterium* or *Escherichia* [2, 12–16]. A few studies indicated species such as *Lactobacillus crispatus*, *Lactobacillus jensenii, Gardnerella vaginalis* or *Streptococcus anginosus*, identified by MALDI-TOF MS analysis, as the most prevalent within healthy FUM [2, 12, 16, 17].

Since it is widely recognized in other body niches that microbiota dysbiosis, i.e., significant change in microbiota composition, may contribute to disease development [18, 19], it is important to assess the scale of microbiota compositional shifts occurring naturally in healthy urogenital tract and evaluate resilience of urogenital microbiota. To date, three-months daily assessment of female urinary microbiota demonstrated that it can be both dynamic and resilient. Moreover, changes in urinary microbiota composition may occur daily and certain shifts are associated with particular physiological or lifestyle factors, such as increased detection of *Streptococcus* and *Staphylococcus* genus associated with vaginal intercourse, or increased detection of e.g., *Corynebacterium* and *Actinomyces* during menstruation [14].

Therefore, fundamental knowledge on urogenital microbiota compositional stability remains incomplete and needs to be enlarged by long-term studies, addressing adequately the species shifts occurring in the urogenital tract and preferably eliminating factors known to alter microbiota structure.

To evaluate compositional stability of healthy FUM at two timepoints within a long period of time (2.5-year interval), we performed a comprehensive culturomic-based analysis (extended number of characterized isolates and improved methodologies for bacterial identification) of voided midstream urine samples of ten reproductive-age asymptomatic women. To the best of our knowledge, this is the first study assessing long-term FUM compositional stability at two distant timepoints in urogenital tract of asymptomatic women.

Results

Cohort overview

Ten asymptomatic reproductive-age women (24–40 years old) provided voided midstream urine samples ($n = 20$) at two time points within the 2.5-year interval. All participants were residents of Portugal, declared to have a balanced diet, and reported good or very good general health conditions according to their individual interpretation. None of the participants had symptoms or discomfort associated with their urogenital tract at either sampling time. Although some participants declared to have had UTI in the past, none of them reported to suffer from recurrent UTIs. Additionally, 2 participants acquired UTI (U7, U23) in the time between first and second sampling. One participant resigned from hormonal contraception (U4) in the interval between samplings. Three individuals reported themselves as active smokers (U9, U15, U26). Detailed demographic information about participants at the first and second sampling time is provided in Table 1.

Culturomic analysis overview

The bacterial load varied from 10^4 to 10^8 CFU/ml with maximum difference of 10^2 CFU/ml for sample pair (4 out of 10 sample pairs). A range of 17–321 (mean = 103, median = 63) isolates per sample was characterized. Identification of 1830 bacterial isolates resulted in detection of 5 phyla, 48 genera and 115 species. Overall, identified taxa distribution at phylum level was characterized by dominance of the Firmicutes (50–51% of total species for first and second sampling, respectively) and Actinobacteria (40%; 30%), and less prevalent Proteobacteria (6%; 11%), Bacteroidetes (3%; 3%) and Fusobacteria (1%; 2%). A list of species detected in each participant during first and second sampling can be found in Additional file 1: Table S1.

Diversity of healthy FUM over time

Overall, alpha diversity represented by mean Shannon index was 1.25 (standard deviation = 0.79; SD), species richness varied within range of 5 to 39 species/sample (mean = 19, SD = 8) and species evenness (Pielou's evenness index) varied within range of 0.0002 to 0.29 per sample (mean = 0.18, SD = 0.1). All values of species richness, evenness and Shannon index per each sample are presented in Additional file 1: Table S2. Species richness for sample pairs differed in a range of 2 to 18 species. Shannon index SD for sample pairs was ranging from 0.12 to 1.44. In 7 out of 10 sample pairs Shannon index SD was higher than 0.5. Graphic representation of alpha diversity measures is presented in Fig. 1.

Sample pairs presented a range of 1–12 (median of 10) species in common. Percentage of shared species within individual over time was in a range of 9.1–41.7%

Table 1 Demographic characteristics of participants

METRIC	First sampling	Second sampling
Age (years)	mean = 30.7 (SD = 4.97)	mean = 32.5 (SD = 4.97)
BMI (kg/m^2)	mean = 21.74 (SD = 2.18)	mean = 21.76 (SD = 2.24)
Smokers	30%	30%
Sexually active	100%	100%
Regular menstrual cycle	90%	90%
Previous pregnancy	40%	40%
Hormonal contraception	90%	80%
Anti-inflammatory drugs usage in week before sampling	30%	20%
UTI in the past	40%	60%

Age and BMI expressed in mean and standard deviation (SD). Remaining parameters expressed in % of positive women

(mean = 29%), with changes in their relative abundance (Fig. 2). Species observed in both samples of at least one participant, corresponded to 38 out of 115 species detected. Those included prevalent *Staphylococcus epidermidis, Micrococcus luteus, Streptococcus anginosus* and *Staphylococcus haemolyticus* (in more than 5 sample pairs) mostly present in low relative abundance. Additionally, *Lactobacillus crispatus, Gardnerella vaginalis, Gardnerella swidsinskii, Gardnerella* genomospecies 3 and *Streptococcus agalactiae* were among shared species however usually present in high relative abundance (range of 29–44% average relative abundance).

Beta diversity is presented with Bray-Curtis dissimilarity matrix (Fig. 3a) and two-dimension non-metric ordination (Fig. 3b). Most samples between individuals (6/10) were different based on Bray-Curtis dissimilarity > 0.5. In NMDS ordination (stress value 0.2) U26 sample pair was observed as the more dissimilar pair and was previously characterized by particularly low species richness. ANOSIM test revealed statistically significant differences between bacterial communities and smoking status of the individuals ($R = 0.25$, $p = 0.03$). Multilevel pattern analysis identified 2 species associated with smoking status and FUM variance, namely *Finegoldia magna* ($p < 0.05$) and *Streptococcus anginosus* ($p < 0.05$). Remaining factors tested did not show statistically significant microbiota composition differences.

Healthy FUM community structure types over time

FUM structure types identified in asymptomatic women are presented in Fig. 4. Ten community structure types (dendrogram representing samples hierarchical

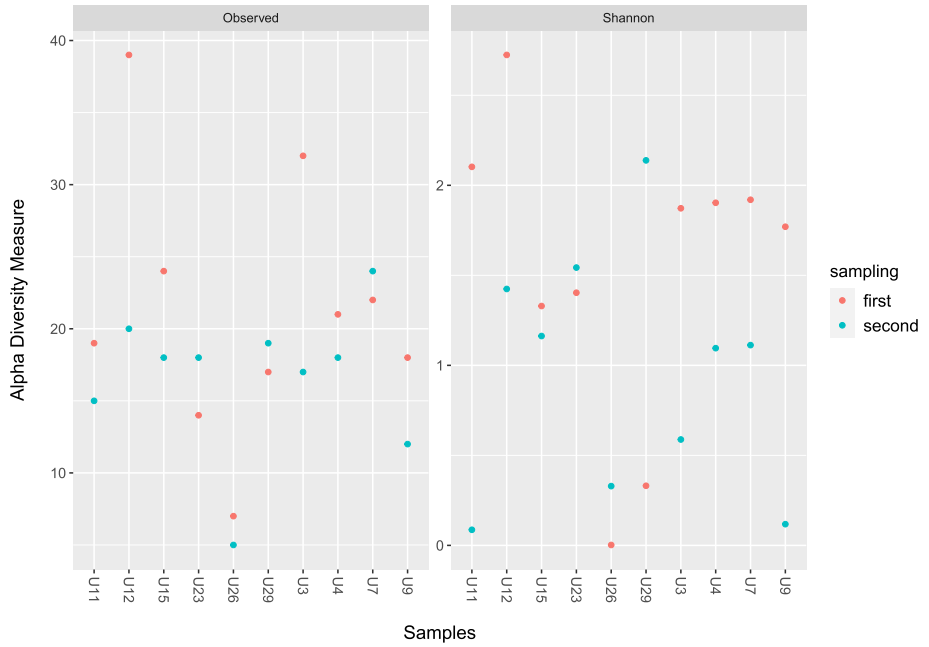


Fig. 1 Alpha diversity among samples measured by observed number of species and Shannon index

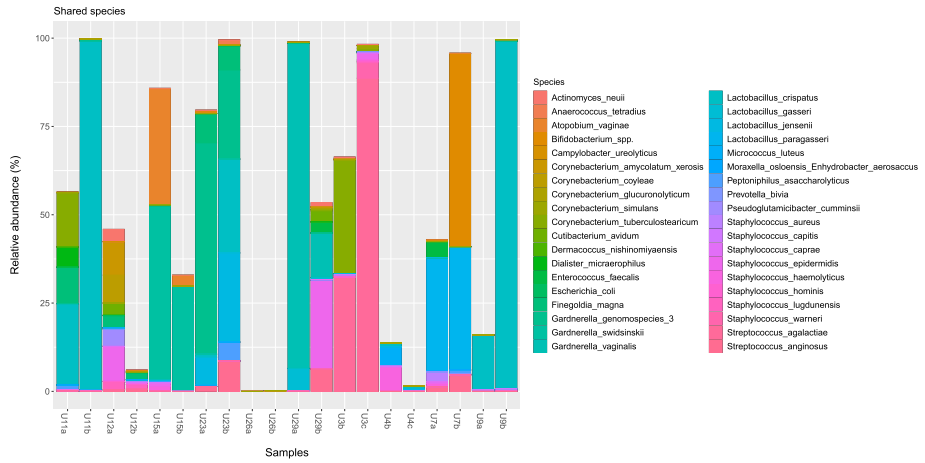


Fig. 2 Summarized relative abundance (%) of shared species in each sample. Sample pairs are ordered next to each other

clustering based on species level identification is available in Additional file 2: Fig. S1) were identified. The largest cluster (6 out of 20 samples) was characterized by *Lactobacillus crispatus,* often in combination with *Gardnerella* spp. namely, *Gardnerella vaginalis* or *Gardnerella* genomospecies 3. The other more common clusters presented abundant *Streptococcus agalactiae* or abundant *Bifidobacterium* spp. and *Lactobacillus paragasseri*, with other Gram-positive bacteria in lower abundance. Summary of clusters with different bacterial combinations characterizing community structure types are presented in Table 2.

Gardnerella vaginalis and the recently described *Gardnerella* species (*Gardnerella swidsinskii* or *Gardnerella* genomospecies 3) were observed in 5 individuals, usually with single species per individual (4/5) and *Gardnerella vaginalis* was the more prevalent one. Moreover, the recently described *Lactobacillus mulieris*, originally isolated from other cohort of our FUM study [20], was also identified in one individual (U26b). Furthermore, two putatively new species close to *Limosilactobacillus vaginalis* were also depicted in 2 individuals (U9a and U11a) (data not shown).

Different community structure types were observed within 6 out of 10 sample pairs, with changes related to genus or species presence and abundance [e.g., *Lactobacillus jensennii, Staphylococcus haemolyticus, Staphylococcus epidermidis* type (U12a) converted to *Lactobacillus crispatus* type (U12b) or *Gardnerella swidsinskii, Atopobium vaginae* and *Dialister microaerophilus* type (U15a) converted to *Gardnerella vaginalis*, *Bifidobacterium* spp., *Cutibacterium avidum* community structure type (U15b)]. Noteworthy, an individual with highly abundant Enterobacteriaceae members (U26) presented a shift in the community structure type (*Citrobacter koseri* to *Escherichia coli*) and shared just one species, the *Lactobacillus jensenii*.

Stable community structure types within individuals were observed in 4 (U3, U7, U11, U23) out of 10 individuals and were represented by *Lactobacillus crispatus, Bifidobacterium* spp. with *Lactobacillus paragasseri* or *Streptococcus agalactiae,* in combination with other Gram-positive bacteria.

Interestingly, two of those sample pairs (U7 and U23) are from individuals that acquired UTI, followed by antibiotic treatment, in the interval between samplings. Maintenance of 9 species (23.7%) including *Lactobacillus paragasseri* was observed in U7 sample pair, and 10 species (41.7%), including *Lactobacillus crispatus* and *Gardnerella* genomospecies 3 was observed in U23 sample pair, despite changes in their relative abundance.

Additional analysis based on genus level revealed higher stability, with 6 out of 10 individuals comprising sample pairs in the same clusters. Overall, the highest number of samples belonged to the cluster represented by *Lactobacillus* genus, followed by clusters characterized by abundant *Gardnerella* or *Streptococcus* genera. Heatmap and dendrogram representing hierarchical clustering at genus level is available in Additional file 2: Fig. S2 and Fig. S3, respectively.

Discussion

In this study, using a comprehensive and extended culturomic approach, we enlarged the knowledge on diversity of FUM and its bacterial community structures in asymptomatic individuals. We also demonstrated FUM stability in two timepoints within long period of time.

Most of FUM studies are based on genus level classification and on most dominant taxa [12–16], while particular functional characteristics are often species

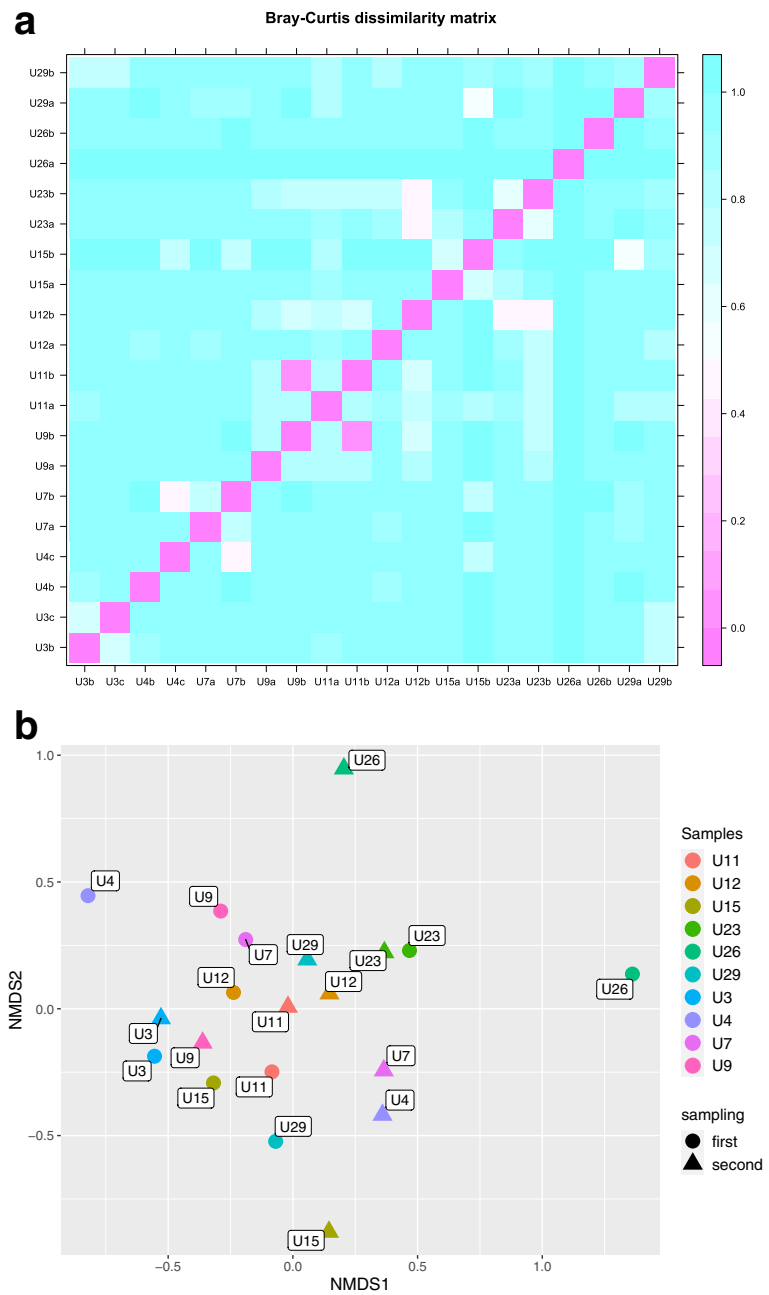


Fig. 3 Beta diversity among samples. **a** Heatmap representing Bray-Curtis dissimilarity matrix between the samples. **b** Relationship between samples presented by 2-dimension NMDS ordination based on Bray-Curtis distance matrix, with 0.2 stress value

specific, e.g., antimicrobials or metabolites production [21, 22].

In our study, due to the higher number of morphotypes characterized per sample and the higher taxonomic resolution conferred by genotypic markers used (e.g, *pheS*), we captured a slightly higher amount of species, however reflecting a similar diversity as previous reports [2, 12]. For instance, *Gardnerella swidsinskii* and *Gardnerella* genomospecies 3 were here identified for the first time within urogenital microbiota of asymptomatic individuals, in addition to *Gardnerella vaginalis*, suggesting their frequent occurrence in healthy FUM. Moreover, higher species diversity within Lactobacillaceae was captured with the recently described *Lactobacillus mulieris* infrequently observed among the samples tested [20].

Although alpha diversity measures (mean Shannon index < 1.5) suggest that FUM is less diverse than other

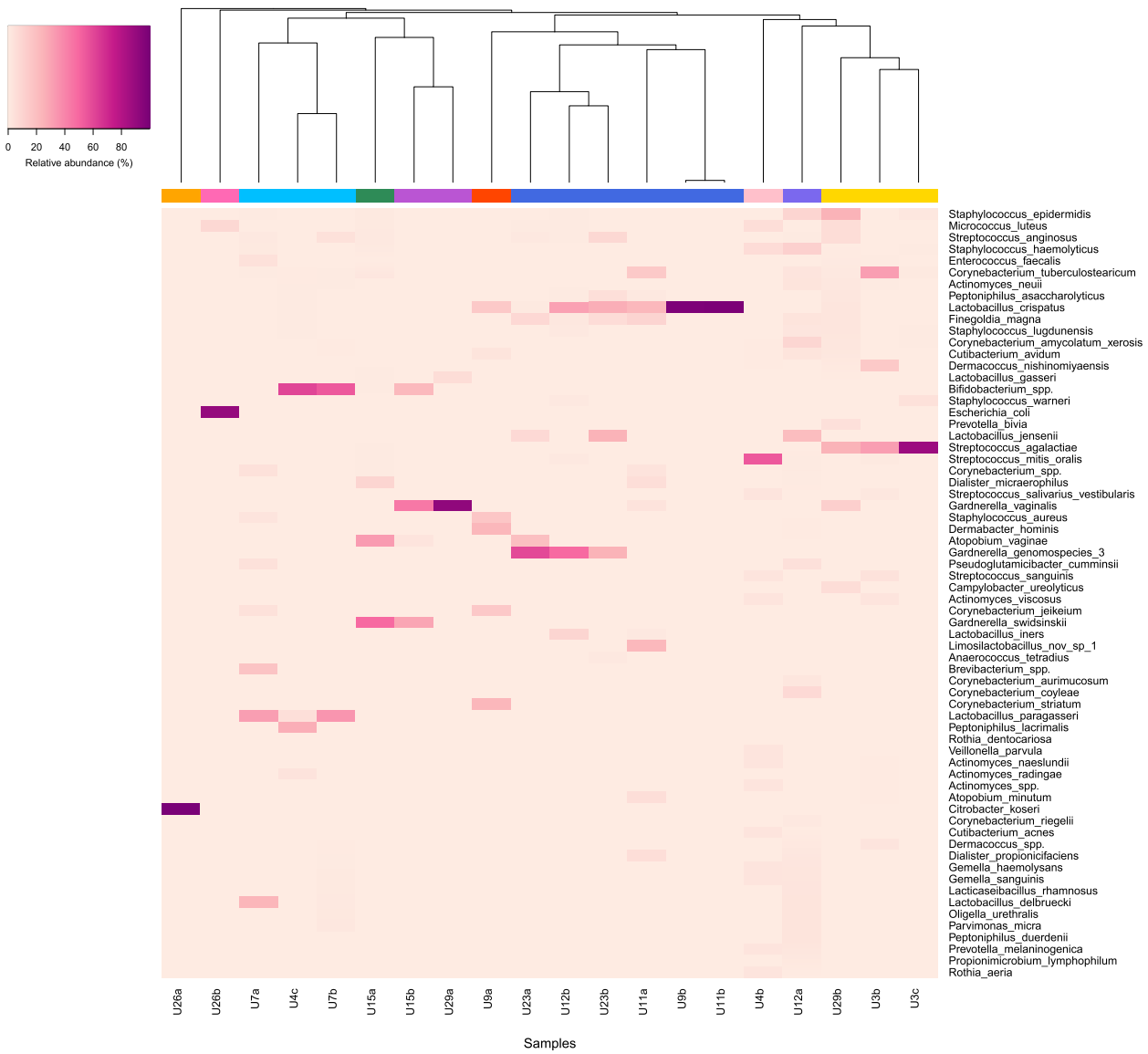


Fig. 4 Heatmap based on abundance (%) of species detected at least once in abundance more than 1%. Species are ordered in decreasing prevalence. Dendrogram presents hierarchical clustering of microbiota profiles into community structure types, based on 0.8 cutoff. Colorful bar below the dendrogram stands for different community structure types

human body niches [17], we observed a high overall species variability and diverse community structure types. Hierarchical community clustering based on Bray-Curtis dissimilarity matrix enabled to capture various community structure types based on bacterial species combinations. The largest cluster was characterized by the commonly described *Lactobacillus crispatus*, often in combination with *Gardnerella vaginalis* or *Gardnerella* genomospecies 3. *Gardnerella* species were also observed in other community structure types, usually with only one species present in individual FUM, confirming previously reported high occurrence of this genus in urinary microbiota [16, 17]. The remaining community structure types were characterized by many diverse species, including species commonly associated with UTI.

Of interest, FUM composition differences, related to *Finegoldia magna* and *Streptococcus anginosus,* according to smoking status were observed, requiring further validation. Those species were previously associated with urinary symptoms presence and/or severity [17, 23].

Within individuals, FUM changes were frequently detected at two distant timepoints (e.g., *Gardnerella swidsinskii, Atopobium vaginae and Dialister microaerophilus* community type converted to *Gardnerella vaginalis, Bifidobacterium* spp., *Cutibacterium avidum* type). This data suggests, the possibility of interchange between certain

Table 2 Summary of community structure types detected within healthy FUM

Community structure type	Characteristic species combination (ordered by decreasing relative abundance, only top 3 shown)	Samples	Shannon index (mean, SD – standard deviation)
1	*Citrobacter koseri* *Enterococcus faecalis* *Lactobacillus jensenii*	U26a	0.002
2	*Escherichia coli* *Micrococcus luteus* *Lactobacillus jensenii*	U26b	0.33
3	*Bifidobacterium* spp. *Lactobacillus paragasseri* *Enterococcus faecalis*	U4c U7a, U7b	1.38 (SD 0.47)
4	*Gardnerella swidsinskii* *Atopobium vaginae* *Dialister microaerophilus*	U15a	1.33
5	*Gardnerella vaginalis* *Bifidobacterium* spp. *Cutibacterium avidum*	U15b U29a	0.75 (SD 0.59)
6	*Corynebacterium striatum* *Dermabacter hominis* *Staphylococcus aureus*	U9a	1.77
7	*Lactobacillus crispatus*	U9b U11a, U11b U12b U23a, U23b	1.11 (SD 0.84)
8	*Streptococcus mitis/oralis* *Staphylococcus haemolyticus* *Micrococcus luteus*	U4b	1.90
9	*Lactobacillus jensenii* *Staphylococcus haemolyticus* *Staphylococcus epidermidis*	U12a	2.72
10	*Streptococcus agalactiae* *Staphylococcus epidermidis* *Corynebacterium tuberculostearicum*	U3b, U3c U29b	1.53 (SD 0.83)

bacterial groups that might share common metabolic functions. Additionally, few communities that maintained their composition at two timepoints were detected and characterized by combinations of *Lactobacillus crispatus, Bifidobacterium* spp. with *Lactobacillus paragasseri* or *Streptoccocus agalactiae*. Further evidence on the resilience of those communities is the maintenance of *Lactobacillus crispatus* or *Bifidobacterium* spp. with *Lactobacillus paragasseri* community structure type in women after antibiotic treatment for a UTI. Studying short-term FUM dynamics, Price et al. also reported resilience of lower urinary tract microbiota in communities with dominance of e.g., *Lactobacillus* or *Lactobacillus* and *Gardnerella* combination [14]. Similarly to their findings, at the genus level, changes in relative abundance of *Lactobacillus*, *Gardnerella* or *Streptococcus* were observed in three individuals, leading to change in community structure type [14].

Moreover, maintenance of *Gardnerella* species, *Lactobacillus gasseri* and *Lactobacillus jensenii* was also observed in certain sample pairs, for which different FUM composition was observed at two time points tested. The protective role of particular strains belonging to *Lactobacillus jensenii* [24, 25], can possibly contribute to the health maintenance of an individual with high abundance of Enterobacteriaceae, including an uropathogenic ST131 *Escherichia coli* (UPEC) according to its gene content [26] (unpublished data). This data also highlights the relevance of strain level characterization to understand FUM role in health and disease, as previously noticed [27, 28].

It is notable that our study focused on the group of well characterized young age European women, contrary to most urinary and urogenital microbiota studies [4, 6, 9–12, 17]. Moreover, samples were collected only on the 3rd week of women's menstrual cycle to prevent interferences (e.g., menstrual discharge), which was recently demonstrated by Price et al., [14] as factor influencing microbiota composition. Women providing samples over time were under similar lifestyle and physiological conditions.

Although our cohort included a small number of participants, we believe that detailed description of a

small group of women may substantially enlarge knowledge originated from studies with large scale cohorts but less detailed analysis. Additionally, we are aware that our choice of using voided urine samples brings a risk of genital contamination, thus representing more accurately the urogenital microbiome. However, Chen et al., recently reported that prevalence of *Lactobacillus* and *Gardnerella* were detected with equal sensitivity in voided urine and urine collected by catheter [15]. Moreover, we are convinced that characterizing microbiota from samples routinely used for screening and diagnosis is highly valuable to facilitate accurate results interpretation and potentiate their use in future diagnostics. Undoubtedly, knowing also genital tract microbiota composition would be highly beneficial to enlarge our understanding of health-associated and pathogenic strains similarity between urinary and vaginal microbiota [28].

Another potential limitation could be the lack of culture-independent DNA-based data, however current diagnostic procedures for urine samples are based on culturing. Moreover, culturomic approach is necessary to assess alive bacterial communities and provide isolates for further characterization.

Conclusions

In this study, we characterized species level stability of the FUM of reproductive age women at two timepoints within a long period of time. We present further evidence that FUM can be dynamic over time and multiple FUM communities may be associated with urogenital tract of some asymptomatic individuals.

Additionally, at 2 sampling points with long time interval, we identified community structure types that seem to indicate persistence of certain species in healthy FUM and provides further evidence on resilient bacterial communities.

We also revealed previously unknown diverse community structure types in healthy FUM. These findings may challenge further identification of eubiotic and dysbiotic states and consequently, diagnostic and treatment strategies for urogenital and urinary tract pathologies.

Moreover, our results support that culturomic analysis with the large-scale isolates characterization is a valuable tool for microbiota diversity description and provides isolates for further analysis.

The future studies focusing on strain level characterization to discern functions contributing to health maintenance in urogenital tract are required. Furthermore, healthy FUM structures characterized by highly abundant species commonly associated with UTI, as here reported, highlight the need for a better understanding of microbiota-host interactions.

Methods

Participant information

Ten asymptomatic women (24–40 years old) were recruited to voluntarily participate in the FUM study conducted at the Faculty of Pharmacy, University of Porto, Portugal, at two time points. All women provided informed written consent for participation in the study and fulfilled a detailed questionnaire containing demographic, health-associated and lifestyle information before both sampling times. The study was developed according to the Helsinki Declaration principles and the protocol was submitted and approved by the Ethical Commission of Faculty of Pharmacy, University of Porto. Inclusion criteria at both sampling times were no pregnancy, no antibiotic treatment in the previous month and no current symptoms or diagnosis of urinary tract infection (UTI).

Sample collection

Ten women provided first morning voided midstream urine samples at two time points (total number of samples = 20; sample pairs = 10). Interval between first and second sample collection varied in a range of 11 and 28 months, depending on donors' availability. Samples were collected in the 3rd week of the menstrual cycle. Participants also provided vaginal swab collected prior to urine sample collection (data not shown). Detailed verbal and written instructions were provided to each woman before sampling. Flyer included written and graphical information on proper wash prior to sampling and vaginal swab collection in order to minimize possible vulvovaginal contamination.

Sample analysis

Urine samples were subjected to an extended culturomic analysis within 2 h from sample collection. The extended culturomic protocol is a modification of the expanded quantitative urine culture (EQUC) previously described [2]. Culture included plating of 100 μl of urine into 140 mm diameter-Petri dishes. Protocol included Columbia Agar with 5% sheep blood (Biogerm, Portugal) and chromogenic agar typically used for uropathogens detection (HiCrome UTI, HiMedia, India) supplemented with previously described nutrients i.e., 2% (w/v) gelatin, 0.5% (w/v) yeast extract, 0.1% (w/v) starch, 0.1% (w/v) glucose and 0.1% (v/v) Tween 80 [29, 30]. Incubation at 37 °C for 48 h was performed under aerobic, microaerophilic, and anaerobic atmospheric conditions for Columbia Agar plates, and aerobic and microaerophilic condition for supplemented chromogenic agar (GENbox MICROAER and GENbox ANAER, bioMérieux, France). Besides culture, dipstick test (Combur-Test, Roche) and microscopy examination were performed. Additionally, when during microscopic examination a higher bacterial load

was suspected, diluted volume of urine was plated and incubated following the same protocol. Each colony morphotype was quantified to obtain a most approximate number of colony forming units per milliliter (CFU/ml) and up to 5 colonies of the same morphotype were isolated, stored and subjected to identification. Multiple representatives were isolated to ensure reliable microbiota profiling. Many species belonging to genera widely present within urogenital microbiota e.g., *Lactobacillus*, *Staphylococcus*, *Corynebacterium* have very similar or equal colony morphology, thus their diversity may be easily underestimated.

Isolates identification

Firstly, all isolates were subjected to matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) VITEK MS system (bioMérieux, France), using in-vitro diagnostic database version 3.0. In case of no identification by MALDI-TOF MS isolates were subjected to 16S rRNA gene sequencing and/or other suitable genotypic biomarkers (*pheS*, *rpoB*, *recN*) [31–34]. Additionally, due to recent taxonomic reclassification of genus *Gardnerella*, all isolates identified as *Gardnerella vaginalis* by MALDI-TOF MS were subjected to *cpn60* gene sequencing [35, 36].

Statistical analysis

Continuous and categorical variables referring to participants' demographic and lifestyle characteristics were interpreted based on descriptive statistics. All community analyses were based on relative abundance (%) calculated as the CFU percentage of identified species from total CFU/ml count. Alpha-diversity (within-samples diversity) represented by the number of observed species and Shannon index (increases when species richness and evenness increase), was performed and visualized using phyloseq package (version 1.30.0) [37] R version 3.6.1 [38]. Pielou's evenness index was calculated; evenness refers to the distribution of species in terms of relative abundance. Pielou's index comprises values between 0 and 1, where lower values stand for lower degree of evenness. Figure representing cumulative relative abundance of shared species was created with phyloseq package. Beta-diversity (between-samples diversity) was represented by Bray-Curtis dissimilarity matrix with values comprised between 0 and 1, where 0 states for high similarity and 1 for high dissimilarity, and 2-dimension Non-metric Multi-dimensional Scaling (NMDS) with samples ordination based on dissimilarity matrix. Stress value was measured using vegan (version 2.5.6) [39] package. Heatmap representing Bray-Curtis dissimilarity matrix was generated with lattice package (version 0.20.38) [40]. NMDS plot was performed using phyloseq package. Statistical significance for age, body mass index, smoking status, previous UTI, usage of anti-inflammatory drugs in a week before sampling, hormonal contraceptives usage, previous pregnancies and presence or absence of menstrual cycle was accessed with analysis of similarities (ANOSIM) performed with vegan package, based on Bray-Curtis dissimilarity matrix. ANOSIM analysis result in significance level (p value) and R value where number close to 0 stands for similarity, and close to 1 stand for dissimilarity. Multilevel pattern analysis for identification of bacterial species responsible for community divergence was accessed using indicspecies package (version 1.7.9) [41]. A heatmap including a dendrogram for hierarchical clustering with cutoff value of 0.8 to define the clusters was generated using vegan (version 2.5.6) and gplots (version 3.0.1.2) [42] R packages. Hierarchical clustering was performed using unweighted pair group method with arithmetic mean (UPGMA) based on Bray-Curtis dissimilarity matrix. Additionally, ggplot2 (version 3.2.1) package [43] was used for data visualization.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-021-02123-3.

Additional file 1: Table S1. Species identified in each participant during first and second sampling and their relative abundance (%). **Table S2.** Number of observed species, Pielou's evenness index and Shannon index per each sample.

Additional file 2: Figure S1. Dendrogram representing samples hierarchical clustering based on species level identification. A cutoff value of 0.8 was used to define the clusters (dashed blue line). **Figure S2.** Heatmap based on abundance (%) of genera detected. Dendrogram presents clustering of microbiota profiles into community structure types, based on 0.8 cutoff. Colorful bar below the dendrogram stands for different community structure types. **Figure S3.** Dendrogram representing samples hierarchical clustering based on genus level identification.

Abbreviations

FUM: Female urogenital microbiota; UTI: Urinary tract infection; CFU: Colony forming units; MALDI-TOF MS: Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; NMDS: Non-metric Multi-dimensional Scaling; SD: Standard deviation; ANOSIM: Analysis of similarities

Acknowledgements

The authors would like to thank all women voluntarily providing the samples and participating in this study and also thank Helena Ramos and Paulo Pinto (Hospital Geral de Santo António, Porto, Portugal) for their technical support with MALDI-TOF MS-based microbial identification. The bioMérieux (Portugal, Lda) provided equipment and material for MALDI-TOF MS analysis, and had no role in the study design, data collection and analysis, decision to publish, or writing of the manuscript. The authors would also like to thank Teresa Gonçalves Ribeiro for the assistance with *Gardnerella* species identification.

Authors' contributions

MK, SUP, JR, FG and LP designed the study and supervised participant recruitment. MK and SUP processed the samples and collected the raw data. MK and SUP performed isolates identification. MK performed data analysis, interpretation, and visualization. MK wrote the manuscript. All authors contributed significantly to manuscript revision and approved the final version of the manuscript.

Funding
This work received financial support from Applied Molecular Biosciences Unit - UCIBIO, which is financed by national funds from FCT/MCTES [UID/Multi/04378/2019, UIDP/04378/2020 and UIDB/04378/2020]. MK is supported by Fundação para a Ciência e Tecnologia, I.P. - FCT [SFRH/BD/132497/2017]. SUP was supported from project NORTH-01-0145-FEDER-000024, JR received support from ICETA [UID/MULTI/04378/2013] and FG is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0034]. The funding bodies had no role in the design of the study; collection, analysis and interpretation of data; writing the manuscript.

Availability of data and materials
The raw datasets generated and analyzed during the current study and accession numbers for DNA sequences deposited in NCBI database are available in the GitHub repository (https://github.com/magksi/FUM_stability.git).

Ethics approval and consent to participate
Approval of the study was obtained from the Faculty of Pharmacy (University of Porto, Porto, Portugal) Ethics Committee. Procedures performed in the study were all in accordance with the ethical standards of the institutional and national research committee, with the 1964 Helsinki Declaration, and its later amendments. All individual participants included in the study had given written informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare that there are no conflicts of interest.

Received: 3 September 2020 Accepted: 9 February 2021
Published online: 25 February 2021

References

1. 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; 50:1376–83.
2. 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; 52:871–6.
3. Neugent ML, Hulyalkar NV, Nguyen VH, Zimmern PE, De Nisco NJ. Advances in understanding the human urinary microbiome and its potential role in urinary tract infection. mBio. 2020;11:e00218–20.
4. Karstens L, Asquith M, Davin S, Stauffer P, Fair D, Gregory WT, et al. Does the urinary microbiome play a role in urgency urinary incontinence and its severity? Front Cell Infect Microbiol. 2016;6:78.
5. Karstens L, Asquith M, Caruso V, Rosenbaum JT, Fair DA, Braun J, et al. Community profiling of the urinary microbiota: considerations for low-biomass samples. Nat Rev Urol. 2018;15:735–49.
6. Khasriya R, Sathiananthamoorthy S, Ismail S, Kelsey M, Wilson M, Rohn JL, et al. Spectrum of bacterial colonization associated with urothelial cells from patients with chronic lower urinary tract symptoms. J Clin Microbiol. 2013; 51:2054–62.
7. Ackerman AL, Underhill DM. The mycobiome of the human urinary tract: potential roles for fungi in urology. Ann Transl Med. 2017;5:31.
8. Pearce MM, Zilliox MJ, Rosenfeld AB, Thomas-White KJ, Richter HE, Nager CW, et al. The female urinary microbiome in urgency urinary incontinence. Am J Obstet Gynecol. 2015;213:347.e1–11.
9. Wu P, Chen Y, Zhao J, Zhang G, Chen J, Wang J, et al. Urinary microbiome and psychological factors in women with overactive bladder. Front Cell Infect Microbiol. 2017;7:488.
10. Gill K, Kang R, Sathiananthamoorthy S, Khasriya R, Malone-Lee J. A blinded observational cohort study of the microbiological ecology associated with pyuria and overactive bladder symptoms. Int Urogynecol J. 2018;29:1493–500.
11. Curtiss N, Balachandran A, Krska L, Peppiatt-Wildman C, Wildman S, Duckett J. A case controlled study examining the bladder microbiome in women with overactive bladder (OAB) and healthy controls. Eur J Obstet Gynecol Reprod Biol. 2017;214:31–5.
12. Coorevits L, Heytens S, Boelens J, Claeys G. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. Eur J Clin Microbiol Infect Dis. 2017;36:635–9.
13. Curtiss N, Balachandran A, Krska L, Peppiatt-Wildman C, Wildman S, Duckett J. Age, menopausal status and the bladder microbiome. Eur J Obstet Gynecol Reprod Biol. 2018;228:126–9.
14. 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. https://doi.org/10.1128/mBio.00475-20.
15. Chen YB, Hochstedler B, Pham TT, Acevedo Alvarez M, Mueller ER, Wolfe AJ. The urethral microbiota - a missing link in the female urinary microbiota. J Urol. 2020. https://doi.org/10.1097/JU.0000000000000910.
16. Price TK, Hilt EE, Thomas-White K, Mueller ER, Wolfe AJ, Brubaker L. The urobiome of continent adult women: a cross-sectional study. BJOG. 2020; 127:193–201.
17. 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;5:e01283–14.
18. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569:655–62.
19. Wilkins LJ, Monga M, Miller AW. Defining Dysbiosis for a cluster of chronic diseases. Sci Rep. 2019;9:1–10.
20. Rocha J, Botelho J, Ksiezarek M, Perovic SU, Machado M, Carriço JA, et al. *Lactobacillus mulieris* sp. nov., a new species of *Lactobacillus delbrueckii* group. Int J Syst Evol Microbiol. 2020;70(3):1522–7.
21. Edelman SM, Lehti TA, Kainulainen V, Antikainen J, Kylväjä R, Baumann M, et al. Identification of a high-molecular-mass lactobacillus epithelium adhesin (LEA) of *Lactobacillus crispatus* ST1 that binds to stratified squamous epithelium. Microbiology. 2012;158(Pt 7):1713–22.
22. Ojala T, Kankainen M, Castro J, Cerca N, Edelman S, Westerlund-Wikström B, et al. Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*. BMC Genomics. 2014;15:1070.
23. Fok CS, Gao X, Lin H, Thomas-White KJ, Mueller ER, Wolfe AJ, et al. Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. Int Urogynecol J. 2018;29:1765–71.
24. Petrova MI, Lievens E, Malik S, Imholz N, Lebeer S. *Lactobacillus* species as biomarkers and agents that can promote various aspects of vaginal health. Front Physiol. 2015;6:81.
25. Sihra N, Goodman A, Zakri R, Sahai A, Malde S. Nonantibiotic prevention and management of recurrent urinary tract infection. Nat Rev Urol. 2018;15:750–76.
26. Spurbeck RR, Dinh PC, Walk ST, Stapleton AE, Hooton TM, Nolan LK, et al. *Escherichia coli* isolates that carry *vat*, *fyuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. Infect Immun. 2012;80:4115–22.
27. 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;11. https://doi.org/10.3389/fmicb.2020.02094.
28. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, et al. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. Nat Commun. 2018;9. https://doi.org/10.1038/s41467-018-03968-5.
29. Alves P, Castro J, Sousa C, Cereija TB, Cerca N. *Gardnerella vaginalis* outcompetes 29 other bacterial species isolated from patients with bacterial vaginosis, using in an in vitro biofilm formation model. J Infect Dis. 2014; 210:593–6.
30. Man JCD, Rogosa M, Sharpe ME. A medium for the cultivation of *Lactobacilli*. J Appl Bacteriol. 1960;23:130–5.
31. Héritier C, Poirel L, Nordmann P. Genetic and biochemical characterization of a chromosome-encoded carbapenem-hydrolyzing ambler class D beta-lactamase from *Shewanella algae*. Antimicrob Agents Chemother. 2004;48: 1670–5.
32. Naser SM, Thompson FL, Hoste B, Gevers D, Dawyndt P, Vancanneyt M, et al. Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus species* based on *rpoA* and *pheS* genes. Microbiology. 2005;151:2141–50.
33. Mellmann A, Becker K, von Eiff C, Keckevoet U, Schumann P, Harmsen D. Sequencing and staphylococci identification. Emerg Infect Dis. 2006;12:333–6.

34. Ribeiro TG, Novais Â, Branquinho R, Machado E, Peixe L. Phylogeny and comparative genomics unveil independent diversification trajectories of *qnrB* and genetic platforms within particular *Citrobacter* species. Antimicrob Agents Chemother. 2015;59:5951–8.
35. Hill JE, Albert AYK, VOGUE Research Group. Resolution and co-occurrence patterns of *Gardnerella leopoldii*, *Gardnerella swidsinskii*, *Gardnerella piotii* and *Gardnerella vaginalis* within the vaginal microbiome. Infect Immun. 2019. https://doi.org/10.1128/IAI.00532-19.
36. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P. Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. Int J Syst Evol Microbiol. 2019;69:679–87.
37. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217.
38. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2018. https://www.R-project.org/
39. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: community ecology package; 2019. https://CRAN.R-project.org/package=vegan
40. Sarkar D. Lattice: multivariate data visualization with R. New York: Springer; 2008. ISBN 978-0-387-75968-5
41. De Caceres M, Legendre P. Associations between species and groups of sites: indices and statistical inference. Ecology. 2009; http://sites.google.com/site/miqueldecaceres/.
42. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. Gplots: various R programming tools for plotting data; 2020. https://CRAN.R-project.org/package=gplots
43. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

Publisher's Note

Additional material associated with the manuscript entitled “Long-term stability of the urogenital microbiota of asymptomatic European women” is available through the following links:

Supplementary Material:
https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-021-02123-3#Sec15

Step towards understanding interconnection between vaginal and urinary tract microbiome

Magdalena Ksiezarek[1,2], Svetlana Ugarcina Perovic[3], Joana Rocha[4], Elisabete Cappelli[1,2], Marcia Sousa[1,2], Teresa Goncalves Ribeiro[1,2], Filipa Grosso[1,2], *Luisa Peixe[1,2]

[1]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[2]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

[3]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

[4]UCGenomics/GenomePT, Laboratório de Sequenciação e Genómica Funcional da Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal

Manuscript in preparation

ABSTRACT

With the increasing knowledge on the composition of microbial communities residing in the urinary tract (UT) it became more clear the need to better understand the relationship and reciprocity between the vaginal and UT microbiomes.

In this study we aimed to analyze microbiome community state types (CSTs) from 29 vaginal samples (20 women) and compare them with their paired UT microbiome data. In ten women the sampling procedures were performed in two different moments (within 2.5-year interval), which allowed to investigate the long-term stability of vaginal microbiome.

Overall, vaginal and urinary microbiome were composed of highly similar taxa and highly similar structure (based on alpha and beta diversity indices). However, the specific differences might be observed regarding species abundance and co-occurrence patterns. For instance, highly abundant *Gardnerella* genomospecies 3, *Citrobacter koseri* or *Escherichia coli* could be more specific for urinary tract microbiome while e.g., *Limosilactobacillus urinaemulieris* or *Corynebacterium* putative novel species 5 could be more specific to vaginal microbiome. Moreover, certain species combinations demonstrated opposite pattern in both niches e.g., positive correlation for *Lactobacillus jensenii* and *Lactobacillus iners* in vaginal microbiome and negative in urinary tract microbiome. We also demonstrated that, besides currently recognized 4 *Lactobacillus* CST (*L. crispatus, L. gasseri, L. jensenii, L. iners*), vaginal microbiome may be also dominated by *Lactobacillus mulieris, Lactobacillus paragasseri* and *Limosilactobacillus urinaemulieris*. Comparison of vaginal microbiome at two distant time points unveiled that the most stable CST was dominated by *Bifidobacterium* sp..

We demonstrated that paired urinary and vaginal samples were substantially different, sharing an average of 38% of species, which also supports the usage of mid-stream voided urine as representative sample for lower urinary tract microbiome characterization.

These findings are essential to better understand the impact of vaginal and urinary microbiomes in the health of both niches, a critical knowledge that may guide in the future the development of microbiome-based modulation therapies.

INTRODUCTION

The existence of urinary microbiome among healthy persons was confirmed by Wolfe et al in 2012 (1), leading to an intensive research aiming to define healthy urinary tract (UT) microbiome composition and evaluate if dysbiosis may be involved in urinary disorders with unknown etiology.

To date, characterization of urinary tract microbiome composition pointed for *Lactobacillus*, *Gardnerella*, *Streptococcus* as the most prevalent and abundant microbiome members (2–6, Perovic SU in preparation), being highly variable among individuals and influenced by certain physiological or behavioral factors e.g., age, menstrual cycle, smoking status (3, 5, 7, 8). Primary longitudinal data for urinary tract microbiome is also available, which suggests that microbiome can undergo both daily as well as long-time compositional variations (3, 5). Available longitudinal studies (short- and long-term) point to certain community structures that seem to be more resilient in UT composed by e.g., *Lactobacillus crispatus, Lactobacillus paragasseri,* and/or *Streptococcus agalactiae*.

According to current knowledge, many species detected in urinary tract microbiome are also common members of vaginal microbiome. Although healthy vaginal microbiome was initially associated mostly with *Lactobacillus* species (e.g., *L. crispatus*, *L. iners*, *L. gasseri*), nowadays more community state types (CSTs) involving other genera, including many anaerobes (e.g., *Atopobium vaginae, Gardnerella* sp., *Prevotella bivia*) are detected among asymptomatic women (9–11). Noteworthy, the same bacterial species are reported from microbiome of UT, independently of urine sample collection method (voided urine or transurethral catheterization) (3–5, 12).

In fact, urinary and genital tract are anatomically linked, and their proximity likely promotes bacterial transition, as already demonstrated for *Escherichia coli* and urinary tract infections (13). Indeed, it was reported that the existing interlink between female urinary and vaginal microbiome refer not only to common uropathogens but also commensal species (12). Thomas-White et al. focused on genomic similarity of selected representatives of urinary tract microbiome members (e.g., *Escherichia coli*, *Streptococcus anginosus*, *Lactobacillus crispatus*) and demonstrated phylogenetic relationship for strains residing in both niches. However, detailed composition of urinary tract and vaginal microbiome have been scarcely characterized. A recent study by Komesu et al. investigated microbiome composition from both niches by sequencing of 1-3 variable regions of 16S rRNA gene (14). It was shown that many bacterial taxa (e.g., *Lactobacillus*, *Gardnerella*, *Prevotella*) have high correlation coefficients between urinary tract and vaginal microbiome and overall *Lactobacillus* are the most abundant bacteria in both niches (14). Due to the limitations of 16S rRNA sequencing, this study presented data mostly at genus level, and only *Lactobacillus* were preliminary identified to species level.

Nevertheless, it is anticipated that the presence of certain bacteria in vaginal environment can have an impact on urinary tract homeostasis (15), being important to comprehensively characterize the bacterial interlink between vaginal and urinary tract microbiome.

This study aimed to characterize at species level the compositional patterns, and evaluate the similarity, of microbiome of healthy vaginal and urinary tracts. A meticulous culturomic analysis of vaginal microbiome composition and over-time stability from reproductive-age European women was performed with further integration of the data from paired voided urine and vaginal samples (originating from the same individual) to understand species-level interlink between vaginal and urinary tract microbiome.

METHODS

Sample collection

Twenty women enrolled in the Female Urogenital Microbiome study (3, Perovic SU in preparation) provided a total of twenty-nine paired urine and vaginal samples. Data regarding urine samples has been previously published (3, Perovic SU in preparation). Also, in a previous study (3), ten women provided two set of samples within 2.5-year interval. Samples were collected in the 3rd week of the menstrual cycle, being collected in the same day the vaginal swabs followed immediately by the urine samples. All participants received Transystem™ collection swab (Copan, Italy) with amies agar gel, without charcoal, and an Eppendorf tube with 1ml of sterile tryptic soy broth (Liofilchem, Italy) supplemented with 20% (v/v) glycerol. Participants were instructed to immerse swab (shortly and gently) directly after sample collection into the Eppendorf tube with microbiological media mixed with glycerol, and just after insert the swab into Transystem™ collection tube. This step was performed to support long-term preservation of putatively sensitive species (internal validation). Collected swabs were stored at -80°C until further processing.

Sample analysis

Immediately before culturing, vaginal swabs were extracted into primary 1ml sterile solution of 0.9% NaCl that constituted the base for further dilutions and quantification (colony forming units/1ml of extracted swab; CFU/ml). The basic protocol included sampling of 0.1µl, 1µl and 100µl of primary solution, but when high bacterial load was suspected after microscopic examination, appropriate dilutions (up to 100 x) were additionally sampled. Prepared aliquots were subjected to an extended culturomic analysis, as previously published (3, Perovic SU in

preparation). In brief, each sample was plated on Columbia agar with 5% sheep blood (bioMérieux, France) and supplemented chromogenic agar (HiCrome UTI, HiMedia, India). Plates were incubated for 48h at 37°C under different atmospheric conditions (3, Perovic SU in preparation). After incubation, each colony morphotype was quantified and up to 3 colonies of the same morphotype were isolated, stored and subjected to identification.

Isolates identification

Isolates were identified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) VITEK MS system (bioMérieux, France), using in-vitro diagnostic database version 3.0. In case of no identification by MALDI-TOF MS, the isolates were further subjected to 16S rRNA gene sequencing and/or other suitable genotypic biomarkers (*pheS*, *cpn60*, *rpoB*, *recN*) (3, Perovic SU in preparation).

Data analysis

Compositional analyses were performed based on relative abundance (RA) data (raw CFU per species/ total CFU observed in the sample). Samples similarity was assessed using Bray-Curtis distances (values between 0 to 1, where the lowest value stands for the highest similarity) and hierarchical clustering using unweighted pair-group method (UPGMA) and applying a cutoff of 0.8 distance for clusters delineation. Diversity analysis included alpha (within samples) and beta diversity (between samples). Alpha diversity was represented by following measures: richness defined by species number, evenness defined by Pielou index (values between 0 to 1, where the lowest values stand for not even distribution) and by Shannon index which combines evenness and richness (value of Shannon index increases when evenness and richness increases). Beta diversity was assessed using Bray-Curtis distances. Samples ordination was performed by nonmetric multidimensional scaling (NMDS) algorithm using original Bray-Curtis and Jaccard distance data as inputs. Samples ordination was visualized in 2-dimensional space and stress function was outputted (values between 0 to 1, where the lowest stress value stands for confident fit). Comparison of differential abundances according to specific metadata variable and its statistical interpretation was performed using permutational multivariate analysis of variance (PERMANOVA; adonis function available in vegan package) and analysis of similarities (ANOSIM; vegan package). Statistical means comparisons were performed with nonparametric Kruskal-Wallis or Wilcoxon test, wherever appropriate.

All analyses and visualizations were performed in R version 4.0.3 (16) using following packages: microbiome v1.12.0, phyloseq v1.34.0, gplots v3.1.1, ggplot2 v3.3.5, vegan v2.5.7, readxl v1.3.1, dplyr v1.0.7, plyr v1.8.6, Heatplus v2.36.0, RColorBrewer v1.1.2, dendextend v1.15.1, tidyverse v1.3.1, patchwork v1.1.1, ggpubr v0.4.0, reshape2 v1.4.4, corrplot v0.9 (17–21).

RESULTS

Vaginal microbiome of asymptomatic women

A total of 1084 isolates from 29 vaginal samples were identified to 5 phyla, 45 genera and 104 species (**Supplementary Table S1**). Bacterial load varied from ~ 10^2-10^8 (mean 10^7, median 10^5) CFU/ml. Bacteria belonging to Firmicutes covered for most of bacterial load (RA = 69.4%), followed by Actinobacteria (28.7%) and in smaller extent by Proteobacteria (1.8%) and others. The highest species diversity was observed among Firmicutes (53 species) and Actinobacteria (38 species), followed by less diverse Proteobacteria (10 species), Bacteroidetes (2 species) and Fusobacteria (1 species).

Alpha diversity of vaginal microbiome was represented by richness (species number), evenness (Pielou index) and Shannon index. We detected from 5 to 22 species/sample (median 16 species/sample), that had a highly variable evenness (Pielou index range 0.03-0.9, median 0.46). Shannon index also appeared to be highly divergent (range 0.08-2.25), with most samples presenting relatively low diversity (median 1.22). We also investigated the trend of alpha diversity measures and associated metadata, as shown in **Figure 1**. Two samples from one donor using a contraceptive implant (without menstrual cycle) showed significantly lower species evenness based on Pielou index (Wilcoxon test, p=0.02) and significantly lower diversity based on Shannon index (Wilcoxon test, p=0.01) of vaginal microbiome, comparing to women with menstrual cycles (**Figure 1**). For the other variables (age, BMI, contraceptive usage, previous pregnancy, usage of anti-inflammatory drugs a week before sampling, smoking status or previous urinary tract infections) we did not observe statistically significant differences.

Overall, the most abundant species were *Lactobacillus crispatus* (11.9% of total RA), *Lactobacillus iners* (7.9%) and *Streptococcus agalactiae* (6.8%) (**Supplementary Table S1**). However, the most prevalent species (present in >50% of the samples) were *Staphylococcus epidermidis* (n=27 samples), *Staphylococcus haemolyticus* (n=21) and

Enterococcus faecalis (n=21), followed by *Corynebacterium tuberculostearicum* (n=18), *Corynebacterium amycolatum* (n=18) and *Actinomyces neuii* (n=16).

Comparison between samples (beta diversity) assessed by clustering of vaginal microbiome based on Bray-Curtis distance matrix and UPGMA algorithm revealed 13 community state types (CSTs) (**Figure 2**). Almost 50% of the samples were dominated by *Lactobacillus* species (6/13 CSTs), i.e., *L. crispatus* (VS24a, VS11a, VS4c, VS9a; mean RA 76.6%), *L. iners* (VS5a, VS10a, VS22a, VS26a; mean RA 42%), *L. mulieris* (VS8a, VS26b; mean RA 44.6%), *L. gasseri* (VS25a, VS29a; mean RA 29.9%), *L. paragasseri* (VS7a; RA 89.2%) and *L. jensenii* (VS12a; RA 90.7%). Additionally, one sample was dominated by related *Limosilactobacillus urinaemulieris* (VS9b; RA 99%).
We also observed dominance of *Streptococcus agalactiae* (VS3c, VS29b; mean RA 81.4%), CSTs with abundant anaerobic species i.e., *Anaerococcus tetradius* and *Finegoldia mangna* (VS23a, VS23b; mean RA 26.4% and 21.6%, respectively), CSTs with abundant *Staphylococcus epidermidis* (VS19a, VS12b, VS6a, VS3b, VS11b; mean RA 25.9%) in combination with *Actinomyces neuii* and often *Corynebacterium* sp..
As per other CSTs, 3 samples (VS15b, VS7b, VS15a) were characterized by highly abundant *Bifidobacterium* sp. (mean RA 63.1%) and/or *Gardnerella* sp. (mean RA 35.8%), specifically *G. swidsinskii* and *G. vaginalis* and low abundant *Staphylococcus epidermidis* and *Streptococcus anginosus* (mean RA <0.01%). Remaining 2 samples demonstrated unique microbiome composition, with one of them (VS1b) being dominated by *Actinomyces neuii* and the second one (VS4b) by *Cutibacterium acnes*.

Stability of vaginal microbiome of asymptomatic women

Since our 20 vaginal swabs originated from 10 women sampled at two different times (within 2,5-year interval), we assessed compositional similarity of paired swabs obtained from the same individuals. By the hierarchical clustering analysis, we observed that only 2 pairs of samples (VS15a/b and VS23a/b) clustered into the same CST while the remaining 8 did not (**Supplementary Figure S1 and S2**). However, some paired samples clustered relatively close, suggesting that their CST did not shift drastically. Two most stable pairs represent a community composed by i) highly abundant *Bifidobacterium* sp. and ii) abundant anaerobic species i.e., *Anaerococcus tetradius* and *Finegoldia magna* (CST defined also in previous paragraph).

Compositional (dis)similarities of paired urinary and vaginal microbiome

In order to better understand the link between urinary and vaginal microbiome, 29 urine samples processed and analyzed as previously reported (3, Perovic SU in preparation) were included for comparison.

The comparison of 29 paired urinary tract and vaginal microbiomes at phylum level demonstrated divergent phyla distribution for many sample pairs (**Supplementary Figure S3**). Overall, we observed less abundant Proteobacteria and Actinobacteria, and more abundant Firmicutes in vaginal microbiome, comparing to urinary tract microbiome (**Supplementary Figure S3**). Dissimilarity between sample pairs increased with more specific taxonomic ranks e.g., order and family level (**Supplementary Figure S4** and **S5**, respectively). In **Table 1** we present species level comparison of CSTs of paired vaginal and urinary tract microbiome obtained from the same women. CSTs are shortly defined as top3 species shared between all members of CST, ordered by decreasing abundance (complete list of species detected in each microbiome is presented in **Supplementary Table 1**). We assessed the number of species overlapping in both sample pairs and identified that 12 to 59% of species were detected in both paired samples (mean=39%; **Table 1**).

Additionally, in **Figure 3** relative and absolute abundance for paired samples of 10 most abundant genera are presented, which slightly differed for RA and CFU data. For instance, *Bifidobacterium*, *Gardnerella*, *Lactobacillus* and *Streptococcus* were consistently found among the most abundant genera in both sites (**Figure 3**). Noteworthy, absolute abundance data demonstrated also that most urinary and vaginal samples presented low bacterial load, with exception of samples characterized by *Bifidobacterium*, *Gardnerella*, and occasionally *Lactobacillus*. Moreover, *Citrobacter* and *Escherichia* were detected in high load in two urinary microbiomes of asymptomatic women (3).

We further performed hierarchical clustering for paired urinary and vaginal samples based on Bray-Curtis distance matrix (**Supplementary Figure S6).** Using UPGMA and cutoff of 0.8 we identified 18 clusters (**Supplementary Figure S7**) comprising the 29 paired urine and vaginal samples. While 62% of the sample pairs (n=18/29) were considered dissimilar, as vaginal and urine samples from the same women were found in different clusters, the remaining (n=11/29; sample pairs: 15a, 15b, 7a, 7b, 8a, 11a, 24a, 3c, 12a, 5a, 22a) presented similar community structure (**Supplementary Figures S6 and S7**).

Alpha diversity of urinary and vaginal microbiome

Alpha diversity indices i.e., richness defined by species number, evenness defined by Pielou index and combined abundance and evenness by Shannon index for urinary and vaginal samples are presented in **Figure 4**. The alpha diversity was only significantly different (p=0.002) concerning the species richness, with mean of 19.7 species/sample (median 19) in urinary microbiome and mean of 15.2 species/sample in vaginal microbiome. We did not observe significant differences for Pielou evenness and Shannon index.

As per overall species detection, from total of 166 species, 80 were identified in both sample types, 62 species were detected only from urine samples (e.g., *Acinetobacter pittii*, *Klebsiella oxytoca*, *Oligella urethralis*) and 24 species were identified only from vaginal samples (e.g., *Anaerococcus vaginalis*, *Atopobium rimae*, *Enterococcus faecium*, *Limosilactobacillus fermentum*).

Beta diversity of urinary and vaginal microbiome

We assessed samples' beta diversity by using two different indices, Jaccard distance was used to visualize community composition (including only data on species presence/absence) and Bray-Curtis distance to visualize community structure (including species abundance). We performed samples ordination using non-metric multi-dimensional scaling (NMDS) with 1000 iterations, using both distance matrices. NMDS based on Jaccard and Bray-Curtis distances is presented in **Figure 5**. The stress value for both 2-dimensional representations was ~ 0.26 which correspond to weak ties. Differences between vaginal and urinary tract microbiome communities were not significant statistically ($p<0.05$) for both Jaccard (ANOSIM p=0.06, adonis p=0.17) and Bray-Curtis distances (ANOSIM p=0.39, adonis p=0.55).

Differential abundance in urinary and vaginal microbiome

Although overall diversity between urinary and vaginal microbiome was not statistically significant, we explored further most abundant taxa, particularly their distribution in both sample types. We compared total RA of 5 most abundant genera identified from urinary and vaginal samples (**Figure 6**). Although visually we could assume higher abundance of *Lactobacillus* and lower abundance of *Gardnerella* in vaginal microbiome comparing to urinary, the Wilcoxon test showed that differences observed are not statistically significant (p=0.93 and p=0.067). There were also no differences for remaining top 3 genera i.e., *Streptococcus*, *Corynebacterium* and *Staphylococcus*. We also observed cumulative relative

and absolute abundance species level trend among most variable genera i.e., *Lactobacillus* and *Gardnerella* (**Figure 7**).

While RA of various *Lactobacillus* species were on similar levels in vaginal and urinary tract microbiome, higher CFU of *Lactobacillus* were observed in vaginal microbiome. *Gardnerella* genomospecies 3 was detected in high RA and only in urinary tract microbiome. *Gardnerella swidsinskii* and *Gardnerella vaginalis* presented higher load (CFU) in vaginal microbiome (**Figure 7**).

Co-occurence patterns in urinary and vaginal microbiome

We further extracted model coefficients from PERMANOVA analysis and identified species that contributed the most to compositional differences between urinary and vaginal microbiome (**Figure 8**). Besides previously mentioned urine specific *Gardnerella* genomospecies 3, based on species abundance, urinary microbiome was likely enriched in e.g., *Citrobacter koseri*, *Escherichia coli* and *Gardnerella vaginalis*, while vaginal microbiome had more *Staphylococcus epidermidis*, *Limosilactobacillus urinaemulieris* and *Actinomyces neuii*, among the others (**Figure 8**).

We also performed Spearman correlation to assess taxonomic co-occurrence in vaginal and urinary tract microbiome (**Supplementary Figure S8**). Looking only at 5 top abundant genera, *Lactobacillus* and *Gardnerella* showed negative correlation (r_s=-0.25) in both urinary and vaginal microbiome, similarly to *Lactobacillus* and *Streptococcus* (r_s=-0.10 urinary, r_s=-0.23 vaginal), *Gardnerella* and *Corynebacterium* (r_s=-0.31), and *Gardnerella* and *Staphylococcus* (r_s=-0.30 urinary, r_s=-0.37 vaginal). Strong positive correlation was observed for *Corynebacterium* and *Staphylococcus* in both niches (r_s=0.68 urinary, r_s=0.62 vaginal) and weaker for *Corynebacterium* and *Streptococcus* (r_s=0.2 urinary, r_s=0.35 vaginal).

Species level co-occurrence analysis revealed additional within-genus patterns such as positive correlation of *Gardnerella vaginalis* and *Gardnerella swidsinskii* or relatively strong positive correlation of *Lactobacillus delbruecki* and *Lactobacillus paragasseri*, in both niches. We also observed differences between niches at species level, e.g., positive correlation for *Lactobacillus jensenii* and *Lactobacillus iners* in vaginal (r_s=0.43) while negative (r_s=-0.20) in urinary microbiome or stronger positive correlation for *Lactobacillus mulieris* and *Lactobacillus iners* in vaginal (r_s=0.47) than urinary microbiome (r_s=0.05). Additionally, in urinary microbiome *Gardnerella leopoldii* and other *Gardnerella* sp. were negatively correlated. Species level Spearman correlation plots for vaginal and urinary tract microbiome are available in **Supplementary Figures S9 and S10**, respectively.

DISCUSSION

In this study we analyzed microbiome composition from 29 vaginal samples and compare it with their paired urinary tract microbiome data. We compared those niches including their alpha and beta diversity, differential abundance of top taxa, and species co-occurrence, evaluating interconnection patterns.

Our data on vaginal microbiome highlight diversity of CST that can be observed among asymptomatic women. Similarly to what is currently known on vaginal microbiome (9, 22, 23), a *Lactobacillus* dominating community was the most commonly observed. However, we identified more CST comprised by *Lactobacillus* species than previously reported i.e., *L. mulieris* and *L. paragasseri*. This difference may be easily explained as those two species are relatively novel (24, 25) and they are closest relative of *L. jensenii* and *L. gasseri*, respectively. However, usually reported lower diversity of *Lactobacillus* in vaginal microbiome may be a consequence of widely applied 16S rRNA community sequencing for microbiome characterization (22, 23, 26, 27) since this methodology does not allow to reliably distinguish those closely related species due to high homology of their 16S rRNA gene sequences (24, 25). Besides *Lactobacillus* dominated CST, we identified CST with highly dominant *Limosilactobacillus urinaemulieris* - novel species of *Limosilactobacillus* sp. (former *Lactobacillus* sp.) (28) and several CST composed of various facultative or obligate anaerobic bacteria. Some of them e.g., *Anaerococcus tetradius*, *Finegoldia magna* or *Streptococcus anginosus*, may be involved in vaginal infections (bacterial vaginosis or aerobic vaginitis) (29, 30). Indeed, communities composed of non-*Lactobacillus* bacteria were previously associated with higher susceptibility to bacterial vaginosis and sexually transmitted infections (31, 32). Interestingly, some of these species are also currently investigated to understand their contribution to urinary tract health e.g., *Finegoldia magna* and *Streptococcus anginosus* putative involvement in urinary incontinence (33, 34). We also observed *Gardnerella vaginalis* and *Gardnerella swidsinskii* in vaginal microbiome of asymptomatic women, however their contribution to vaginal and urinary tract health needs to be evaluated and is currently the subject of high interest.

We also observed higher alpha diversity (Pielou evenness and Shannon index) in women with menstrual cycle, comparing to the one without. This data can be linked to previous reports showing higher Shannon index during menses (26), however due to our small sample size, it needs to be further evaluated.

We further evaluated the long-term stability of vaginal microbiome at two time points (within 2,5-year interval). The temporal dynamics of vaginal microbiome has been studied previously by several groups (22, 23, 26, 35–38) but they assessed multiple samples daily over maximum period of 4 months (22). These previous reports suggested that vaginal communities dominated by most of *Lactobacillus* species tend to be more stable (specifically *L. crispatus*) (39) and shifts were observed more often in microbiomes dominated by *L. iners* or non-*Lactobacillus* genera (22). Curiously, in our cohort the most stable vaginal communities were composed by abundant *Bifidobacterium*, but also anaerobic bacteria. These discrepancies may be the result of methodologies applied for CST delineation. Here, we subjected complete microbiome community for hierarchical clustering, which may result in CST characterized by species combinations instead of most abundant taxa. Moreover, with our study design (sampling specifically at 3rd week of menstrual cycle) we eliminated possible variations according to menstrual cycle (22, 26). In fact, during menstrual cycle, vaginal microbiome diversity increases while abundance of *Lactobacillus* may decrease and *Gardnerella* increase (26, 35). Song et al also demonstrated that vaginal microbiome variations might be associated with usage of hormonal contraceptives and menstrual cycle, but also diet and exercise (26). Interestingly, Gajer et al demonstrated that metabolome, especially lactate level, did not change in woman undergoing CST shift from *Lactobacillus* to *Streptococcus*, demonstrating that although taxonomic variation was observed, in some individuals the microbiome functionality did not change (22).

Here we reported that more than half paired vaginal and urinary tract microbiome were classified in different CST (3, this study), which demonstrated that microbiome composition identified from voided urine and vaginal swabs from many individuals is substantially different. Moreover, approximately one third of species overlapped between sample pairs. These data suggest that contrary to what have been proposed by Brown et al (40), microbiome characterized from voided urine should not be used as a reliable proxy for estimation of vaginal microbiome composition. Notably, Brown et al followed CST classification, which determines low-*Lactobacillus* communities as one group. Considering high diversity of communities non-dominated by *Lactobacillus* sp. presented here, we believe they should be reported separately. Moreover, Brown et al and the other studies (26, 40) analyzed the data by 16S rRNA sequencing, which is prone to misidentified *Lactobacillus* species closely related to those commonly reported from vaginal environment (see above).

We report here that overall, both microbiomes show high similarity regarding most dominant taxa, similarly to findings by Komesu et al (14). Moreover, our alpha and beta diversity analysis shows that overall vaginal and urinary tract microbiome are highly similar, with only lower species richness detected for vaginal microbiome. Ranges of Pielou's evenness suggested more uneven species distribution in microbiome of both niches and Shannon index was slightly higher for urinary tract microbiome, however in our cohort was not found to be statistically significant. Lack of significance may be here explained by small sample size, since higher mean Shannon index for urinary tract microbiome than for vaginal microbiome was observed in other cohorts (14).

Despite high overall similarity (e.g., species often detected in both niches such as *Lactobacillus crispatus*, *Streptococcus anginosus*, *Escherichia coli* (12, this study), microbiome of vagina and urinary tract seem to have certain specific particularities. We observed that our vaginal samples have *Lactobacillus* species in higher RA, comparing to urinary tract microbiome, which is in agreement with previous findings (14). We also observed that *Gardnerella* sp. could be in higher RA in urinary tract microbiome than in vaginal microbiome. Although in our cohort this observation was not significant statistically, it was previously reported that *Lactobacillus* and *Gardnerella* are inversely related in microbiome (35). Nevertheless, this correlation should be further investigated. Of note, most species from both genera (*Lactobacillus* and *Gardnerella*) were in much higher bacterial load (CFU) in vaginal microbiome, confirming that urinary samples have low bacterial biomass.

Importantly, the similarities and differences between these niches are more clear at species level. For instance, *Gardnerella* genomospecies 3 and *Gardnerella leopoldii* were identified only in urinary tract microbiome. These findings require validation on larger cohort, since it seems that these *Gardnerella* species were previously isolated from vagina in other study (41). Further identification of species contributing to compositional differences between both microbiomes highlighted species specificity and abundance according to niche e.g., highly abundant *Gardnerella* genomospecies 3, *Citrobacter koseri* or *Escherichia coli* being more specific for urinary tract microbiome, and e.g., *Limosilactobacillus urinaemulieris* or *Corynebacterium* putative novel species 5 (Perovic SU in preparation) being more specific to vaginal microbiome.
Furthermore, with co-occurrence analysis and Pearson correlation we shown positive correlation in both niches for *Corynebacterium* with *Staphylococcus* and *Streptococcus*,

while negative for the other most abundant genera. This was expected since non-*Lactobacillus* dominated communities are often composed of species belonging to these Gram-positive genera (3). We found that *Lactobacillus* and *Gardnerella* are negatively correlated in both niches, which support their previously mentioned inverse relation (35) and dominance of one or another in vaginal and urinary tract microbiome. Moreover, we reported that certain species are likely to co-occur in both niches e.g., *Gardnerella vaginalis* and *Gardnerella swidsinskii*, while other species combinations demonstrated opposite pattern e.g., positive correlation for *Lactobacillus jensenii* and *Lactobacillus iners* in vaginal microbiome and negative in urinary tract microbiome. These might be particularly important since *L. iners* role in urogenital health is controversial (42). Unveiling these bacterial interactions and co-occurrence patterns in vaginal and urinary tract microbiome will likely contribute to further understanding of their function.

Noteworthy, although the use of voided urine as a representative of lower urinary tract microbiome is still controversial, here we demonstrated that the microbiome composition identified from voided urine and vaginal swab differs. Moreover, Komesu et al analyzed microbiome of urine obtained by transurethral catheterization which resulted in highly similar results to ours (genus level) (14).

CONCLUSIONS

Here we present data regarding interlink of vaginal and urinary tract microbiome obtained from well-characterized cohort of reproductive-age asymptomatic women. We demonstrate specific patterns of bacterial abundance and species co-occurrence that might be characteristic to urinary tract or vaginal microbiome. Our findings also suggest that overall structure of both bacterial communities is similar, thus urogenital microbiome may function as a whole, and specific microbes likely may interchange their exact location and act complementary in both niches. Moreover, our data demonstrate that usage of midstream voided urine is suitable for lower urinary tract microbiome investigation.

This data supports existence of interconnection between vaginal and urinary tract microbiome and will contribute to further development of therapeutic approaches involving microbiome interventions

List of abbreviations

ANOSIM, analysis of similarities; CFU, colony forming unit; CST, community state types; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; NMDS, nonmetric multidimensional scaling; PERMANOVA, permutational multivariate analysis of variance; RA, relative abundance; UPGMA, unweighted pair-group method algorithm; UT, urinary tract; UTI, urinary tract infection; VS, vaginal swab

Declarations

Ethics approval and consent to participate

Approval of the study was obtained from the Faculty of Pharmacy (University of Porto, Porto, Portugal) Ethics Committee. Procedures performed in the study were all in accordance with the ethical standards of the institutional and national research committee, with the 1964 Helsinki Declaration, and its later amendments. All individual participants included in the study had given written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal, and by the UCIBIO – Applied Molecular Biosciences Unit, Laboratory of Microbiology, Department of Biological Sciences, REQUIMTE, Faculty of Pharmacy, University of Porto, 4050-313 Porto, which are financed by national funds from FCT (Fundação para a Ciência e a Tecnologia, I.P.) (UIDB/MULTI/04378/2020 and UIDP/04378/2020). SUP was supported as researcher from the project NORTH-01-0145-FEDER-000024, MK and MS by an FCT PhD grant (SFRH/BD/132497/2017 and SFRH/BD/05038/2020, respectively), and JR by a postdoctoral fellowship from ICETA (UID/MULTI/04378/2013). FG is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0034].

Authors' contributions

SUP, MK, FG and LP designed the study and supervised participant recruitment. SUP, MK and JR processed the samples and collected the data. SUP, MK, JR, MS, EC and TGR performed the isolates' identification. TGR supervised MS and EC. MK conducted the data analysis, visualization, interpreted the data and wrote the manuscript. TGR, FG and LP revised the article. All authors read and approved the manuscript.

Acknowledgements

We thank the women who participated and contributed time and effort to make this study a success. We thank Helena Ramos and Paulo Pinto (Hospital Geral de Santo António, Porto, Portugal) for their technical support with MALDI-TOF MS-based microbial identification. The bioMérieux (Portugal, Lda.) provided equipment and material for MALDI-TOF MS analysis, and had no role in the study design, data collection and analysis, decision to publish, or writing of the manuscript.

References

1. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, Fitzgerald M, Mueller ER, Schreckenberger P, Dong Q, Nelson DE, Brubaker L. 2012. Evidence of uncultivated bacteria in the adult female bladder. J Clin Microbiol 50:1376–1383.
2. Price TK, Hilt EE, Thomas-White K, Mueller ER, Wolfe AJ, Brubaker L. 2020. The urobiome of continent adult women: a cross-sectional study. BJOG 127:193–201.
3. Ksiezarek M, Ugarcina-Perovic S, Rocha J, Grosso F, Peixe L. 2021. Long-term stability of the urogenital microbiota of asymptomatic European women. BMC Microbiol 21:64.
4. Coorevits L, Heytens S, Boelens J, Claeys G. 2017. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. Eur J Clin Microbiol Infect Dis 36:635–639.
5. Price TK, Wolff B, Halverson T, Limeira R, Brubaker L, Dong Q, Mueller ER, Wolfe AJ. 2020. Temporal Dynamics of the Adult Female Lower Urinary Tract Microbiota. mBio 11.
6. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. J Clin Microbiol 52:871–876.

7. Ammitzbøll N, Bau BPJ, Bundgaard-Nielsen C, Villadsen AB, Jensen A-M, Leutscher PDC, Glavind K, Hagstrøm S, Arenholt LTS, Sørensen S. 2021. Pre- and postmenopausal women have different core urinary microbiota. 1. Sci Rep 11:2212.
8. Adebayo AS, Ackermann G, Bowyer RCE, Wells PM, Humphreys G, Knight R, Spector TD, Steves CJ. 2020. The Urinary Tract Microbiome in Older Women Exhibits Host Genetic and Environmental Influences. Cell Host Microbe 28:298-305.e3.
9. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 108 Suppl 1:4680–4687.
10. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Nikita L, Galuppi M, Lamont RF, Chaemsaithong P, Miranda J, Chaiworapongsa T, Ravel J. 2014. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. Microbiome 2:4.
11. Moosa Y, Kwon D, de Oliveira T, Wong EB. 2020. Determinants of Vaginal Microbiota Composition. Front Cell Infect Microbiol 10:467.
12. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, Hilt EE, Price TK, Wolfe AJ, Lawley TD. 2018. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. 1. Nat Commun 9:1557.
13. Brannon JR, Dunigan TL, Beebout CJ, Ross T, Wiebe MA, Reynolds WS, Hadjifrangiskou M. 2020. Invasion of vaginal epithelial cells by uropathogenic *Escherichia coli*. Nat Commun 11:2803.
14. Komesu YM, Dinwiddie DL, Richter HE, Lukacz ES, Sung VW, Siddiqui NY, Zyczynski HM, Ridgeway B, Rogers RG, Arya LA, Mazloomdoost D, Levy J, Carper B, Gantz MG. 2020. Defining the relationship between vaginal and urinary microbiomes. Am J Obstet Gynecol 222:154.e1-154.e10.
15. Stapleton AE. 2016. The Vaginal Microbiota and Urinary Tract Infection. Microbiol Spectr 4.
16. RCoreTeam. 2019. R: a Language and Environment for Statistical Computing.
17. McMurdie PJ, Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8:e61217.
18. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B, Galili T. 2020. gplots: Various R Programming Tools for Plotting Data.

19. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
20. Oksanen J, Guillaume B, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevens H, Szoecs E, Wagner H. 2019. Vegan: Community Ecology Package.
21. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. Journal of Open Source Software 4:1686.
22. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, Koenig SSK, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal dynamics of the human vaginal microbiota. Sci Transl Med 4:132ra52.
23. Ravel J, Brotman RM, Gajer P, Ma B, Nandy M, Fadrosh DW, Sakamoto J, Koenig SS, Fu L, Zhou X, Hickey RJ, Schwebke JR, Forney LJ. 2013. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. Microbiome 1:29.
24. Rocha J, Botelho J, Ksiezarek M, Perovic SU, Machado M, Carriço JA, Pimentel LL, Salsinha S, Rodríguez-Alcalá LM, Pintado M, Ribeiro TG, Peixe L. 2020. *Lactobacillus mulieris* sp. nov., a new species of *Lactobacillus delbrueckii* group. Int J Syst Evol Microbiol 70(3):1522-1527.
25. Tanizawa Y, Tada I, Kobayashi H, Endo A, Maeno S, Toyoda A, Arita M, Nakamura Y, Sakamoto M, Ohkuma M, Tohno M. 2018. *Lactobacillus paragasseri* sp. nov., a sister taxon of *Lactobacillus gasseri*, based on whole-genome sequence analyses. Int J Syst Evol Microbiol 68:3512–3517.
26. Song SD, Acharya KD, Zhu JE, Deveney CM, Walther-Antonio MRS, Tetel MJ, Chia N. 2020. Daily Vaginal Microbiota Fluctuations Associated with Natural Hormonal Cycle, Contraceptives, Diet, and Exercise. mSphere 5.
27. Smith BC, McAndrew T, Chen Z, Harari A, Barris DM, Viswanathan S, Rodriguez AC, Castle P, Herrero R, Schiffman M, Burk RD. 2012. The cervical microbiome over 7 years and a comparison of methodologies for its characterization. PLoS One 7:e40425.
28. Ksiezarek M, Ribeiro TG, Rocha J, Grosso F, Perovic SU, Peixe L. 2021. *Limosilactobacillus urinaemulieris* sp. nov. and *Limosilactobacillus portuensis* sp. nov. isolated from urine of healthy women. Int J Syst Evol Microbiol 71:004726.

29. Neumann A, Björck L, Frick I-M. 2020. *Finegoldia magna*, an Anaerobic Gram-Positive Bacterium of the Normal Human Microbiota, Induces Inflammation by Activating Neutrophils. Front Microbiol 11:65.
30. Tao Z, Zhang L, Zhang Q, Lv T, Chen R, Wang L, Huang Z, Hu L, Liao Q. 2019. The Pathogenesis Of *Streptococcus anginosus* In Aerobic Vaginitis. Infect Drug Resist 12:3745–3754.
31. McKinnon LR, Achilles SL, Bradshaw CS, Burgener A, Crucitti T, Fredricks DN, Jaspan HB, Kaul R, Kaushic C, Klatt N, Kwon DS, Marrazzo JM, Masson L, McClelland RS, Ravel J, van de Wijgert JHHM, Vodstrcil LA, Tachedjian G. 2019. The Evolving Facets of Bacterial Vaginosis: Implications for HIV Transmission. AIDS Res Hum Retroviruses 35:219–228.
32. Gosmann C, Anahtar MN, Handley SA, Farcasanu M, Abu-Ali G, Bowman BA, Padavattan N, Desai C, Droit L, Moodley A, Dong M, Chen Y, Ismail N, Ndung'u T, Ghebremichael MS, Wesemann DR, Mitchell C, Dong KL, Huttenhower C, Walker BD, Virgin HW, Kwon DS. 2017. *Lactobacillus*-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women. Immunity 46:29–37.
33. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. mBio 5:e01283-01214.
34. Fok CS, Gao X, Lin H, Thomas-White KJ, Mueller ER, Wolfe AJ, Dong Q, Brubaker L. 2018. Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. Int Urogynecol J 29:1765–1771.
35. Srinivasan S, Liu C, Mitchell CM, Fiedler TL, Thomas KK, Agnew KJ, Marrazzo JM, Fredricks DN. 2010. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. PLoS One 5:e10197.
36. Mirmonsef P, Modur S, Burgad D, Gilbert D, Golub ET, French AL, McCotter K, Landay AL, Spear GT. 2015. Exploratory comparison of vaginal glycogen and *Lactobacillus* levels in premenopausal and postmenopausal women. Menopause 22:702–709.
37. Muzny CA, Blanchard E, Taylor CM, Aaron KJ, Talluri R, Griswold ME, Redden DT, Luo M, Welsh DA, Van Der Pol WJ, Lefkowitz EJ, Martin DH, Schwebke JR. 2018. Identification of Key Bacteria Involved in the Induction of Incident Bacterial Vaginosis: A Prospective Study. J Infect Dis 218:966–978.

38. Kaur H, Merchant M, Haque MM, Mande SS. 2020. Crosstalk Between Female Gonadal Hormones and Vaginal Microbiota Across Various Phases of Women's Gynecological Lifecycle. Front Microbiol 11:551.
39. Verstraelen H, Verhelst R, Claeys G, De Backer E, Temmerman M, Vaneechoutte M. 2009. Longitudinal analysis of the vaginal microflora in pregnancy suggests that *L. crispatus* promotes the stability of the normal vaginal microflora and that *L. gasseri* and/or *L. iners* are more conducive to the occurrence of abnormal vaginal microflora. BMC Microbiol 9:116.
40. Brown SE, Robinson CK, Shardell MD, Holm JB, Ravel J, Ghanem KG, Brotman RM. 2021. Assessing the Concordance Between Urogenital and Vaginal Microbiota: Can Urine Specimens Be Used as a Proxy for Vaginal Samples? Front Cell Infect Microbiol 11:671413.
41. Putonti C, Thomas-White K, Crum E, Hilt EE, Price TK, Wolfe AJ. 2021. Genome Investigation of Urinary *Gardnerella* Strains and Their Relationship to Isolates of the Vaginal Microbiota. mSphere 6.
42. Petrova MI, Reid G, Vaneechoutte M, Lebeer S. 2017. *Lactobacillus iners*: Friend or Foe? Trends in Microbiology 25:182–191.

Table 1. Species level comparison of community structure types (CSTs) of paired vaginal and urinary tract microbiome obtained from the same women. CSTs are shortly defined as top 3 species shared between all members of CST, ordered by decreasing abundance.

Donor/ sampling	Vaginal CSTs	Urinary CSTs	Nr of overlapping species (%)
1b	*Actinomyces neuii* *Staphylococcus haemolyticus* *Escherichia coli*	*Gardnerella leopoldii* *Alloscardovia omnicolens* *Bifidobacterium* spp.	3 (12)
3b	*Staphylococcus epidermidis* *Actinomyces neuii*	*Streptococcus agalactiae* *Staphylococcus epidermidis* *Corynebacterium tuberculostearicum*	12 (37.5)
3c	*Streptococcus agalactiae* *Finegoldia magna* *Streptococcus anginosus*	*Streptococcus agalactiae* *Staphylococcus epidermidis* *Corynebacterium tuberculostearicum*	13 (57)
4b	*Cutibacterium acnes* *Corynebacterium* sp. nov. 5 *Staphylococcus epidermidis*	*Streptococcus mitis/oralis* *Staphylococcus haemolyticus* *Micrococcus luteus*	5 (17)
4c	*Lactobacillus crispatus* *Staphylococcus epidermidis* *Corynebacterium tuberculostearicum*	*Bifidobacterium* spp. *Lactobacillus paragasseri* *Enterococcus faecalis*	9 (43)
5a	*Lactobacillus iners* *Staphylococcus epidermidis*	*Lactobacillus iners* *Corynebacterium tuberculostearicum* *Staphylococcus epidermidis*	8 (38)
6a	*Staphylococcus epidermidis* *Actinomyces neuii*	*Streptococcus agalactiae* *Streptococcus salivarius/vestibularis* *Micrococcus luteus*	5 (28)
7a	*Lactobacillus paragasseri* *Lactobacillus delbruecki*	*Bifidobacterium* spp. *Lactobacillus paragasseri* *Enterococcus faecalis*	13 (54)
7b	**Bifidobacterium* spp. *Streptococcus anginosus* *Staphylococcus epidermidis*	*Bifidobacterium* spp. *Lactobacillus paragasseri* *Enterococcus faecalis*	12 (48)
8a	*Lactobacillus mulieris* *Lactobacillus iners* *Staphylococcus epidermidis*	*Lactobacillus crispatus* *Lactobacillus mulieris* *Staphylococcus epidermidis*	10 (27)
9a	*Lactobacillus crispatus* *Staphylococcus epidermidis* *Corynebacterium tuberculostearicum*	*Corynebacterium striatum* *Dermabacter hominis* *Staphylococcus aureus*	6 (25)
9b	*Limosilactobacillus urinaemulieris* *Staphylococcus warneri* *Staphylococcus epidermidis*	*Lactobacillus crispatus*	5 (29)
10a	*Lactobacillus iners* *Staphylococcus epidermidis*	*Lactobacillus crispatus* *Lactobacillus mulieris* *Staphylococcus epidermidis*	13 (48)
11a	*Lactobacillus crispatus* *Staphylococcus epidermidis*	*Lactobacillus crispatus*	11 (44)

	Corynebacterium tuberculostearicum		
11b	*Staphylococcus epidermidis* *Actinomyces neuii*	*Lactobacillus crispatus*	9 (53)
12a	*Lactobacillus jensenii* *Corynebacterium amycolatum* *Staphylococcus haemolyticus*	*Lactobacillus jensenii* *Staphylococcus haemolyticus* *Staphylococcus epidermidis*	10 (24)
12b	*Staphylococcus epidermidis* *Actinomyces neuii*	*Lactobacillus crispatus*	9 (36)
15a	**Bifidobacterium* spp. **Gardnerella swidsinskii* *Streptococcus anginosus*	*Gardnerella swidsinskii* *Atopobium vaginae* *Dialister microaerophilus*	13 (46)
15b	**Gardnerella vaginalis* **Gardnerella swidsinskii* *Streptococcus anginosus*	*Gardnerella vaginalis* *Bifidobacterium* spp. *Cutibacterium avidum*	10 (34)
19a	*Staphylococcus epidermidis* *Actinomyces neuii*	*Lactobacillus crispatus* *Lactobacillus mulieris* *Staphylococcus epidermidis*	14 (58)
22a	*Lactobacillus iners* *Staphylococcus epidermidis*	*Lactobacillus iners* *Corynebacterium tuberculostearicum* *Staphylococcus epidermidis*	10 (37)
23a	*Finegoldia magna* *Corynebacterium* sp. nov. 5 *Anaerococcus tetradius*	*Lactobacillus crispatus*	6 (21)
23b	*Finegoldia magna* *Corynebacterium* sp. nov. 5 *Anaerococcus tetradius*	*Lactobacillus crispatus*	14 (52)
24a	*Lactobacillus crispatus* *Staphylococcus epidermidis* *Corynebacterium tuberculostearicum*	*Lactobacillus crispatus* *Corynebacterium tuberculostearicum* *Finegoldia magna*	15 (50)
25a	*Lactobacillus gasseri* *Corynebacterium tuberculostearicum* *Actinomyces neuii*	*Lactobacillus iners* *Corynebacterium tuberculostearicum* *Staphylococcus epidermidis*	13 (59)
26a	*Lactobacillus iners* *Staphylococcus epidermidis*	*Citrobacter koseri* *Enterococcus faecalis* *Lactobacillus jensenii*	4 (20)
26b	*Lactobacillus mulieris* *Lactobacillus iners* *Staphylococcus epidermidis*	*Escherichia coli* *Micrococcus luteus* *Lactobacillus jensenii*	4 (29)
29a	*Lactobacillus gasseri* *Corynebacterium tuberculostearicum* *Actinomyces neuii*	*Gardnerella vaginalis* *Bifidobacterium* spp. *Cutibacterium avidum*	10 (42)
29b	*Streptococcus agalactiae* *Finegoldia magna* *Streptococcus anginosus*	*Streptococcus agalactiae* *Staphylococcus epidermidis* *Corynebacterium tuberculostearicum*	12 (50)

*dominant in the sample, but not common to all samples in CST

Figure 1. Alpha diversity measures and associated metadata.

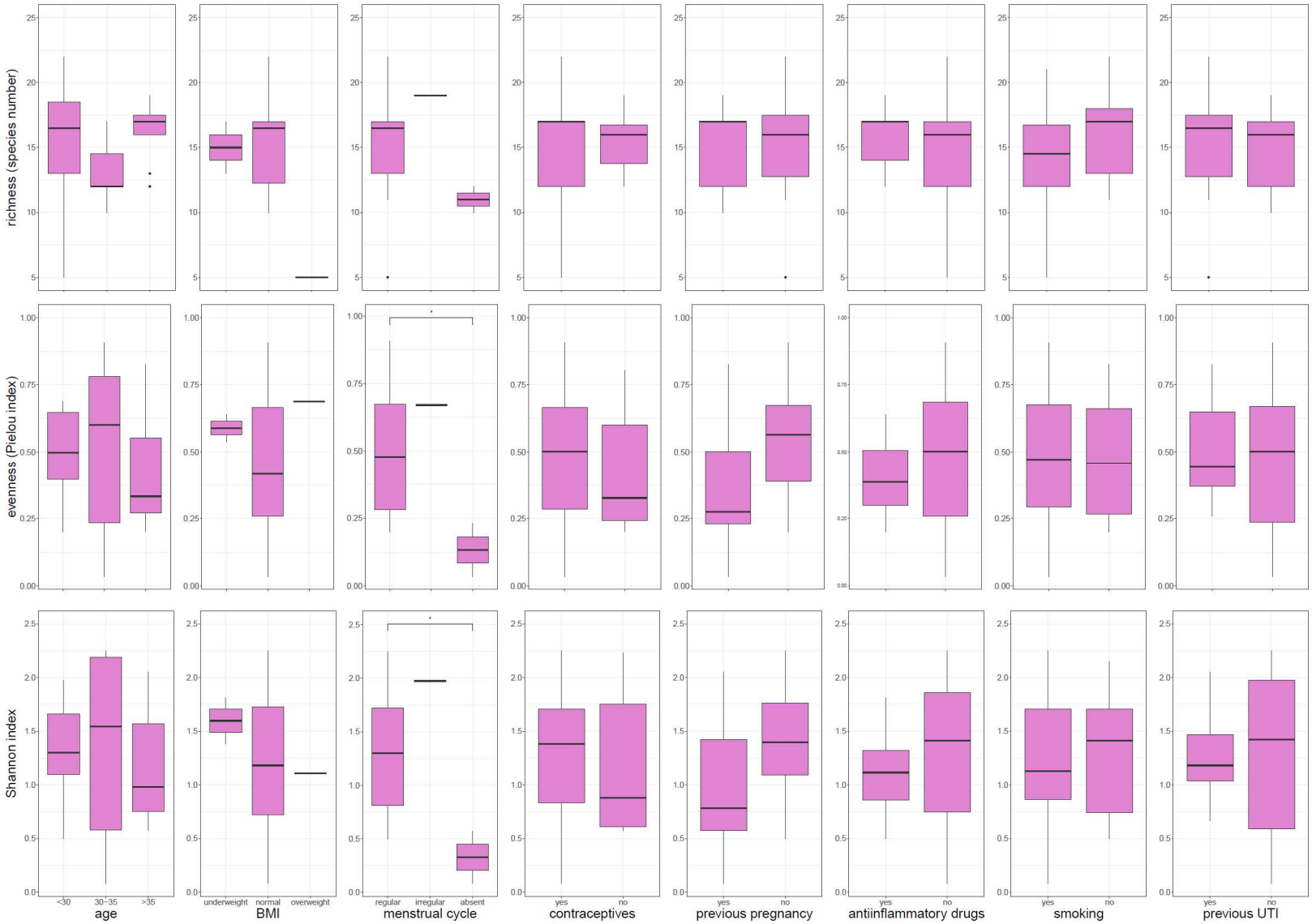


Figure 2. Heatmap representing vaginal community structure types (CSTs) assessed by UPGMA hierarchical clustering based on Bray-Curtis distance matrix. Colorful bar below the dendrogram define CSTs.

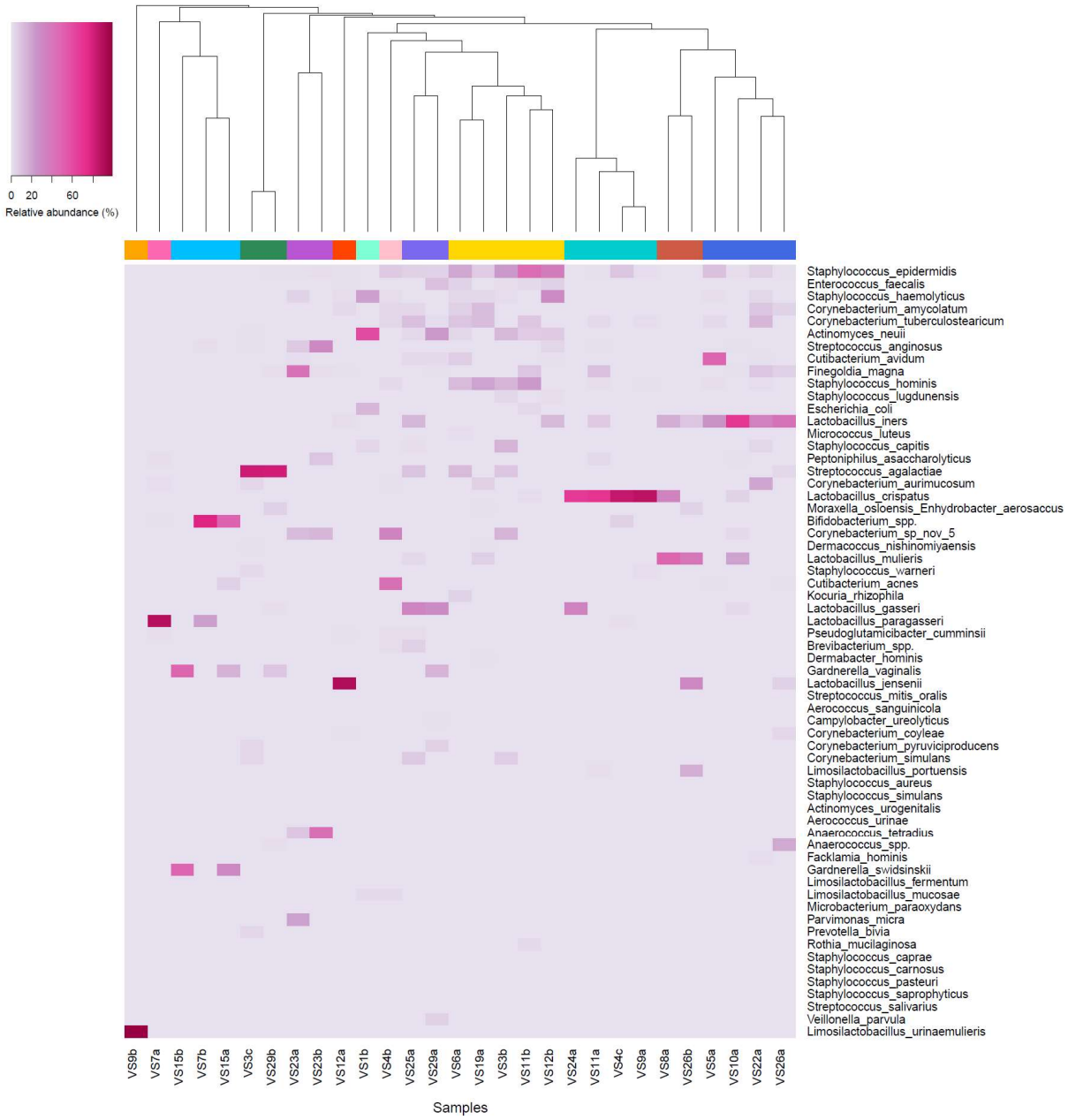


Figure 3. Relative and absolute abundance of 10 most abundant genera detected in paired samples.

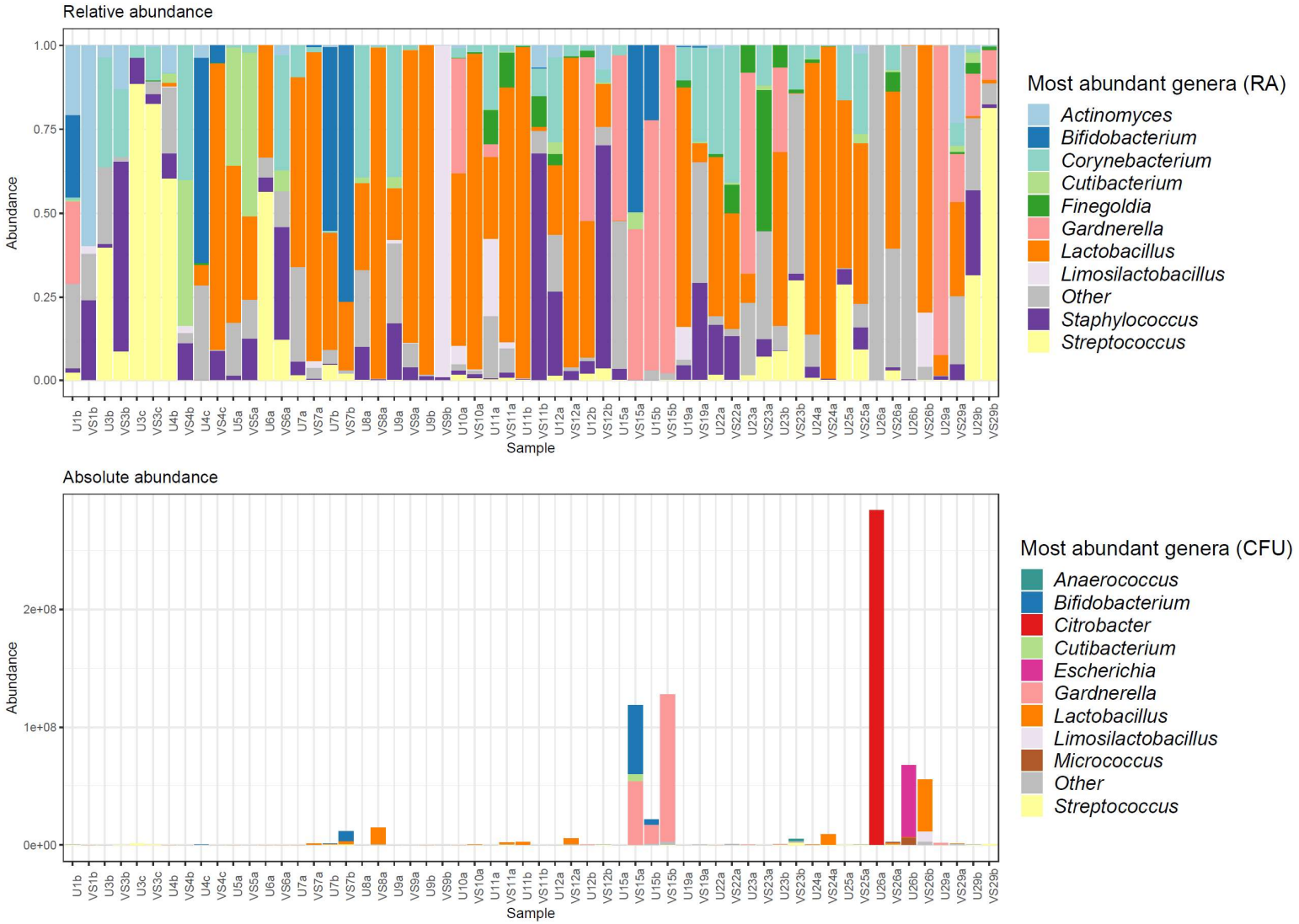


Figure 4. Distribution of alpha diversity indices for urinary tract and vaginal microbiome.

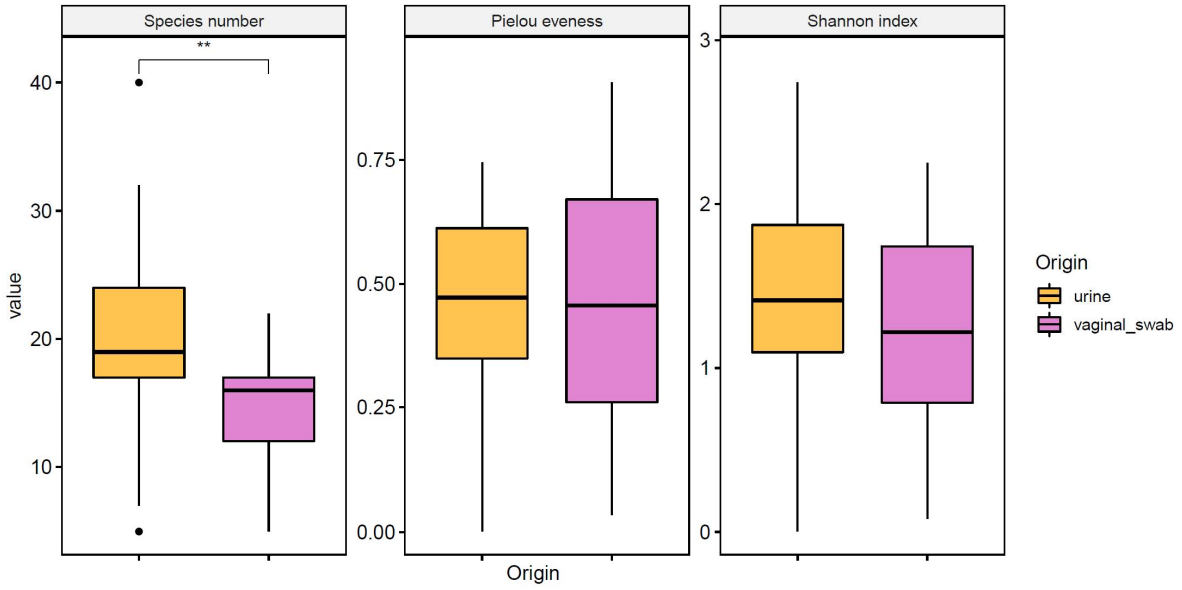


Figure 5. NMDS urinary and vaginal samples ordination based on Jaccard and Bray-Curtis distances. The stress value for both 2-dimensional representations was ~0.26.

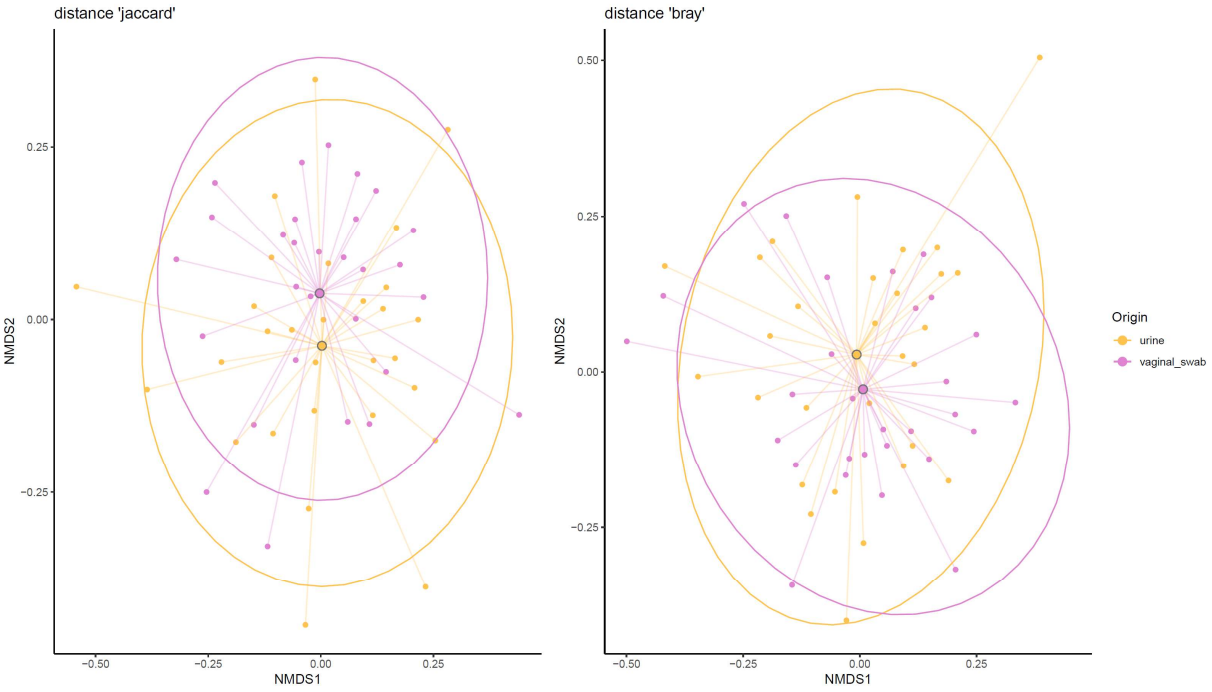


Figure 6. Distribution of relative abundance of 5 most abundant genera identified from urinary and vaginal samples.

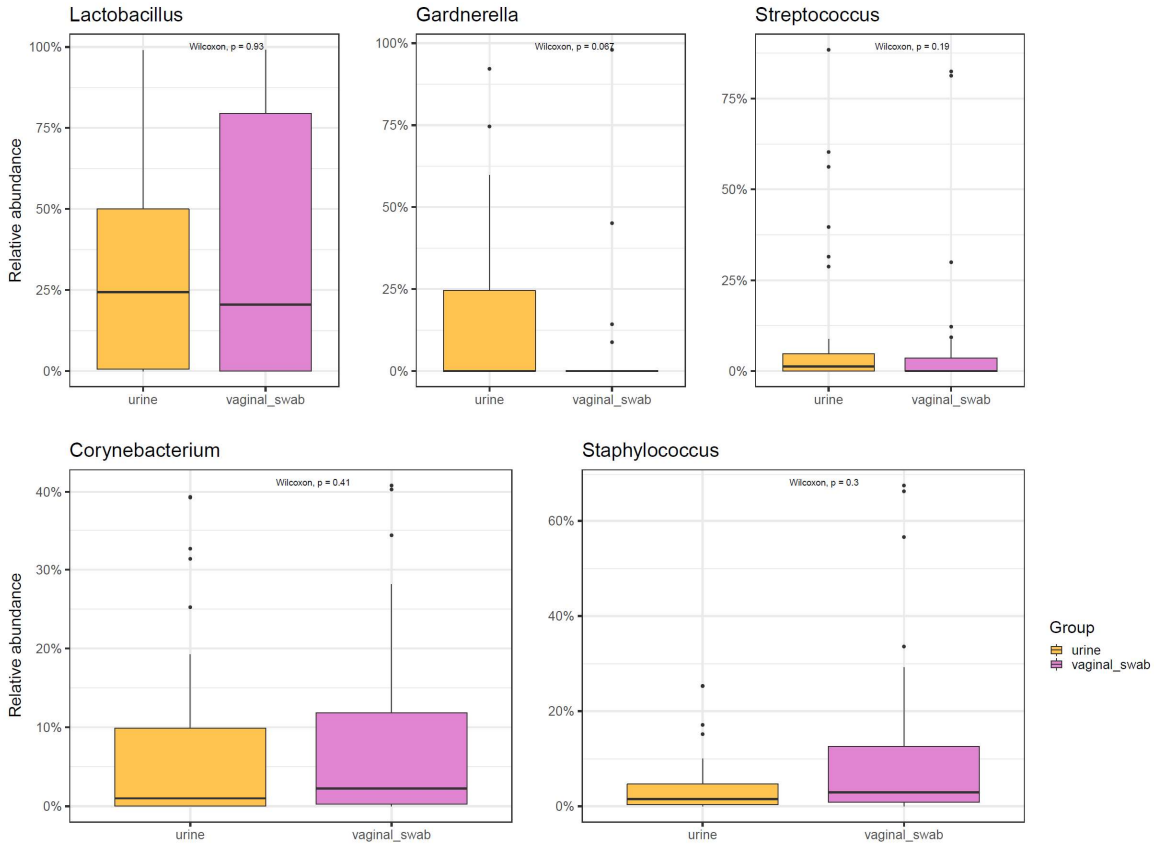


Figure 7. Cumulative relative and absolute abundance of *Lactobacillus* and *Gardnerella* species.

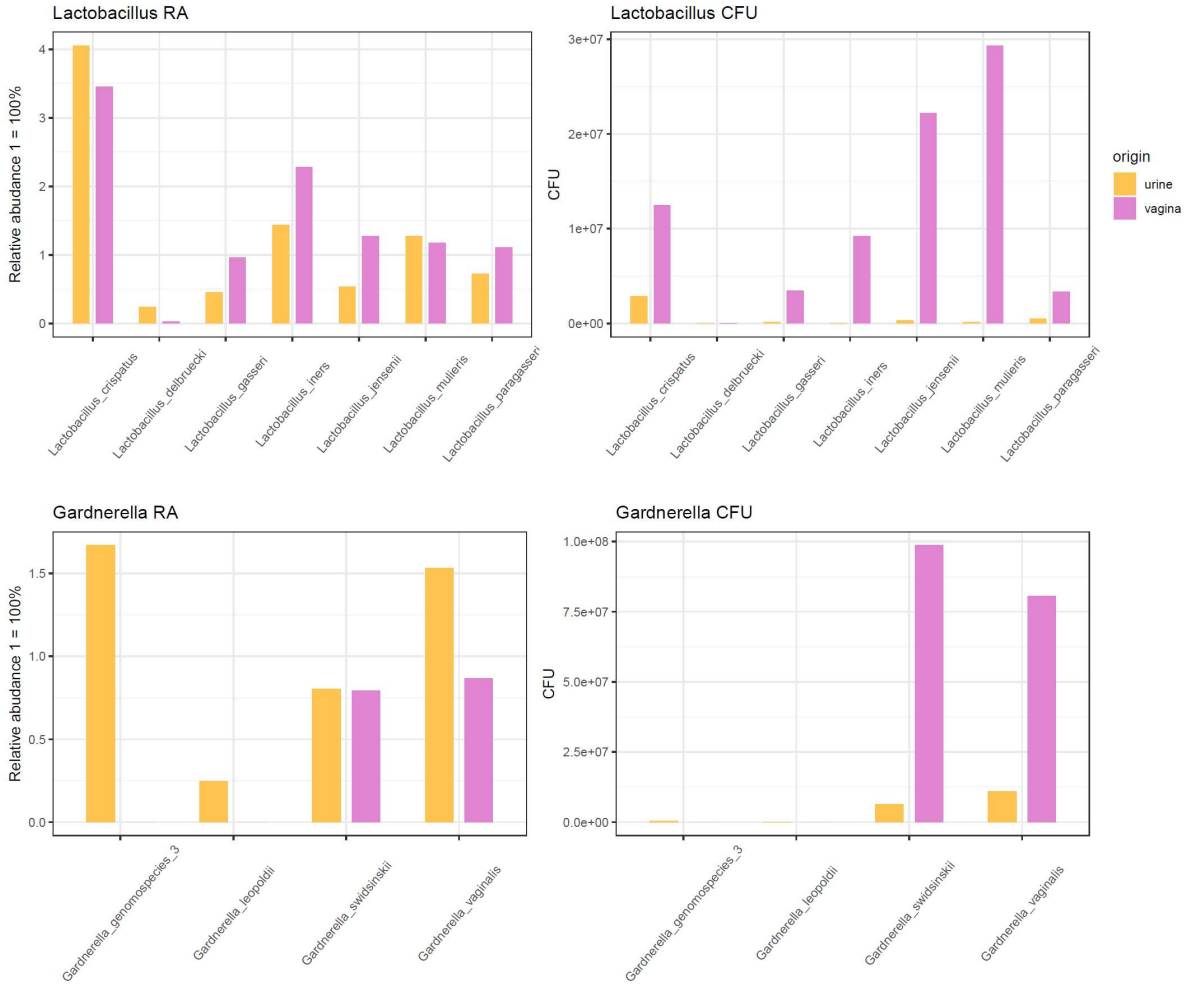
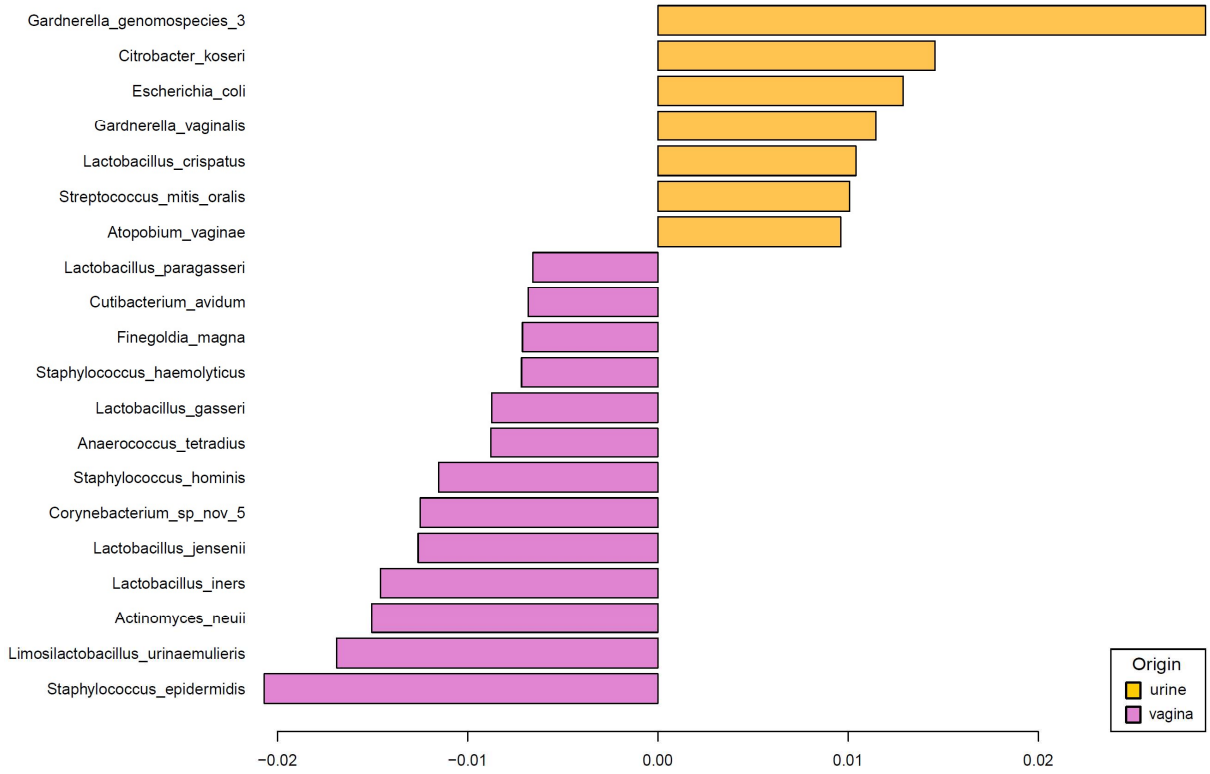


Figure 8. Species that contributed the most to compositional differences between urinary and vaginal microbiome.



Additional material associated with the manuscript entitled “Step towards understanding interconnection between vaginal and urinary tract microbiome” is available through the following links:

Supplementary Table S1:
https://docs.google.com/spreadsheets/d/1J1_6g0QkIsNNGfITvsaOU-tR_cRpL7Ye/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Supplementary Figures S1-S10:
https://drive.google.com/drive/folders/1RneYTXyxeCBU_325vheuw1PgEwHOXYSZ?usp=sharing

3.2. Female Urinary Tract Microbiome in diseased host

A sneak peek at recurrent urinary tract infections in young adult women

Urinary microbiome of overactive bladder: toward a better understanding through improved high-throughput community analyses

A sneak peek at recurrent urinary tract infections in young adult women

Ksiezarek M[1,2], Ugarcina Perovic S[3], Grosso F[1,2], Peixe L.[1,2]

[1]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[2]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

[3]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

Manuscript in preparation

Sir,

Urinary tract infections (UTIs) are one of the most common infections, being estimated that 50% of women will develop at least one UTI in their lifetime (1). Regarding recurrent urinary tract infections (rUTIs) there is no consensus on their definition in the existing literature; however, for the purpose of this study we will consider the European Association of Urology guidelines that refer rUTI as recurrences of uncomplicated and/or complicated UTIs, with a frequency of at least three UTIs per year or two UTIs in six months (2).

Although incidence of rUTI greatly increases at postmenopausal age, which is explained as the result of estrogen levels decrease with consequent reduction of glycogen, leading to female urogenital microbiome changes (e.g., decrease in *Lactobacillus* sp. relative abundance - RA), rUTI affect women at different ages with great impact on their well-being (1). Still, there is no universally accepted treatment algorithm (nor truly effective) for rUTI in women, with current approaches varying among broadening the antibiotic spectrum, increasing treatment length, and/or increasing antibiotic dosage.

The increasing knowledge on the female urogenital microbiome (FUM) also grew the interest in investigating its compositional differences in both healthy and patients suffering from rUTI. This study constitutes another step forward in the understanding of rUTI, and how FUM may contribute for this impactful condition in women's life.

This study included four urine samples from three reproductive-age women (range 25-36 years) suffering from rUTIs. Samples 13a, 14b and 32a were obtained in the absence of UTI symptoms while 32b was obtained from a woman during the course of UTI. First morning voided midstream urine samples were subjected to an extended culturomic analysis within 2 hours from sample collection (3). Each colony morphotype was quantified to approximate number of colonies forming units per milliliter (CFU/ml) and up to 5 colonies of the same morphotype were isolated and subjected to identification. In addition, dipstick test (Combur-Test, Roche) and microscopy examination were performed. All isolates were subjected to matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) VITEK MS system (bioMérieux, France), using in-vitro diagnostic database version 3.0. In case of no identification by MALDI-TOF MS isolates were subjected to 16S rRNA gene sequencing and/or other suitable genotypic biomarkers (*phe*S, *rpo*B, *rec*N, *cpn*60). All community analyses were based on RA-% calculated as the CFU percentage of identified species from total CFU/ml count. The bacterial load varied between 10^4-10^6 CFU/mL, with a range of 17-30 of isolates being characterized per sample. It is of note that it was observed a slightly increased in the mean number of species/sample detected in comparison with our previous study with a healthy cohort (24

vs 19 species/sample, respectively), with all but one (14b, dominance of *Corynebacterium coyleae*) presenting community structures dominated by *Lactobacillus*: *Lactobacillus iners* (13a - RA 49%), *Lactobacillus crispatus* (32a – RA 34%), *L. iners* and *Lactobacillus jensennii* (32b – RA 38% and 19%, respectively) (Table S1) (3). The great abundance of *L. iners* in the analyzed samples has already been linked with vaginal dysbiosis, being speculated that its presence is not always indicative of a balanced microbiome (3, 4). Moreover, the community analysis also revealed some UTI-associated species such as *Escherichia coli*, *Klebsiella pneumoniae* or *Varibaculum* sp..

The comparison of the two samples belonging to the same donor but obtained without symptoms (32a) and during the course of UTI (32b) revealed that, during symptomology, was observed a marked increase in the bacterial load, from 10^4 to 10^6 CFU/ml together with an increase in species diversity (17 to 22 species). Moreover, although *E. coli* absolute abundance increased during the course of UTI, its relative abundance decreased, comparing to the state where no symptoms were present (from 25.4% to 13.4%, Fig.1)., associated with appearance of *L. iners* and *Lactobacillus jensennii* in high abundance (32b – RA 38% and 19% respectively (Fig.1; Table S1) and reduction in *L. crispatus* abundance This could imply that, similarly to what is known for vaginal microbiome, increased FUM diversity may be associated with particular urological conditions. Remarkably, *E. coli* detected in both samples of this donor represented the same strain of extraintestinal pathogenic *E. coli* belonging to ST12 and B2 phylogenetic group (5). These *E. coli* possessed a variety of virulence-associated determinants, for instance adhesins e.g., *fimH*, *sfa/focDE*, *papC*, *papG* or toxins e.g., *hlyA*, *cnf1*, *vat* that could contribute to their attachment to urothelium and putative pathogenicity in urinary tract (5). Thus, the symptoms may result not only from *E. coli* activity but also from the synergistic action of other microbiome members.

This pilot study's results reinforce the idea that FUM may be a target for evaluation, prevention and treatment approaches for rUTI. Nevertheless, data needs to be further explored, namely concerning functional characterization, and also involving larger well-characterized cohorts.

Acknowledgements

The authors would like to thank all women voluntarily providing the samples and participating in this study and also thank Helena Ramos and Paulo Pinto (Hospital Geral de Santo António, Porto, Portugal) for their technical support with MALDI-TOF MS-based microbial identification. The bioMérieux (Portugal, Lda) provided equipment and material for MALDI-TOF MS analysis, and had no role in the study design, data collection and analysis, decision to publish, or writing of the manuscript.

Funding

This work was financed by national funds from FCT - Fundação para a Ciência e a Tecnologia, I.P., in the scope of the project UIDP/04378/2020 and UIDB/04378/2020 of the Research Unit on Applied Molecular Biosciences - UCIBIO and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy - i4HB. MK is supported by Fundação para a Ciência e Tecnologia, I.P. - FCT [SFRH/BD/132497/2017]. SUP was supported from project NORTH-01-0145-FEDER-000024, and FG is supported by national funds through FCT in the context of the transitional norm [DL57/2016/ CP1346/CT0034]. The funding bodies had no role in the design of the study; collection, analysis and interpretation of data; or writing the manuscript.

Ethics approval and consent to participate

Approval of the study was obtained from the Faculty of Pharmacy (University of Porto, Porto, Portugal) Ethics Committee. Procedures performed in the study were all in accordance with the ethical standards of the institutional and national research committee, with the 1964 Helsinki Declaration, and its later amendments. All individual participants included in the study had given written informed consent.

Competing interests

The authors declare that there are no conflicts of interest.

References

1. Jung C, Brubaker L. 2019. The etiology and management of recurrent urinary tract infections in postmenopausal women. Climacteric. 22(3):242-249. doi: 10.1080/13697137.2018.1551871.

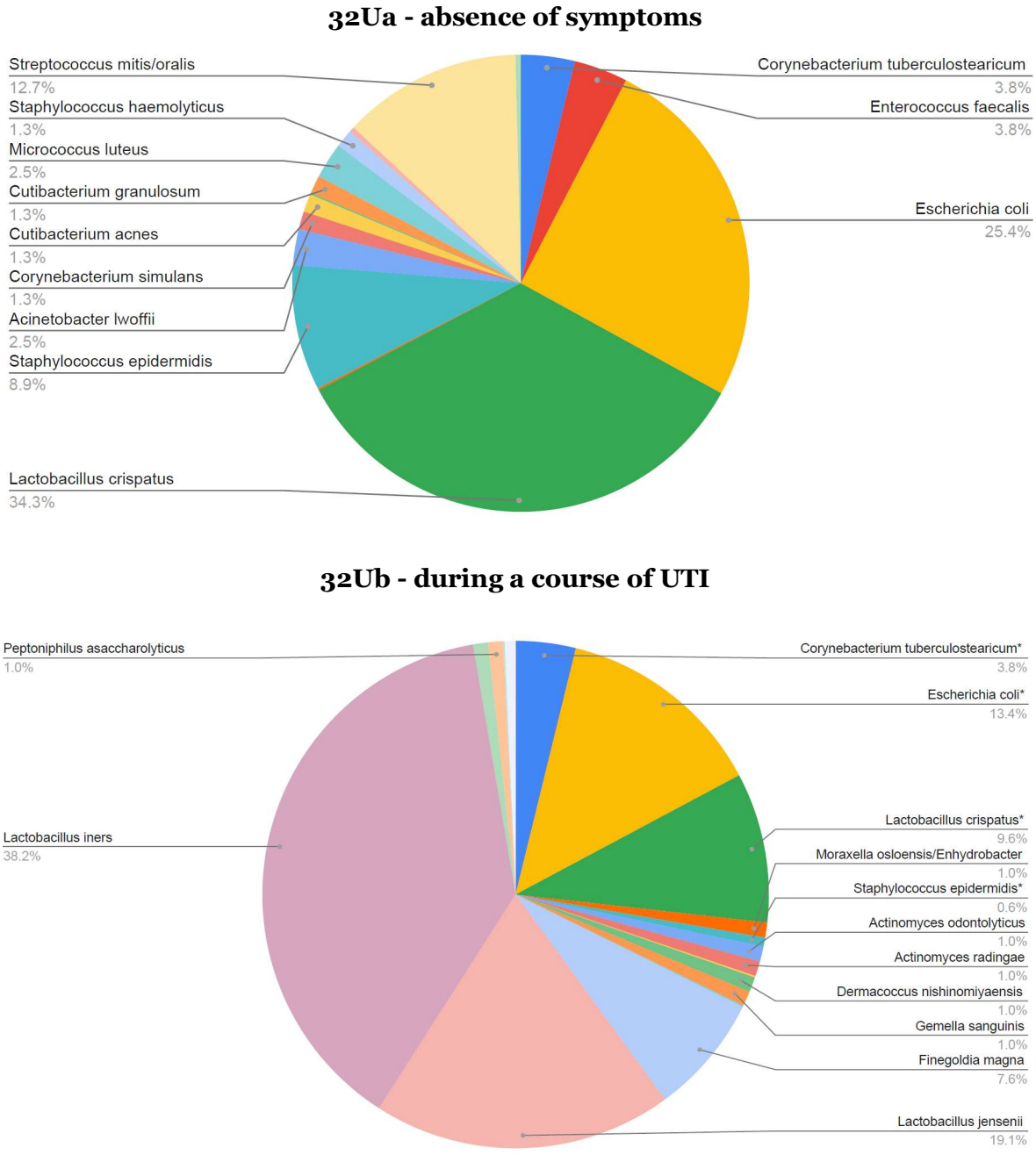
2. Bonkat G, Bartoletti R, Bruyère F, Cai T, Geerlings SE, Köves B, Schubert S, Wagenlehner F., *et al*. 2020. EAU guidelines on Urological Infections. Edn. presented at the EAU Annual Congress Amsterdam the Netherlands 2020. ISBN 978-94- 92671-07-3.

3. Ksiezarek M, Ugarcina-Perovic S, Rocha J, *et al.* 2021. Long-term stability of the urogenital microbiota of asymptomatic European women. *BMC Microbiol* 21, 64 https://doi.org/10.1186/s12866-021-02123-3.

4. Petrova M I, Reid G, Vaneechoutte M, Lebeer S. 2017. *Lactobacillus iners*: Friend or Foe? Trends in Microbiology, 25(3), 182–191. doi:10.1016/j.tim.2016.11.007.

5. Ksiezarek M, Novais A, Peixe L. 2021. The darkest place is under the candlestick - healthy urogenital tract as a source of UTI-related *Escherichia coli* lineages. bioRxiv 2021.08.06.455400; doi: https://doi.org/10.1101/2021.08.06.455400.

Figure 1. Urinary tract microbiome composition (relative abundance) in rUTI woman in the absence of symptoms (sample 32Ua) and during a course of rUTI (32Ub).



Additional material associated with the manuscript entitled “A sneak peek at recurrent urinary tract infections in young adult women” is available through the following links:

Supplementary Material:
Table S1
https://docs.google.com/spreadsheets/d/1KBxccItym-U-TwqP41s-sri8VjnVTsUa/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Urinary microbiome of overactive bladder: toward a better understanding through improved high-throughput community analyses

Svetlana Ugarcina-Perovic[1], Magdalena Ksiezarek[2,3], Joana Rocha[4], Márcia Sousa[2,3], Teresa Gonçalves Ribeiro[2,3], Filipa Grosso[2,3], Luís Vale[5,6], Carlos Silva[5,6], Paulo Dinis[5,6], Tiago Antunes-Lopes[5,6,7], Luísa Peixe[2,3]#

[1]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

[2]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[3]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

[4]UCGenomics/GenomePT, Laboratório de Sequenciação e Genómica Funcional da Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal

[5]Department of Urology, Hospital de S. João, Porto, Portugal

[6]Faculty of Medicine, University of Porto, Portugal

[7]i3S – Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal

Manuscript in preparation

Abstract

Introduction. Recent studies based on advanced culture-independent and -dependent methodological approaches have suggested possibility of urinary tract microbiome dysbiosis contributing to overactive bladder syndrome (OAB). However, compositional data on OAB microbiome at species level is still critically lacking, and the pathogenesis of OAB remains unclear.

Aim. In this preliminary study, we aim to comprehensively characterize the urinary tract microbiome of heterogeneous group of OAB patients, using two complementary methodologies.

Methodology. Mid-stream urine samples were collected from 6 pre- and postmenopausal women diagnosed with OAB symptoms in Hospital São João (Porto, Portugal). Symptoms severity was assessed with Overactive Bladder Symptom Score (OBSS). To characterize urinary tract microbiome, we used extended culturomics – different media and incubation conditions, MALDI-TOF/MS and/or genotypic biomarkers for isolate identification. Concurrently, all samples were subjected to amplicon sequencing using V1-V8 regions of 16S rRNA by PacBio SMRT cell technology.

Results: In all patients (OBSS range = 7 - 9), 217 species were identified, with high richness (median=58) and high load of bacteria (10^4-10^8 CFU/ml), although no criteria for urinary tract infection were found based on routine urinalysis and standard culture. We identified a dominance of *Lactobacillus iners*, *Gardnerella leopoldii, Gardnerella vaginalis* and *Gardnerella* genomospecies 3 and diverse putative urogenital pathogens (e.g., *Escherichia coli*, *Klebsiella pneumoniae, Enterococcus faecalis*, *Ureaplasma urealyticum, Ureaplasma parvum*) and others more recently recognized (e.g., *Atopobium vaginae*, *Actinotignum schaalii*). Each patient had at least one opportunistic pathogen (median=5). Preliminary molecular screening showed that some strains possessed virulence-associated features (*E. coli*). The patient with the most severe symptoms had 71 species of which some are putatively relevant e.g., *Gardnerella vaginalis*, *Ureaplasma urealyticum*, *Atopobium vaginae*.

Conclusion. All OAB patients presented a diverse urinary microbiome enriched in bacterial species known to be involved in urinary tract infection. Analysis of our small cohort indicated that OAB urinary tract microbiome can be characterized by high abundance of *Lactobacillus iners* and diverse *Gardnerella* species.

Introduction

Healthy bladder has its own distinct community of bacteria – urinary microbiome – that has been detected by advanced culture-independent sequencing techniques and enhanced urine culturing (1–3), but missed by standard urine culture (4). Furthermore, there is growing evidence that the urinary microbiome may play a critical role in human urinary tract health and disease state, including common lower urinary tract conditions [(e.g., urinary tract infection (UTI) and overactive bladder syndrome (OAB)] (4–9). Although OAB causes significant distress and impairment on quality of life, this complex symptom with a higher prevalence in women than men is still of unknown etiology and undefined effective treatment plan (10–12).

In women with OAB, a particular urinary microbiome with an increase in abundance of *Gardnerella* and decrease of *Lactobacillus* was evidenced (4, 7). Also, OAB studies have reported associations between urinary symptom severity and several genera, e.g., *Proteus* (7) and *Aerococcus* (8), and several emerging urogenital pathogens, e.g., *Actinotignum schaalli* (4, 13) and *Atopobium vaginae* (14). Interestingly, previously recognized commensal species, *Lactobacillus gasseri*, has been also reported as a OAB specific member (4). Moreover, a study conducted by Wu et al. (8) suggested the connection between urinary microbiome and central nervous system. Nevertheless, further evidence of particular urinary tract microbiome and accurate bacterial species identification in OAB patients that could guide diagnostic and treatment are required. Moreover, it is a prerequisite for the diagnoses of OAB the absence of UTI obtained by standard urine culture protocol, which has recognized detection limitations (15, 16) may lead to flaws in OAB diagnoses. In this study, we aim to comprehensively investigate the bacterial composition of urinary tract microbiome in an age-heterogeneous group of women diagnosed with OAB, using extended culturomics and improved amplicon sequencing approaches.

Methods

Participant recruitment and sample collection

This study was approved by ethical committee of the Hospital São João (Porto, Portugal). Mid-stream urine samples were collected between September 2017 and June 2018 from 6 female patients (36-80 years) diagnosed with OAB. Symptoms severity was assessed with Overactive Bladder Symptom Score (5). Exclusion criteria included: current UTI (based on urine dipstick, microscopic exam, and standard urine culture), antibiotic exposure in the past 4 weeks, pregnancy, history of pelvic radiotherapy, bladder tumour, urolithiasis, and neurogenic

voiding dysfunction. A questionnaire was conducted concerning personal and health information that was encrypted, ensuring data confidentiality. Urine samples were subjected to standard urine culture at the hospital microbiology laboratory, whereas extended culturomic protocol was performed at the research laboratory (Faculty of Pharmacy, University of Porto, Porto, Portugal) up to 2 hours after sample collection and transport at 4 °C. Concurrently, samples were pre-treated for community amplicon sequencing analysis and stored at -80 °C.

Standard culture and extended culturomic analysis

The standard culture protocol involved inoculation of ChromID CPS Elite chromogenic culture medium (bioMérieux, France) with 1 µl of uncentrifuged urine and aerobic incubation at 37 °C, for 24 hours. Only the confluent growth of a single organism, with a count of ≥ 10^5 colony forming units (CFU)/ml, was presumed as relevant growth.

The extended culturomic protocol included distribution of 100 µl of the urine samples on Columbia agar with 5% sheep blood (BAP, Biogerm, Portugal) and chromogenic agar plate (CAP, HiCrome UTI, HiMedia, India) supplemented with gelatine, yeast extract, starch, glucose and Tween 80 as previously described (17, 18). BAPs and CAPs were incubated under aerobic and microaerophilic conditions (GENbox MICROAER, bioMérieux, France), at 37 °C for up to 48 h. Additionally, BAPs were incubated under anaerobic conditions (GENbox ANAER, bioMérieux, France), at 37 °C for 48 h. Each morphologically distinct colony type was counted. All selected isolates (1-5 colonies of each morphology) were subjected to matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), using the VITEK® MS and In-Vitro Diagnostic database version 3.0 (bioMérieux, France). Unidentified isolates and isolates belonging to species for which MALDI-TOF MS has insufficient resolution power (Perovic SU in preparation) were further subjected to 16S rRNA gene and other genetic markers (*pheS*, *cpn60*, *rpoB* or *recN*) and/or PCR assays for the detection of species-specific genes (*dltS*, *sodA* and *malB*).

DNA extraction and community amplicon sequencing

Sample pre-treatment, prior to DNA extraction, included 20 ml of urine centrifuged at 5,500 rpm for 15 min, with the resulting pellet suspended in 1 ml of phosphate buffered saline and stored at -80 °C until further processing. Genomic DNA was extracted using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Germany), according to the manufacturer's protocol and following the pre-treatment protocol for Gram-negative bacteria. Extracted DNA was stored

at 4 °C until sent for targeted amplicon sequencing of V1-V8 hyper-variable regions of the 16S rRNA gene using the Pacific Biosciences Single-Molecule Real-Time technology (PacBio SMRT, GATC Biotech AG, Germany). The obtained sequences were quality filtered by UCHIME (version 4.2.40) (19). Chimera-free sequences were subjected to BLASTn analysis using non-redundant 16S rRNA reference sequences with an E-value cutoff of 1e-06 (20). If the final number of high-quality reads after all filtering steps was less than 1000, the corresponding sample was excluded. Reference 16S rRNA gene sequences were obtained from the Ribosomal Database Project Classifier (21). Taxonomic classification was based on the NCBI Taxonomy (22).

Results

In all patients (OBSS range = 7 - 9, median = 8), we detected a high species richness (217 species, range = 16-105 species/sample, median = 58 species/sample), with community amplicon sequencing revealing a higher number of species than extended culturomics (Table S1, S2). High bacterial load (1x10^4 - 1x10^8 CFU/ml) was observed in all samples. The information on OAB patients and species detected by extended culturomics and community amplicon sequencing are provided in Table S1 and Table S2, respectively.

The most abundant five species per urine sample detected by culturomics and amplicon sequencing are presented in Figure 1A and 1B, respectively. We detected a predominance of *Gardnerella* genomospecies 3 and *Gardnerella leopoldii* by culturomics, and *Lactobacillus iners* by amplicon sequencing in **OAB1 and OAB2** (Figure 1A and 1B), and *Bifidobacterium* spp. and *Corynebacterium aurimucosum* by culturomics with *Lactobacillus gasseri* detected in high relative abundance by amplicon sequencing in **OAB5** (Figure 1A and B). Remarkably, predominance of *Escherichia coli* (1x10^6 CFU/ml) and *Enterococcus faecalis* (1x10^8 CFU/ml), recognized as putative uropathogens, were identified in two samples (**OAB4 and OAB6**, respectively; Figure 1A). One urine sample collected from a patient with the most severe OAB symptoms (OBSS = 9) presented a urinary microbiome dominated by *Gardnerella vaginalis* (46% by culturomics) and *Ureaplasma urealyticum* (44% relative abundance by sequencing) (**OAB3**; Figure 1A and 1B, respectively).

In most samples diverse bacterial species with urogenital pathogenic potential were detected (**Table 1**). Each sample had at least one (median = 5) urogenital opportunistic pathogens detected in a range of 10^1-10^8 CFU/ml. The most prevalent urogenital pathogenic species among patients were *Enterococcus faecalis* (n=5) and *Streptococcus anginosus* (n=5), followed by *Campylobacter ureolyticus* (n=4) and *Ureaplasma parvum* (n=4). The

community amplicon sequencing of **OAB5** sample revealed the highest species richness (105 species), including a high diversity of emerging pathogens *Campylobacter ureolyticus, Ureaplasma parvum, Oligella urethralis* and *Alloscardovia omnicolens* that were missed by extended culturomics. In other samples, emerging urogenital pathogens were detected with more than 1x10^3 CFU/ml, such as *Atopobium vaginae* and *Corynebacterium coyleae*, while others, namely *Morganella morganii* and *Proteus mirabilis,* were detected in less than 1x10^3 CFU/ml. Noteworthy, many of these putative pathogenic species were not detected in our cohort of asymptomatic women such as *Morganella morgani, Klebsiella pneumoniae, Proteus mirabilis*, *Pseudomonas luteola* and *Citrobacter portucalensis* (Perovic SU in preparation, Ksiezarek 2021 BMC).

Preliminary strain level analysis of *E. coli* isolated from OAB women revealed that only *E. coli* from **OAB5** belong to phylogenetic group B2 and was predicted as extraintestinal pathogenic *E. coli* (ExPEC), with several virulence-associated features e.g., *fimH*, *sfa/focDE*, *cnf1*, *vat*, *fyuA*, *iroN*, *usp*, *ompT*, *malX* (unpublished data). Remaining *E. coli* strains from **OAB2** sample were not classified as ExPEC and presented just a few putatively virulence-associated genes (unpublished data).

Discussion

In the present study, we capture a higher number of bacterial species than previously reported for OAB patients (4, 7). Moreover, each patient had at least one recognized urogenital pathogen, occasionally in high load (up to 10^8 CFU/ml).

Contrary to the previous findings of a low *Lactobacillus* abundance in OAB patients (4) we detected *Lactobacillus iners* in high relative abundance. It is important to note that, beside *Gardnerella vaginalis* and *Gardnerella leopoldii*, we identified, for the first time in OAB patients *Gardnerella* genomospecies 3 with still unknown pathogenicity (23, 24). Significant impact of the virulence potential of diverse bacterial strains on UTI risk and outcome was already observed (25), stressing the ambiguity of strain level pathogenicity that should be explored especially in OAB-associated microbiome. We demonstrated a sneak peek data on genomic characteristics of *E. coli* strains isolated from OAB microbiome. Interestingly, not all *E. coli* strains demonstrated equally relevant pathogenic potential. However, our additional genomic analysis of *Prevotella brunnea* isolated from **OAB1** sample (RA 2%) showed that this strain was predicted as a human pathogen and possessed a variety of virulence-associated traits e.g., proteins required for sialic acid synthesis, outer membrane proteins, clostripain family protein, proteins involved in lipopolysaccharide modification (Ksiezarek et al., in

preparation). *P. brunnea* appear to be able to stimulate host-microbe interactions that might lead to an unhealthy status. This could support hypothesis that microbial interactions and/or low abundant members play a role in OAB aetiology.

Comparing to the healthy urinary microbiome (Perovic SU in preparation, 3), although of not similarly aged women, OAB patients presented slightly higher median of species/sample (58 vs 53). Microbiome composition enriched in *L. iners* identified here in OAB women was also identified in asymptomatic cohort (Perovic SU in preparation). *L. iners* seems to have a variety of genomic determinants for both, promotion of health status and those associated with host disease, thus its activity is likely strain specific and should be further investigated (26). OAB microbiome was also characterized by different *Streptococcus* and *Staphylococcus* species and some species were only detected in OAB patients, e.g., *Proteus mirabilis*, *Citrobacter portucalensis*. This preliminary comparison would suggest that OAB microbiome has more and/or higher prevalence of putative pathogens, however this must be verified on larger cohort with appropriately selected control group.

Our study based both on 16S rRNA gene amplicon sequencing and advanced urine culture showed that OAB urinary microbiome is diverse and enriched in members carrying potential virulent features. Species detection by third generation DNA sequencing using PacBio technology confirmed and complemented detection by extended culturomics, but it still lacks quantitative specificity. Moreover, some pathogenic species were missed by extended culture, likely due to special culture and metabolic requirements (e.g., *Ureaplasma parvum*) (27).

Conclusions

Our preliminary study provides a detailed composition of heterogeneous bacterial community in urine samples from female OAB patients. All OAB patients presented a diverse urinary microbiome enriched in bacterial species known to be involved in urinary tract infection. For the first time, we identified *Gardnerella* genomospecies 3 in OAB women. Our findings indicated that OAB urinary community can be characterized by high abundance of *Lactobacillus iners* and diverse *Gardnerella* species. Finally, our study based on accurate and reliable species level profiling stress the importance on further strain level exploration in order to understand bacterial involvement in OAB syndrome.

Funding

This work was supported by the Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal, and by the UCIBIO – Applied Molecular Biosciences Unit, Laboratory of Microbiology, Department of Biological Sciences, REQUIMTE, Faculty of Pharmacy, University of Porto, 4050-313 Porto, which are financed by national funds from FCT (Fundação para a Ciência e a Tecnologia, I.P.) (UIDB/MULTI/04378/2020 and UIDP/04378/2020). SUP was supported as researcher from the project NORTH-01-0145-FEDER-000024, MK and MS by an FCT PhD grant (SFRH/BD/132497/2017 and SFRH/BD/05038/2020, respectively), and JR by a postdoctoral fellowship from ICETA (UID/MULTI/04378/2013). FG is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0034].

Authors' contributions

SUP performed MALDI-TOF MS analysis and interpreted the data. MK and SUP performed culturomics analysis, isolate collection, data analysis and wrote the manuscript. MK, TGR, JR, MS, EC and FG performed genotypic isolates' identification. TGR supervised MS and EC. AN contributed to MALDI-TOF MS analysis, *Escherichia* and *Klebsiella* species identification. LP contributed to reviewing the manuscript, supervising and funding the project. All authors read and approved the manuscript.

References

1. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, Fitzgerald M, Mueller ER, Schreckenberger P, Dong Q, Nelson DE, Brubaker L. 2012. Evidence of uncultivated bacteria in the adult female bladder. J Clin Microbiol 50:1376–1383.
2. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. Journal of Clinical Microbiology 52:871–876.
3. Ksiezarek M, Ugarcina-Perovic S, Rocha J, Grosso F, Peixe L. 2021. Long-term stability of the urogenital microbiota of asymptomatic European women. BMC Microbiol 21:64.
4. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014. The Female Urinary Microbiome: a Comparison of Women with and without Urgency Urinary Incontinence. mBio 5.

5. Khasriya R, Sathiananthamoorthy S, Ismail S, Kelsey M, Wilson M, Rohn JL, Malone-Lee J. 2013. Spectrum of bacterial colonization associated with urothelial cells from patients with chronic lower urinary tract symptoms. J Clin Microbiol 51:2054–2062.
6. Karstens L, Asquith M, Davin S, Stauffer P, Fair D, Gregory WT, Rosenbaum JT, McWeeney SK, Nardos R. 2016. Does the Urinary Microbiome Play a Role in Urgency Urinary Incontinence and Its Severity? Front Cell Infect Microbiol 6:78.
7. Curtiss N, Balachandran A, Krska L, Peppiatt-Wildman C, Wildman S, Duckett J. 2017. A case controlled study examining the bladder microbiome in women with Overactive Bladder (OAB) and healthy controls. Eur J Obstet Gynecol Reprod Biol 214:31–35.
8. Wu P, Chen Y, Zhao J, Zhang G, Chen J, Wang J, Zhang H. 2017. Urinary Microbiome and Psychological Factors in Women with Overactive Bladder. Front Cell Infect Microbiol 7:488.
9. Gill K, Kang R, Sathiananthamoorthy S, Khasriya R, Malone-Lee J. 2018. A blinded observational cohort study of the microbiological ecology associated with pyuria and overactive bladder symptoms. Int Urogynecol J 29:1493–1500.
10. Antunes-Lopes T, Cruz F. 2019. Urinary Biomarkers in Overactive Bladder: Revisiting the Evidence in 2019. Eur Urol Focus 5:329–336.
11. Peyronnet B, Mironska E, Chapple C, Cardozo L, Oelke M, Dmochowski R, Amarenco G, Gamé X, Kirby R, Aa FVD, Cornu J-N. 2019. A Comprehensive Review of Overactive Bladder Pathophysiology: On the Way to Tailored Treatment. European Urology 75:988–1000.
12. Ackerman AL, Chai TC. 2019. The Bladder is Not Sterile: an Update on the Urinary Microbiome. Curr Bladder Dysfunct Rep https://doi.org/10.1007/s11884-019-00543-6.
13. Price TK, Hilt EE, Thomas-White K, Mueller ER, Wolfe AJ, Brubaker L. 2020. The urobiome of continent adult women: a cross-sectional study. BJOG 127:193–201.
14. Fok CS, Gao X, Lin H, Thomas-White KJ, Mueller ER, Wolfe AJ, Dong Q, Brubaker L. 2018. Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. Int Urogynecol J 29:1765–1771.
15. Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, Brubaker L, Wolfe AJ, Mueller ER, Schreckenberger PC. 2016. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. J Clin Microbiol 54:1216–1222.
16. Sathiananthamoorthy S, Malone-Lee J, Gill K, Tymon A, Nguyen TK, Gurung S, Collins L, Kupelian AS, Swamy S, Khasriya R, Spratt DA, Rohn JL. Reassessment of Routine

Midstream Culture in Diagnosis of Urinary Tract Infection. Journal of Clinical Microbiology 57:e01452-18.

17. Alves P, Castro J, Sousa C, Cereija TB, Cerca N. 2014. *Gardnerella vaginalis* outcompetes 29 other bacterial species isolated from patients with bacterial vaginosis, using in an in vitro biofilm formation model. J Infect Dis 210:593–596.
18. Man JCD, Rogosa M, Sharpe ME. 1960. A Medium for the Cultivation of Lactobacilli. Journal of Applied Bacteriology 23:130–135.
19. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200.
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.
21. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141-145.
22. Federhen S. 2012. The NCBI Taxonomy database. Nucleic Acids Res 40:D136-143.
23. Putonti C, Thomas-White K, Crum E, Hilt EE, Price TK, Wolfe AJ. 2021. Genome Investigation of Urinary *Gardnerella* Strains and Their Relationship to Isolates of the Vaginal Microbiota. mSphere 6.
24. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P. 2019. Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. Int J Syst Evol Microbiol 69:679–687.
25. Schreiber HL, Spaulding CN, Dodson KW, Livny J, Hultgren SJ. 2017. One size doesn't fit all: unraveling the diversity of factors and interactions that drive *E. coli* urovirulence. Ann Transl Med 5:28.
26. Petrova MI, Reid G, Vaneechoutte M, Lebeer S. 2017. *Lactobacillus iners*: Friend or Foe? Trends in Microbiology 25:182–191.
27. Frølund M, Björnelius E, Lidbrink P, Ahrens P, Jensen JS. 2014. Comparison between Culture and a Multiplex Quantitative Real-Time Polymerase Chain Reaction Assay Detecting *Ureaplasma urealyticum* and *U. parvum*. PLOS ONE 9:e102743.

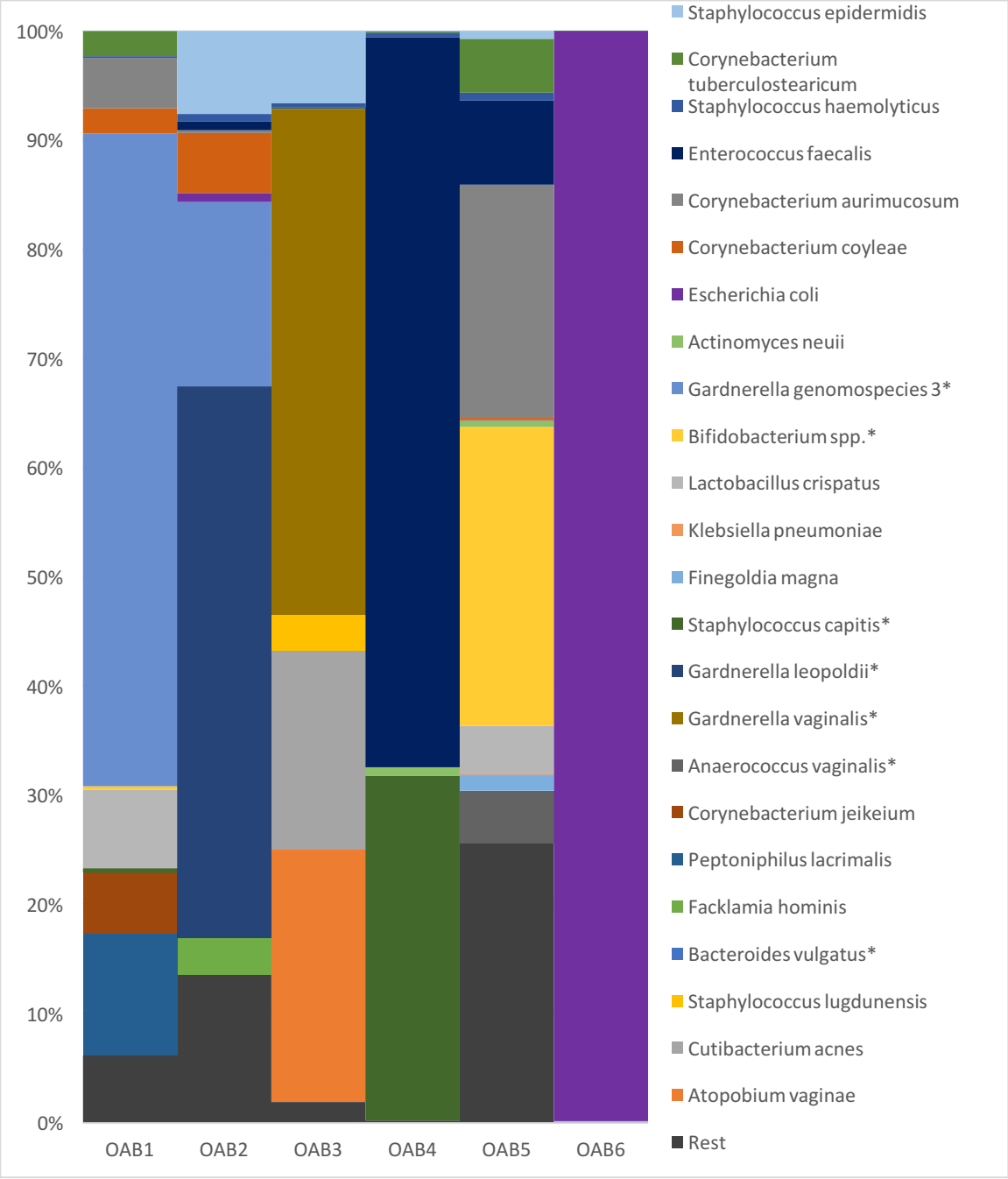
Table 1. Opportunistic pathogens associated with the urogenital tract. Species are listed in order of decreasing detection frequency in OAB samples. ND - not detected; RA - relative abundance.

Species	Culturomics			Amplicon sequencing	
	Samples	CFU/ml	RA (%)	Samples	RA (%)
Enterococcus faecalis	OAB2, OAB3, OAB4, OAB5, OAB6	10^1-10^8	0.01-66.9	OAB3, OAB4, OAB5	0.16-34.04
Streptococcus anginosus	OAB1, OAB2, OAB3, OAB5	10^2-10^3	0.2-4.3	OAB1, OAB3, OAB4, OAB5	0.04-3.68
Campylobacter ureolyticus	OAB4	10^2	<0.01	OAB1, OAB2, OAB4, OAB5	0.11-2.37
Ureaplasma parvum	ND	ND	ND	OAB1, OAB2, OAB4, OAB5	0.07-6.68
Corynebacterium coyleae	OAB1, OAB2, OAB5	10^2-10^4	0.23-5.48	OAB1, OAB2, OAB5	0.06-3.07
Escherichia coli	OAB2, OAB5, OAB6	10^2-10^6	0.06-99.82	OAB3, OAB5, OAB6	0.03-99.35
Ureaplasma urealyticum	ND	ND	ND	OAB1, OAB2, OAB3	0.09-44.37
Gardnerella genomospecies 3	OAB1, OAB2	10^4-10^5	16.85-59.78	ND	ND
Pseudoglutamicibacter cumminsii	OAB1, OAB2	10^3	0.52-1.47	OAB1	0.57
Corynebacterium urealyticum	OAB2	10^3	2.95	OAB1, OAB2	0.06-0.12
Klebsiella pneumoniae	OAB5, OAB6	10^1-10^3	0.04-0.15	OAB5	0.09
Morganella morganii	OAB2, OAB3	10^1	0.02-0.03	OAB3	0.03
Pseudomonas aeruginosa	OAB6	10^1	<0.01	OAB3	0.08
Staphylococcus aureus	OAB2, OAB3	10^1-10^2	0.04-0.13	OAB3	0.05
Atopobium vaginae	OAB3	10^4	23.16	OAB3	1.07
Aerococcus urinae	OAB5	10^1	0.15	OAB5	0.04
Gardnerella leopoldii	OAB2	10^5	50.54	ND	ND
Proteus mirabilis	OAB2	10^2	0.34	ND	ND
Gardnerella vaginalis	OAB3	10^4	46.32	ND	ND
Anaerococcus vaginalis	OAB5	10^3	4.98	ND	ND
Actinotignum schaalii	OAB5	10^2	0.8	ND	ND
Pseudomonas luteola	OAB5	10^1	0.02	ND	ND
Staphylococcus saprophyticus	OAB6	10^2	0.01	ND	ND
Oligella urethralis	ND	ND	ND	OAB1	0.12
Pseudomonas putida	ND	ND	ND	OAB3	0.03
Stenotrophomonas maltophilia	ND	ND	ND	OAB3	0.03

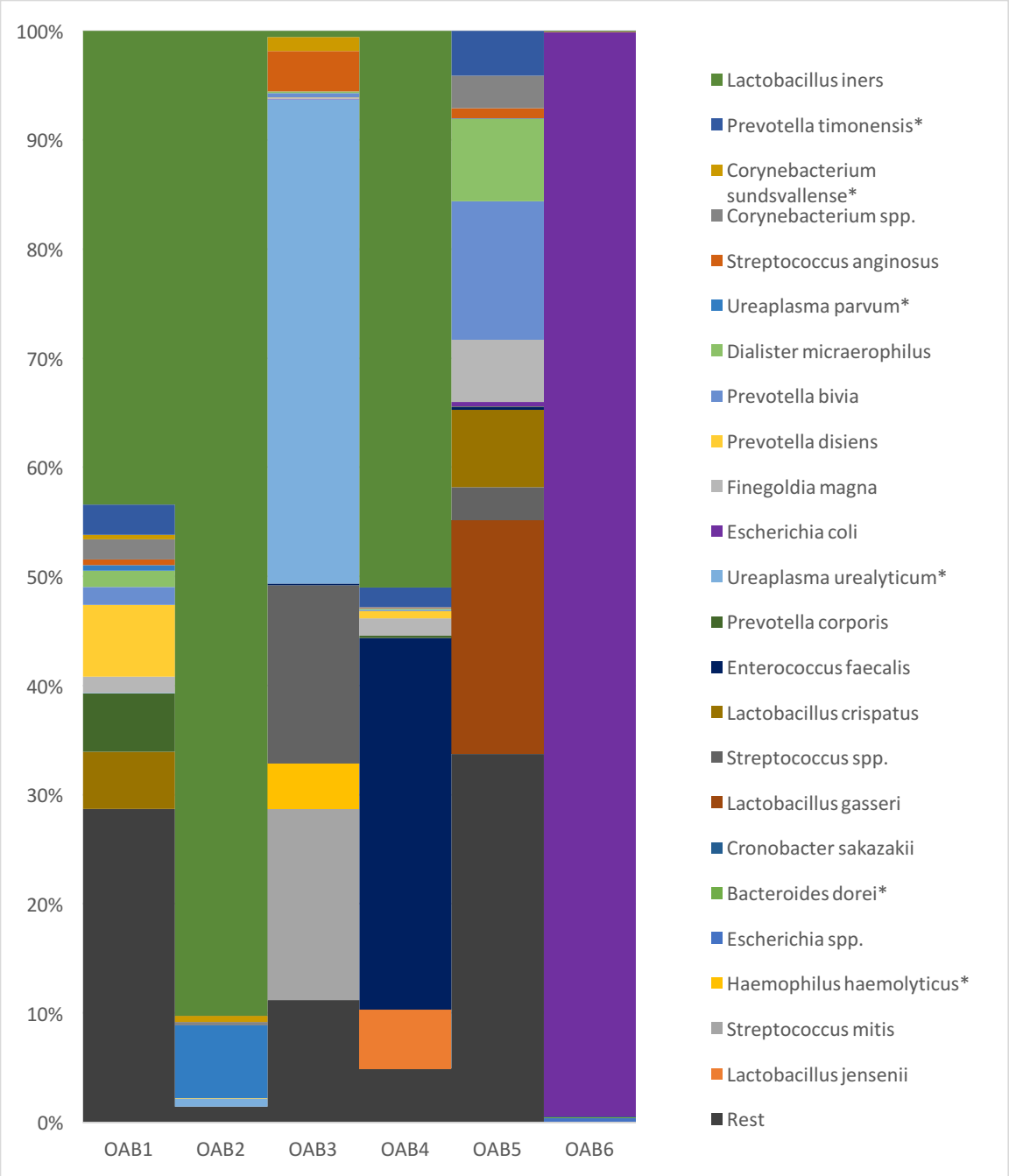
Streptococcus agalactiae	ND	ND	ND	OAB3	0.03
Aerococcus sanguinicola	ND	ND	ND	OAB5	0.09
Alloscardovia omnicolens	ND	ND	ND	OAB5	0.17

Figure 1. The most abundant species detected by culturomics (A) or community amplicon sequencing (B). The most abundant 5 species per urine sample detected by culturomics (A) or community amplicon sequencing (B) were selected and listed in order of decreasing prevalence. Asterisk denotes species detected only by culturomics and amplicon sequencing.

A)



B)


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
OAB1
OAB2
OAB3
OAB4
OAB5
OAB6
Lactobacillus iners
Prevotella timonensis*
Corynebacterium sundsvallense*
Corynebacterium spp.
Streptococcus anginosus
Ureaplasma parvum*
Dialister micraerophilus
Prevotella bivia
Prevotella disiens
Finegoldia magna
Escherichia coli
Ureaplasma urealyticum*
Prevotella corporis
Enterococcus faecalis
Lactobacillus crispatus
Streptococcus spp.
Lactobacillus gasseri
Cronobacter sakazakii
Bacteroides dorei*
Escherichia spp.
Haemophilus haemolyticus*
Streptococcus mitis
Lactobacillus jensenii
Rest

Additional material associated with the manuscript entitled “Urinary microbiome of overactive bladder: toward a better understanding through improved high-throughput community analyses” is available through the following links:

Supplementary Material:
Table S1
https://docs.google.com/spreadsheets/d/14Es55_QbkLW73GWyGTyzIH3jceVJNe9r/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Table S2
https://docs.google.com/spreadsheets/d/1MfiHyzGHHhuEuzBP7aZwHI5WR8swwKHz/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

3.3. Novel species in Female Urinary Tract Microbiome

Limosilactobacillus urinaemulieris sp. nov. and *Limosilactobacillus portuensis* sp. nov. isolated from urine of healthy women

Lactobacillus mulieris sp. nov., a new species of *Lactobacillus delbrueckii* group

The status of the species *Lactobacillus fornicalis* Dicks *et al*. 2000. Request for an opinion

Effectiveness Of MALDI-TOF Mass Spectrometry For Bacterial Identification In The Urogenital Microbiome Studies

Limosilactobacillus urinaemulieris sp. nov. and *Limosilactobacillus portuensis* sp. nov. isolated from urine of healthy women

Magdalena Ksiezarek[1*], Teresa Gonçalves Ribeiro[1*#], Joana Rocha[1], Filipa Grosso[1], Svetlana Ugarcina Perovic[1], Luisa Peixe[1].

[1]UCIBIO-REQUIMTE. Departamento de Ciências Biológicas, Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal

International Journal of Systematic and Evolutionary Microbiology 2019 Jun;71(3).

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION
Ksiezarek *et al.*, *Int. J. Syst. Evol. Microbiol.*
DOI 10.1099/ijsem.0.004726

MICROBIOLOGY SOCIETY

Limosilactobacillus urinaemulieris sp. nov. and *Limosilactobacillus portuensis* sp. nov. isolated from urine of healthy women

Magdalena Ksiezarek†, Teresa Gonçalves Ribeiro*,†, Joana Rocha, Filipa Grosso, Svetlana Ugarcina Perovic and Luisa Peixe*

Abstract

Two Gram-stain-positive strains, c9Ua_26_M^T and c11Ua_112_M^T, were isolated from voided urine samples from two healthy women. Comparative 16S rRNA gene sequences demonstrated that these novel strains were members of the genus *Limosilactobacillus*. Phylogenetic analysis based on *pheS* gene sequences and core genomes showed that each strain formed a separated branch and are closest to *Limosilactobacillus vaginalis* DSM 5837^T. The average nucleotide identity (ANI) and Genome-to-Genome Distance Calculator (GGDC) values between c9Ua_26_M^T and the closest relative DSM 5837^T were 90.7 and 42.9%, respectively. The ANI and GGDC values between c11Ua_112_M^T and the closest relative DSM 5837^T were 91.2 and 45.0%, and those among the strains were 92.9% and 51,0%, respectively. The major fatty acids were $C_{12:0}$ (40.2%), $C_{16:0}$ (26.7%) and $C_{18:1}$ $\omega 9c$ (17.7%) for strain c9Ua_26_M^T, and $C_{18:1}$ $\omega 9c$ (38.0%), $C_{16:0}$ (33.3%) and $C_{12:0}$ (17.6%) for strain c11Ua_112_M^T. The genomic DNA G+C content of strains c9Ua_26_M^T and c11Ua_112_M^T was 39.9 and 39.7 mol%, respectively. On the basis of the data presented here, strains c9Ua_26_M^T and c11Ua_112_M^T represent two novel species of the genus *Limosilactobacillus*, for which the names *Limosilactobacillus urinaemulieris* sp. nov. (c9Ua_26_M^T=CECT 30144^T=LMG 31899^T) and *Limosilactobacillus portuensis* sp. nov. (c11Ua_112_M^T=CECT 30145^T=LMG 31898^T) are proposed.

The genus *Lactobacillus* was recently reclassified into 25 genera, including the novel genus *Limosilactobacillus* proposed by Zheng *et al.* in 2020 [1]. *Limosilactobacillus* is classified in the phylum *Firmicutes*, class *Bacilli*, order *Lactobacillales*, family *Lactobacillaceae*, and comprises 17 species. The genus includes Gram-positive, rod- or coccoid-shaped, catalase-negative, heterofermentative, anaerobic or aerotolerant micro-organisms. Although members of this genus are found in fermented cereals and other fermenting plant materials, dairy products, manure and sewage, the majority are vertebrate host adapted microorganisms [1].

Limosilactobacillus comprises species generally included in the qualitative presumption of safety list (*Limosilactobacillus reuteri*, *Limosilactobacillus panis*, *Limosilactobacillus pontis*, *Limosilactobacillus mucosae*, *Limosilactobacillus fermentum*) with several strains being widely used in food and human nutrition, due to their contribution to fermented food production or their probiotic usage [2, 3]. Recently, *Limosilactobacillus* species (*Limosilactobacillus reuteri*, *Limosilactobacillus fermentum*, *Limosilactobacillus vaginalis*, *Limosilactobacillus coleohominis*) were also detected as members of the human urinary tract microbiota, although not identified in the list of the top 10 most prevalent bacteria retrieved from urinary samples from both male and female specimens [4]. In the present study, we characterized two strains of the genus *Limosilactobacillus* isolated from the urine of healthy women based on a polyphasic approach.

ISOLATION AND ECOLOGY

Two bacterial strains, c9Ua_26_M^T and c11Ua_112_M^T, were isolated from voided urine samples from two healthy women within the scope of an ongoing Female Urinary Microbiota project at the Faculty of Pharmacy, University

Author affiliations: [1]UCIBIO-REQUIMTE. Departamento de Ciências Biológicas, Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal.
***Correspondence:** Luisa Peixe, lpeixe@ffup.pt; Teresa Gonçalves Ribeiro, teresampg84@gmail.com
Keywords: 16S rRNA; *pheS*; genome; fatty acids; *Lactobacillaceae*.
Abbreviations: ANI, average nucleotide identity; GGDC, Genome-to-Genome Distance Calculator; MRS, de Man, Rogosa and Sharpe.
The Whole Genome Shotgun project of *Limosilactobacillus urinaemulieris* sp. nov. c9Ua_26_M^T has been deposited at DDBJ/ENA/GenBank under the accession JABUXR000000000 and *Limosilactobacillus portuensis* sp. nov. c11Ua_112_M^T under the accession JABUXQ000000000. The versions described in this paper are JABUXR010000000 and JABUXQ010000000, respectively.
†These authors contributed equally to this work
Two supplementary figures are available with the online version of this article.

of Porto, Portugal (2017). Urine samples were inoculated on Columbia agar with 5% sheep blood (Biogerm), and colonies were selected after 48 h at 37 °C under microaerophilic conditions (GENbox MICROAER, bioMérieux). The strains were subjected to MALDI-TOF MS (VITEK MS; bioMérieux) analysis with *in vitro* diagnostic database version 3.0, but could not be identified at species or genus level. The strains were maintained on de Man, Rogosa and Sharpe medium (MRS; Liofilchem) for short-term storage and in tryptic soy broth (Liofilchem) supplemented with 20% (v/v) glycerol at −80 °C for long-term storage.

16S rRNA AND *PHES* GENES PHYLOGENY

PCR amplification of *rrs* (16S ribosomal RNA) and *pheS* (phenylalanyl-tRNA synthetase alpha subunit) genes was performed using primers described previously [5, 6]. The sequences of the two genes were subsequently confirmed from the genomic sequences, aligned and the similarity scores were generated using MEGA software version 7.0 (www.megasoftware.net) [7]. Phylogenetic trees based on 16S rRNA or *pheS* gene sequences were reconstructed according to the neighbour-joining method [8], and genetic distances were estimated using Kimura's two-parameter model [9]. The reliability of internal branches was assessed from bootstrapping based on 1000 resamplings [10]. Phylogenetic analysis based on 16S rRNA gene sequences revealed that strains c9Ua_26_M^T and c11Ua_112_M^T clustered with the type strain of *L. vaginalis* (Fig. S1). 16S rRNA gene sequence similarities between the two novel strains c9Ua_26_M^T and c11Ua_112_M^T was 99.9%, between c9Ua_26_M^T and their closest neighbours *L. vaginalis* or *Limosilactobacillus frumenti* were 99.4 and 98.5% (respectively), and between c11Ua_112_M^T and their closest neighbours *L. vaginalis* or *L. frumenti* were 99.5 and 98.6% (respectively). The phylogenetic tree based on *pheS* gene revealed that strains c9Ua_26_M^T and c11Ua_112_M^T clustered with the type strains of *L. vaginalis* and *L. frumenti*, and were 93.3% similar between them (Fig. S2). Additionally, sequence similarity between strain c9Ua_26_M^T and its closest neighbours *L. vaginalis* or *L. frumenti* were 92.1 and 77.5% (respectively), while strain c11Ua_112_M^T was 93.9 and 77.2% similar to *L. vaginalis* and *L. frumenti*, respectively (Fig. S2). Based on these results, *L. vaginalis* was considered the closest relative species of c9Ua_26_M^T, and c11Ua_112_M^T.

GENOME FEATURES

Whole-genome sequencing of strains c9Ua_26_M^T and c11Ua_112_M^T was performed using an Illumina HiSeq system (2×150 nt). Reads were trimmed by Trimmomatic version 0.39 [11] and quality checked using FastQC 0.11.9 (www.bioinformatics.babraham.ac.uk/projects/fastqc/). *De novo* assembly was performed by SPAdes version 3.13.0 [12], and quality assessed by QUAST 5.0.2 [13]. Annotation of the draft genome was provided by the NCBI Prokaryotic Genome Annotation Pipeline [14]. Average nucleotide identity (ANI) based on BLAST+ was calculated by JSpeciesWS [15] and genome-to-genome comparison was performed with the Genome-to-Genome Distance Calculator (GGDC) following the recommended Formula 2 [16].

The genome size of isolate c9Ua_26_M^T was 1898259 bp and assembled in 66 contigs with an N50 value of 90957 bp, while the genome size of isolate c11Ua_112_M^T was 2115452 bp and assembled in 169 contigs with an N50 value of 64968 bp. The ANI value between strains c9Ua_26_M^T and c11Ua_112_M^T was 92.90%, and between strains c9Ua_26_M^T and c11Ua_112_M^T and the closest related type strain of *L. vaginalis* DSM 5837^T it was below the species cut-off level of 95% (90.7% and 91.2%, respectively). Likewise, the intergenomic distance between strains c9Ua_26_M^T and c11Ua_112_M^T revealed a GGDC value of 51%, and between strain c9Ua_26_M^T or c11Ua_112_M^T and the closest related type strain of *L. vaginalis* presented GGDC values of 42.9 and 45.0%, respectively, which is clearly below the proposed criterion for bacterial species delineation (70%) [17]. Bioinformatics analyses of these genomes predicted a G+C content of 39.9 mol% for strain c9Ua_26_M^T, 39.7 mol% for c11Ua_112_M^T and 39.1 mol% for *L. vaginalis*.

A core-genome analysis derived from the genomes of 16 *Limosilactobacillus* species that were available in NCBI Assembly database, as well as the genomes of strains c9Ua_26_M^T and c11Ua_112_M^T, was carried out. The genomes were annotated with Prokka version 1.14.6 [18], and the Roary pan genome pipeline (version 3.13.0) [19] was used to identify core genes with a 70% nucleotide identity cut-off. A phylogenetic tree was generated by FastTree version 2.1.10 [20], using the maximum-likelihood model (Generalized Time-Reversible+CAT), and 1000 resamples based on the Shimodaira–Hasegawa test (SH-like local support values). A total of 264 core genes of *Limosilactobacillus* species were included in the analysis (Fig. 1) . The phylogenomic tree based on the core genome confirmed the positioning of the two isolated strains within the genus *Limosilactobacillus*, as observed in the phylogenetic analyses based on 16S rRNA gene. These analyses demonstrated a close phylogenetic relatedness of strains c9Ua_26_M^T and c11Ua_112_M^T, which formed a distinct cluster from *L. vaginalis*, despite 99.4 and 99.5% identity observed at the 16S rRNA level.

PHYSIOLOGY AND CHEMOTAXONOMY

Cell and colony morphology were observed with cells grown on Columbia agar with 5% sheep blood agar plates (bioMérieux) at 37 °C under anaerobic conditions (GENbox ANAER, bioMérieux) for 48 h. Gram-staining was carried out using Gram-staining kit (bioMérieux). Motility, catalase activity, spore formation and gas production from glucose were determined using established procedures [21]. Growth at different oxygen conditions (aerobic, microaerophilic and anaerobic) were evaluated on MRS agar, and estimated by monitoring the optical density at 600 nm after 48 h of incubation. Growth at different temperatures (20, 25, 30, 37, 40 and 45 °C), pH (pH 3.0, 3.5, 4.0, 5.0, 6.0, 7.0, 7.5, 8.0, 8.5, 9.0, 10.0, 11.0 and 12.0) and NaCl concentrations (0.5, 1.0, 1.5,

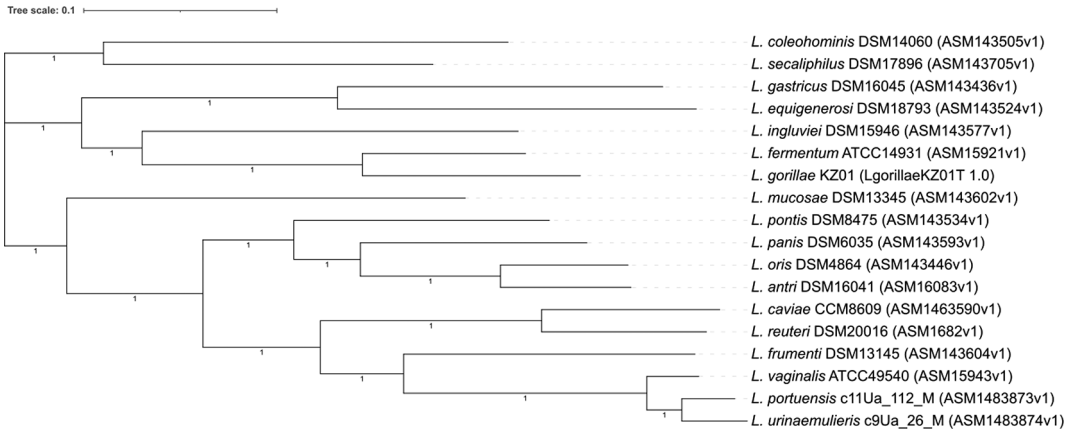


Fig. 1. Maximum-likelihood tree based on a core genome derived from 264 genes showing the phylogenetic relationships between Limosilactobacillus *urinaemulieris* c9Ua_26_M^T, Limosilactobacillus *portuensis* c11Ua_112_M^T and type strains of the genus *Limosilactobacillus*. The tree displayed is unrooted and branch labels indicate SH-like local support values. The scale bar represents 0.1 substitutions per nucleotide position. Genome sequence accession numbers are indicated in parentheses.

2.0, 3.0, 4.0 and 5.0%, w/v) were evaluated for cultures in MRS broth incubated for 7 days under anaerobic conditions, and estimated by monitoring the optical density at 600 nm after 72 h of incubation and after 7 days. Biochemical characterization was performed using standardized API 50 CH, API 20 A and API ZYM strips (bioMérieux) following the manufacturer's instructions (incubation at 37 °C). The results were recorded after 24 h, 48 h and 7 days for API 50 CH and API 20 A and after 4.5 h for API ZYM. Different phenotypic characteristics of strains c9Ua_26_M^T and c11Ua_112_M^T as compared to the type strains of closely related species are summarized in Table 1. *Limosilactobacillus* species ferment a relatively broad spectrum of carbohydrates in comparison with other heterofermentative lactobacilli [1], yet strain c11Ua_112_M^T was only able to metabolize L-arabinose, contrasting with strain c9Ua_26_M^T and *L. vaginalis*. Strain c9Ua_26_M^T was nonreactive for all of the tests in the API 50 CH and API 20 A systems, which is in agreement with phenotypic reports of some *Limosilactobacillus* strains [22]. However, some genes associated with carbohydrate utilization [e.g. sucrose phosphorylase (*scrP*), fructokinase (*fruK*) and the operon repressor *scrR* genes associated with sucrose fermentation] were detected by *in silico* analysis in the genome of strains c11Ua_112_M^T and c9Ua_26_M^T, questioning the data obtained with API 50 CH/API 20 A. Future assays using different media and incubation conditions could clarify the carbohydrate utilization profiles of the proposed *Limosilactobacillus* novel species here described.

Fatty acid profiling of strains c9Ua_26_M^T and c11Ua_112_M^T and the closest related type strain of *L. vaginalis* LMG 12891^T was performed and analysed at the BCCM/LMG Bacteria collection (Ghent, Belgium; https://bccm.belspo.be/lmg/fame-characterization), with cells grown on Columbia agar with 5% sheep blood for 48 h. Cellular fatty acid profiles revealed that strains c9Ua_26_M^T and c11Ua_112_M^T contain major fatty acids $C_{12:0}$, $C_{16:0}$ and $C_{18:1}$ ω9*c*, consistent with the analysis of representative closely relative *L. vaginalis* (Table 2). There was no particular fatty acid that could provide a taxonomically useful marker for the differentiation of these two isolates, but the difference in their amounts might be an indicator to discriminate between and among them from the most closely related valid species such as *L. vaginalis* (Table 2).

Our data support the conclusion that the two novel strains, c9Ua_26_M^T and c11Ua_112_M^T, represent independent

Table 1. Differential phenotypic characteristic of *Limosilactobacillus urinaemulieris* sp. nov., *Limosilactobacillus portuensis* sp. nov. and their closest related species

Strains: 1, c9Ua_26_M^T (this study); 2, c11Ua_112_M^T (this study), 3, *Limosilactobacillus vaginalis* DSM 5837^T [23]. +, Positive; w, weakly positive; –, negative reaction. ND, not determined.

Characteristics	1	2	3
Isolation source	Urine	Urine	Vaginal swab
Growth at:			
45 °C	–	–	+
API 50 CH:			
L-Arabinose	–	+	–
D-Galactose	–	–	+
D-Glucose	–	–	+
D-Fructose	–	–	+
D-Mannose	–	–	+
Sucrose	–	–	+
Raffinose	–	–	+
API ZYM:			
Esterase lipase (C 8)	w	–	ND

Table 2. Comparative cellular fatty acid content (%) of strains c9Ua_26_M^T and c11Ua_112_M^T and their closest related type strain *L. vaginalis* LMG 12891^T

Strains: 1, c9U_a26M^T; 2, c11Ua_112_M^T; 3, *L. vaginalis* LMG 12891^T. All data are from this study. Fatty acids present at >10% are indicated in bold. –, Not detected.

Fatty acid	1	2	3
$C_{12:0}$	**40.2**	**17.6**	**74.0**
$C_{16:0}$	**26.7**	**33.3**	**12.0**
$C_{18:0}$	5.9	7.6	5.2
$C_{18:1}$ $\omega 9c$	**17.7**	**38.0**	**8.8**
Summed feature 7*	9.6	3.6	–

*Summed features consist of one or more fatty acids that cannot be separated by the Sherlock Microbial Identification system. Summed feature 7: $C_{19:0}$ cyclo $\omega 10c/19\omega 6$, $C_{19:1}$ $\omega 6c$ and/or unknown 18.846.

two novel species within the genus *Limosilactobacillus*, for which the names *Limosilactobacillus urinaemulieris* sp. nov. (c9Ua_26_M^T=CECT 30144^T=LMG 31899^T) and *Limosilactobacillus portuensis* sp. nov. (c11Ua_112_M^T=CECT 30145^T=LMG 31898^T) are proposed.

DESCRIPTION OF *LIMOSILACTOBACILLUS URINAEMULIERIS* SP. NOV.

Limosilactobacillus urinaemulieris (u.ri.nae.mu.li'e.ris. L. fem. n. *urina* urine; L. gen. n. *mulieris* of a woman; M. L. gen. n. *urinaemulieris* pertaining to female urine from which the bacterium was first isolated).

Cells are Gram-stain-positive, non-spore-forming and motile bacilli (1.5–2.0 µm long). Colonies are circular, smooth and flat. Growth is observed under anaerobic and microaerophilic conditions. After 48 h of incubation on Columbia agar with 5% sheep blood, colonies are white to grey and 0.5 mm in diameter. Catalase and oxidase are not produced. Growth occurs between 30–40 °C (optimum, 37 °C) until the maximum of 3.0% (w/v) NaCl and in the range of pH 5.0–7.5. In the API 50 CH panel, acidification is not observed for carbohydrates included in the panel, which is congruent with the results in the API 20 A system. Acid is not produced from glycerol, erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl *β*-D-xylopyranoside, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl α-D-mannopyranoside, methyl α-D-glucopyranoside, *N*-acetylglucosamine, amygdalin, arbutin, aesculin ferric citrate, salicin, cellobiose, maltose, lactose, melibiose, sucrose, trehalose, inulin, melezitose, raffinose, starch, glycogen, xylitol, gentiobiose, turanose, D-lyxose, D-tagatose, D-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate. Additionally, cells give negative reactions to L-tryptophane, urea and gelatin. In the API ZYM test system, cells are positive for esterase (C4), leucine arylamidase, acid phosphatase, α-galactosidase and *β*-galactosidase and weakly positive for esterase lipase (C8) and naphthol-AS-BI-phosphohydrolase. Cells are negative for alkaline phosphatase, lipase (C14), valine arylamidase, cysteine arylamidase, trypsin, α-chymotrypsin, *β*-glucuronidase, α-glucosidase, *β*-glucosidase, *N*-acetyl-*β*-glucosaminidase, α-mannosidase and α-fucosidase. The major cellular fatty acids are $C_{12:0}$, $C_{16:0}$ and $C_{18:1}$ $\omega 9c$.

The type strain is c9Ua_26_M^T (=CECT 30144^T=LMG 31899^T), isolated from the urine of a healthy woman in Portugal, 2017. The DNA G+C content of the type strain is 39.9 mol%. 16S rRNA and *pheS* nucleotide sequences are deposited in the GenBank database under the accession numbers MW016377 and MW014289, respectively. The annotated genomic sequence of strain c9Ua_26_M^T was deposited in the DDBJ/ENA/GenBank databases and is available under accession number JABUXR000000000, BioSample SAMN15159665 (BioProject PRJNA548360).

DESCRIPTION OF *LIMOSILACTOBACILLUS PORTUENSIS* SP. NOV.

Limosilactobacillus portuensis (por.tu.en'sis. N.L. masc. adj. *portuensis* referring to the city of Porto, Portugal, from where the bacterium was isolated).

Cells are Gram-stain-positive, non-spore-forming and motile bacilli (2.0–2.5 µm long). Colonies are circular, smooth and flat. Growth is observed under anaerobic and microaerophilic conditions. After 48 h of incubation on Columbia agar with 5% sheep blood, colonies are white to grey and 0.5 mm in diameter. Catalase and oxidase are not produced. Growth occurs between 30–40 °C (optimum, 37 °C) until the maximum of 2.0% (w/v) NaCl, and in the range of pH 4.0–8.0. In the API 50 CH and API 20 A panels, acid is produced from L-arabinose. Acid is not produced from glycerol, erythritol, D-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl *β*-D-xylopyranoside, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl α-D-mannopyranoside, methyl α-D-glucopyranoside, *N*-acetylglucosamine, amygdalin, arbutin, aesculin ferric citrate, salicin, cellobiose, maltose, lactose, melibiose, sucrose, trehalose, inulin, melezitose, raffinose, starch, glycogen, xylitol, gentiobiose, turanose, D-lyxose, D-tagatose, D-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate. Additionally, cells give negative reactions to L-tryptophane, urea and gelatin. In the API ZYM test system, cells are positive for esterase (C4), leucine arylamidase, acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase and *β*-galactosidase. Cells are negative for alkaline phosphatase, esterase lipase (C8), lipase (C14), valine arylamidase, cysteine arylamidase, trypsin, α-chymotrypsin, *β*-glucuronidase, α-glucosidase, *β*-glucosidase, *N*-acetyl-*β*-glucosaminidase, α-mannosidase and α-fucosidase. The major cellular fatty acids are $C_{18:1}$ $\omega 9c$, $C_{16:0}$ and $C_{12:0}$.

The type strain is c11Ua_112_M^T (=CECT 30145^T=LMG 31898^T), which was isolated from the urine of a healthy woman in Portugal, 2017. The DNA G+C content of the type strain is 39.7 mol%. 16S rRNA and *pheS* nucleotide sequences are deposited in the GenBank database under accessions numbers MW016036 and MW014290, respectively. The annotated genomic sequence of strain c11Ua_112_M^T was deposited in the DDBJ/ENA/GenBank and is available under accession number JABUXQ000000000, BioSample SAMN15159667 (BioProject PRJNA548360).

Funding information
This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is financed by national funds from FCT (UIDP/04378/2020 and UIDB/04378/2020). M. K. was supported by FCT (Fundação para a Ciência e Tecnologia) (SFRH/BD/132497/2017). S. U. P. was supported as a researcher from the project NORTH-01–0145-FEDER-000024, and F. G. was supported through FCT under the Transitional Norm DL57/2016/CP1346/CT0034.

Acknowledgements
We thank Helena Ramos and Paulo Pinto (Hospital Geral de Santo António, Porto, Portugal) for their technical support with MALDI-TOF MS-based isolate identification, and bioMérieux (Portugal, Lda) for equipment and material provided. We also would like to thank our female volunteers who provided samples for the analysis.

Author contributions
M. K. and T. G. R., performed the phenotypic, biochemical and genomic characterization, and wrote and edited the original draft of the manuscript. M. K. contributed to strain isolation, genomic analysis and submission of genomes to the public database. T. G. R. performed phylogenetic characterization and designed the study. J. R. extracted genomic DNA. F. G. performed the measurement of strains and contributed to reviewing the manuscript. S. U. P. performed MALDI-TOF MS analysis and contributed to the manuscript revision. L. P. contributed to reviewing the manuscript, project design and administration, and funding.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB *et al*. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 2020;70:2782–2858.
2. Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D *et al*, EFSA Panel on Biological Hazards (BIOHAZ). Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 10: suitability of taxonomic units notified to EFSA until March 2019. *EFSA J* 2019;17:e05753.
3. Salvetti E, O'Toole PW. The genomic basis of lactobacilli as health-promoting organisms. *Microbiol Spectr* 2017;5.
4. Dubourg G, Morand A, Mekhalif F, Godefroy R, Corthier A *et al*. Deciphering the urinary microbiota repertoire by Culturomics reveals mostly anaerobic bacteria from the gut. *Front Microbiol* 2020;11:513305.
5. Héritier C, Poirel L, Nordmann P. Genetic and biochemical characterization of a chromosome-encoded carbapenem-hydrolyzing ambler class D β-lactamase from *Shewanella algae*. *Antimicrob Agents Chemother* 2004;48:1670–1675.
6. Naser SM, Thompson FL, Hoste B, Gevers D, Dawyndt P *et al*. Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiology* 2005;151:2141–2150.
7. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
8. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425.
9. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16:111–120.
10. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39:783–791.
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
13. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
14. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V *et al*. Refseq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 2018;46:D851–D860.
15. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 2016;32:929–931.
16. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:60.
17. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V *et al*. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci* 2014;9:2.
18. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
19. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
20. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5:e9490.
21. Tohno M, Kitahara M, Uegaki R, Irisawa T, Ohkuma M *et al*. *Lactobacillus hokkaidonensis* sp. nov., isolated from subarctic timothy grass (Phleum pratense L.) silage. *Int J Syst Evol Microbiol* 2013;63:2526–2531.
22. Boyd MA, Antonio MAD, Hillier SL. Comparison of API 50 CH strips to whole-chromosomal DNA probes for identification of *Lactobacillus* species. *J Clin Microbiol* 2005;43:5309–5311.
23. Embley TM, Faquir N, Bossart W, Collins MD. *Lactobacillus vaginalis* sp. nov. from the human vagina. *Int J Syst Bacteriol* 1989;39:368–370.

Lactobacillus mulieris sp. nov., a new species of *Lactobacillus delbrueckii* group

Joana Rocha[1], João Botelho[1], Magdalena Ksiezarek[1], Svetlana Ugarcina Perovic[1], Miguel Machado[2], João André Carriço[2], Lígia L. Pimentel[3], Sofia Salsinha[3], Luís M. Rodríguez-Alcalá[3], Manuela Pintado[3], Teresa G. Ribeiro[1] and Luísa Peixe[1]*

[1]UCIBIO-REQUIMTE. Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal;
[2]Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal
[3]CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Porto, Portugal

International Journal of Systematic and Evolutionary Microbiology 2020 Mar;70(3):1522-1527.

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION
Rocha *et al.*, *Int. J. Syst. Evol. Microbiol.*
DOI 10.1099/ijsem.0.003901

MICROBIOLOGY SOCIETY

Lactobacillus mulieris sp. nov., a new species of *Lactobacillus delbrueckii* group

Joana Rocha[1], João Botelho[1], Magdalena Ksiezarek[1], Svetlana Ugarcina Perovic[1], Miguel Machado[2], João André Carriço[2], Lígia L. Pimentel[3], Sofia Salsinha[3], Luís M. Rodríguez-Alcalá[3], Manuela Pintado[3], Teresa G. Ribeiro[1] and Luísa Peixe[1,*]

Abstract

One Gram-stain-positive, non-motile, non-spore-forming, catalase-negative, and coccobacilli-shaped strain, designated c10Ua161M^T, was isolated from a urine sample from a reproductive-age healthy woman. Comparative 16S rRNA gene sequence analysis indicated that strain c10Ua161M^T belonged to the genus *Lactobacillus*. Phylogenetic analysis based on *pheS* and *rpoA* gene sequences strongly supported a clade encompassing strains c10Ua161M^T and eight other strains from public databases, distinct from currently recognized species of the genus *Lactobacillus*. *In silico* Average Nucleotide Identity (ANI) and Genome-to-Genome Distance Calculator (GGDC), showed 87.9 and 34.3% identity to the closest relative *Lactobacillus jensenii*, respectively. The major fatty acids of strain c10Ua161M^T were $C_{18:1}\omega9c$ (65.0%), $C_{16:0}$ (17.8%), and summed feature 8 (10.2%; comprising $C_{18:1}\omega7c$, and/or $C_{18:1}\omega6c$). The DNA G+C content of the strains is 34.2 mol%. On the basis of data presented here, strain c10Ua161M^T represents a novel species of the genus *Lactobacillus*, for which the name *Lactobacillus mulieris* sp. nov. is proposed. The type strain is c10Ua161M^T (=CECT 9755^T=DSM 108704^T).

The genus *Lactobacillus* is a paraphyletic group of lactic acid bacteria comprising more than 200 species [http://www.bacterio.net/lactobacillus.html], belonging to the phylum *Firmicutes*. *Lactobacillus* species are anaerobic, facultative anaerobic or microaerophilic, Gram-stain-positive, non-spore-forming rods, with a wide habitat range attributed to their metabolic versatility [1, 2].

Lactobacillus comprises a high number of species generally recognized as safe and/or included in the qualitative presumption of safety list, with several strains being widely used in food and human nutrition, due to their contribution to fermented food production or their probiotic usage [3, 4]. Members of this genus contribute to the health status of different body sites, namely the gastrointestinal tract and vagina [5].

Recently, *Lactobacillus* species (e.g. *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus jensenii*, *Lactobacillus iners*), amongst other bacterial genera, were also detected as members of the human urinary tract microbiota, changing the prevailing dogma that urine from healthy individuals is sterile [6, 7]. In the present study, we characterized a novel species of the genus *Lactobacillus* isolated from the urine of a healthy female individual based on a polyphasic approach.

Strain c10Ua161M^T was isolated from a urine sample in the course of a study on the urinary microbiota of healthy reproductive-age women in Porto, Portugal (2017). Urine was inoculated on blood agar, and colonies were selected after 48 h at 37 °C under microaerophilic conditions. Strain was preliminarily identified as *Lactobacillus jensenii* (99.9%) by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), analysis using VITEK MS (bioMérieux, France) system and *In Vitro* Diagnostic database version 3.0. The strain was maintained on de Man Rogosa Sharpe medium (MRS; Liofilchem) for short-term storage and in tryptic soy broth (TSB; Liofilchem) supplemented with 20 % (v/v) glycerol at −80 °C for long-term storage.

Complete nucleotide sequences of the housekeeping genes *rrs* (16S rRNA), *pheS* (phenylalanyl-tRNA synthase alpha

Author affiliations: [1]UCIBIO-REQUIMTE. Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal; [2]Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal; [3]CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Porto, Portugal.
***Correspondence:** Luísa Peixe, lpeixe@ff.up.pt
Keywords: 16S rRNA; pheS; rpoA; genome; fatty acid.
Abbreviations: ANI, Average Nucleotide Identity; FA, fatty acid; GGDC, genome-to-genome distance calculator; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; MRS, de Man Rogosa Sharpe medium; pheS, phenylalanyl-tRNA synthase alpha subunit gene; rpoA, RNA polymerase alpha subunit gene; TSB, Tryptic Soy Broth.
One supplementary figure is available with the online version of this article.

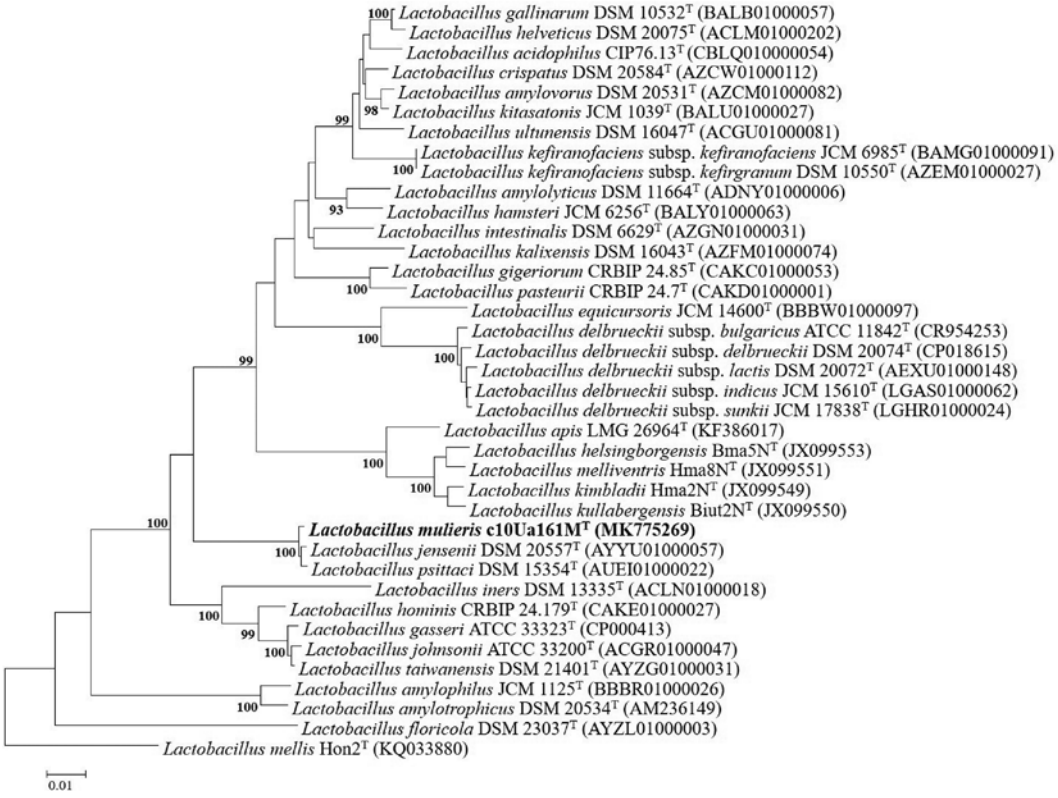


Fig. 1. Neighbour-joining tree (Kimura's two-parameter model and pairwise-deletion option) based on 16S rRNA gene sequences showing the phylogenetic relationships of Lactobacillus *mulieris* c10Ua161M^T (boldface type), and type strains of the genus Lactobacillus. Lactobacillus mellis Hon2 was used as the outgroup. Bootstrap percentages (based on 1000 replications) are shown at nodes. Only values above 90% are shown. Bar, 0.01 substitutions per nucleotide position.

subunit), and *rpoA* (RNA polymerase alpha subunit) of strain c10Ua161M^T and closely related *Lactobacillus* species were extracted, after genome annotation provided by the NCBI Prokaryotic Genome Annotation Pipeline [8], aligned and similarity scores were generated using MEGA version 7.0 (http://www.megasoftware.net/) [9]. Phylogenetic trees based on 16S rRNA gene sequences were constructed according to two different methods, neighbour-joining and maximum-likelihood [10], and genetic distances were estimated using Kimura's two-parameter model [11]. The reliability of internal branches was assessed from bootstrapping based on 1000 resamplings [12]. The 16S rRNA gene sequence variation provides limited resolution to discriminate among closely related species of the genus *Lactobacillus* [13, 14]. Indeed, phylogenetic analysis based on 16S rRNA gene sequences, obtained by the two methods, revealed that strain c10Ua161M^T clustered with type strains of *Lactobacillus jensenii* and *Lactobacillus psittaci* (Figs 1 and S1, available in the online version of this article). Phylogenetic analyses based on *pheS* or *rpoA* genes were constructed according to the neighbour-joining method [10], and genetic distances were estimated using Kimura's two-parameter model [11]. The reliability of internal branches was assessed from bootstrapping based on 1000 resamplings [12]. The phylogenetic trees based on *pheS* or *rpoA* genes showed that strain c10Ua161M^T and eight other strains with currently available genomes in public databases were grouped together and shared high similarity (100% sequence similarity), representing a well-separated lineage supported by bootstrap values of 100% (Figs 2 and 3). Furthermore, Figs 2 and 3 clearly delineate strain c10Ua161M^T and closely related strains in a separate and distinct cluster from the type strains of *Lactobacillus jensenii,* and *Lactobacillus psittaci* (91.6 and 80.3% of *pheS* sequence similarity and 97.7 and 88.7% of *rpoA* sequence similarity, respectively). These *pheS* and *rpoA* sequence divergence values observed indicate that strain c10Ua161M^T represents a novel species within the genus *Lactobacillus* [13, 15].

Whole-genome sequencing of strain c10Ua161M^T was performed by Illumina MiSeq 2×250 nt. The draft-genomes were obtained using INNUca-INNUENDO Reads Control and Assembly (https://github.com/INNUENDOCON/INNUca), which provides a pipeline to check reads quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by *de novo* assembly with SPADes [16]. Annotation of the draft-genome was provided by the NCBI Prokaryotic Genome Annotation Pipeline [8]. *In silico* genome-to-genome comparison was assessed by ANI based on BLAST+ using PyANI v0.2.7 (https://github.com/widdowquinn/pyani) [17], and Genome-to-Genome

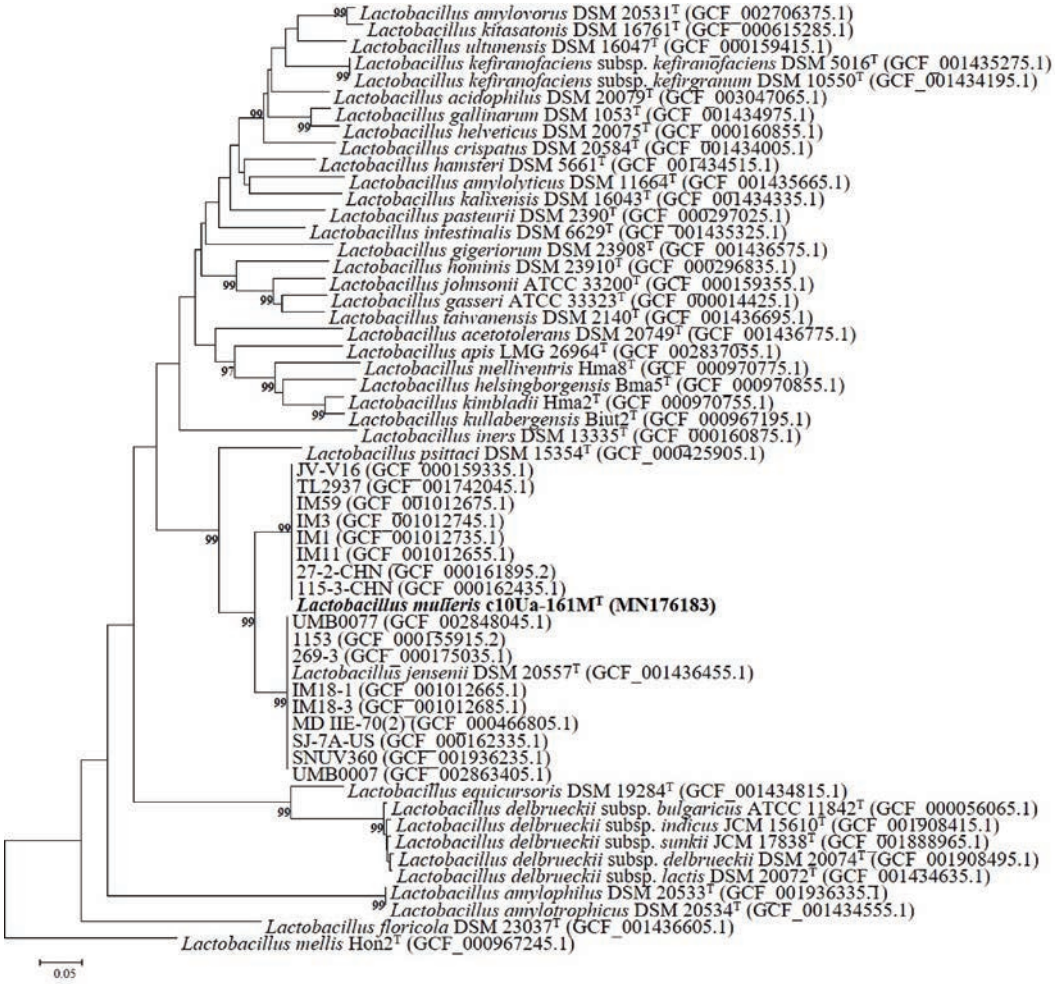


Fig. 2. Neighbour-joining tree (Kimura's two-parameter model and pairwise-deletion option) based on *pheS* gene sequences showing the phylogenetic relationships between Lactobacillus *mulieris* c10Ua161M^T (boldface type), closely related strains, and type strains of the genus Lactobacillus. Lactobacillus mellis Hon2 was used as the outgroup. Nucleotide sequences were extracted from draft/complete genomes obtained from the NCBI Assembly Database, for which the accession numbers are shown in the parenthesis. Bootstrap percentages (based on 1000 replications) are shown at nodes. Only values above 90% are shown. Bar, 0.05 substitutions per nucleotide position.

Distance Calculator (GGDC 2.1) under the recommended Formula 2 (http://ggdc.dsmz.de/distcalc2.php) [18]. The ANI value between strain c10Ua161M^T and type strain of *L. psittaci* (80.5%) and even *L. jensenii* (87.9%) were below the species cut-off level of 95% [19]. Remarkably, publicly available genomes of strains IM1, IM3, TL2937, 115–3-CHN, JV-V16, IM11, IM59, 27–2-CHN deposited as *L. jenseni* should be reclassified as *L. mulieris* based on whole-genome relatedness, since ANI values were all above 95% (99.7%, 99.5 %, 99.5 %, 99.6 %, 99.9 %, 99.6 %, 99.7 %, 99.6 %, respectively). Likewise, the intergenomic distance between strain c10Ua161M^T and the closest relative type strain of *L. jensenii* presented a GGDC value of 34.3%, which is clearly below the proposed criterion for bacterial species delineation (70 %) [20], and supports our findings of strain c10Ua161M^T as a new species.

Cell and colony morphology were observed with cells grown on MRS agar plates at 37 °C under anaerobic conditions for 72 h. Gram-staining was carried out using Gram-Staining kit (bioMérieux). Motility, catalase activity, spore formation, and gas production from glucose were determined using established procedures [21]. Growth at different temperatures [4, 10, 20, 30, 37, 40, 45 and 50 °C], pH [3.0, 3.5, 4.0, 5.0, 6.0, 7.0, 7.5, 8.0, 8.5 and 9.0] and NaCl concentrations [0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 4.5 and 5.0% (w/v)] were evaluated for cultures in MRS broth incubated for 7 days under aerobic conditions. Growth was estimated by monitoring the optical density at 600 nm. Biochemical characterization was performed using the standardized API 50 CH and API ZYM strips (incubation at 37 °C for 48 h) (bioMérieux) following the manufacturer's instructions. The novel isolates stained as Gram-positive. The coccobacilli to rod-shaped cells (0.9–3.0 µm) were non-motile.

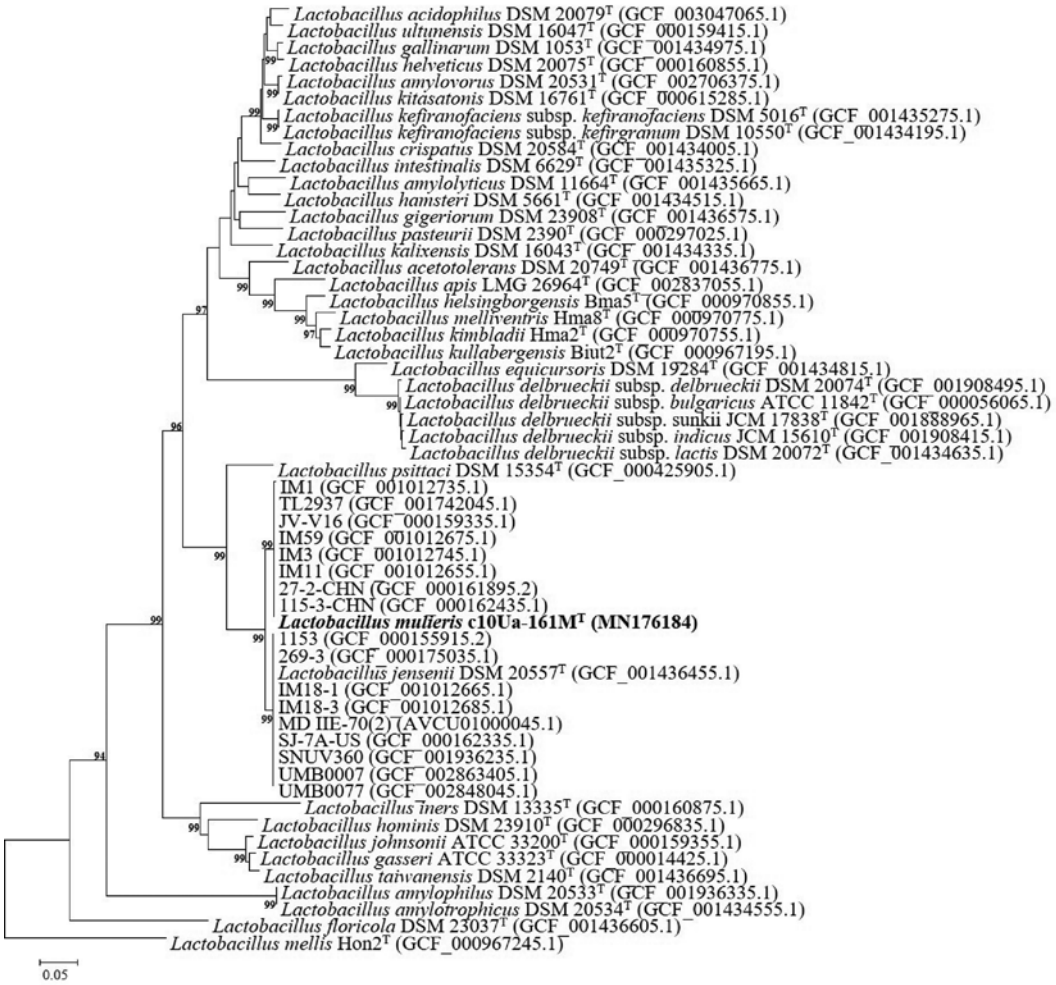


Fig. 3. Neighbour-joining tree (Kimura's two-parameter model and pairwise-deletion option) based on *rpoA* gene sequences showing the phylogenetic relationships between Lactobacillus *mulieris* c10Ua161M^T (boldface type), closely related strains and type strains of the genus Lactobacillus. Lactobacillus mellis Hon2 was used as the outgroup. Nucleotide sequences were extracted from draft/complete genomes obtained from the NCBI Assembly Database, for which the accession numbers are shown in the parenthesis. Bootstrap percentages (based on 1000 replications) are shown at nodes. Only values above 90% are shown. Bar, 0.05 substitutions per nucleotide position.

Growth occurred between 30–45 °C, with optimum temperature at 37 °C, until the maximum of 3.0% (w/v) NaCl, and in the range of pH 5.0–8.5. Different biochemical characteristics of strain c10Ua161M^T as compared to type strains of closely related species are summarized in Table 1. In particular, strain c10Ua161M^T was able to metabolize melibiose and starch, contrasting to *L. jensenii* and *L. psittaci*.

The fatty acid profile of strain c10Ua161M^T and closely related type strains of *L. jensenii* and *L. psittaci* were performed and analysed at the Leibniz Institute DSMZ (German Collection of Microorganisms and Cell Cultures GmbH) [22, 23, https://www.dsmz.de/services/microorganisms/microbiological-analysis/fatty-acid-fingerprint]. The major cellular fatty acids contained by c10Ua161M^T were $C_{18:1}\omega9c$ (65.0 %), $C_{16:0}$ (17.8 %), and summed feature 8 (10.2%; comprising $C_{18:1}\omega7c$, and/or $C_{18:1}\omega6c$) (Table 2). The fatty acid composition of c10Ua161M^T was similar to that of type strains of *L. jensenii* and *L. psittaci*, with small variations in the proportion. (Table 2).

Our data support the conclusion that strain c10Ua161M^T represents a novel species of the genus *Lactobacillus*, for which the name *Lactobacillus mulieris* sp. nov. is proposed. The type strain is c10Ua161M^T.

DESCRIPTION OF *LACTOBACILLUS MULIERIS* SP. NOV.

Lactobacillus mulieris (mu.li'.e.ris. L. gen. n. *mulieris* of a woman, from where the bacterium was isolated).

Gram-stain-positive, non-motile, non-spore-forming, catalase-negative, coccobacilli-shaped bacterium. Colonies

Table 1. Differential characteristics of strain c10Ua161M^T compared with those of closely related type strains of the genus *Lactobacillus*. Strains: 1 - c10Ua161M^T (this study); 2 L. *jensenii* DSM 20557^T [24], 3 L. *psittaci* CCUG 42378^T [25]. +, Positive; –, negative; w, weakly positive; ND, no data available

Characteristics	1	2	3
Isolation source	urine	vaginal discharge	bird lung
Morphology	coccobacilli	rod-shaped	coccobacilli to rod-shaped
Growth at/with:			
15 °C	–	ND	+
45 °C	+	–	+
G+C content (mol%)	34.2	34.3	35.65
API 50 CH:			
Ribose	–	+	–
N-Acetylglucosamine	+	–	ND
Arbutin	–	+	ND
Aesculin/ferric citrate	+	+	–
Melibiose	+	–	–
Trehalose	–	+	–
Raffinose	+	–	+
Amylum (starch)	+	w	ND
API ZYM:			
Alkaline phosphatase	–	ND	+
Leucine arylamidase	+	ND	+
Valine arylamidase	–	ND	w
Cystine arylamidase	–	ND	+
Naphthol-AS-BI-phosphohydrolase	+	ND	

Table 2. Comparative cellular fatty acid content (%) of strain c10Ua161M^T and related type strains of species of the genus *Lactobacillus*. Strains: 1, *L. mulieris* c10Ua161M^T; 2, *L. jensenii* DSM 20557^T; 3, *L. psittaci* DSM 15354^T. All data are from this study. Fatty acids present at >10% are indicated in bold. -, Not detected

Fatty acid	1	2	3
$C_{10:0}$	0.2	0.2	0.2
$C_{12:0}$	0.1	–	0.2
$C_{14:0}$	0.3	0.3	0.3
$C_{16:0}$	**17.8**	**15.6**	**15.9**
$C_{18:1}\omega9c$	**65.0**	**70.5**	**68.4**
$C_{18:0}$	2.9	3.1	2.6
$C_{19:1}$ iso	–	–	0.4
$C_{19:0}$ iso	2.5	–	2.1
Summed Features*			
3: $C_{16:1}\omega7c/_{C16:1}\omega6c/_{C15:0}$ iso 2-OH	1.0	0.9	1.0
8: $C_{18:1}\omega7c/_{C18:1}\omega6c$	**10.2**	9.4	8.9

*Summed features consist of one or more fatty acids that cannot be separated by the Sherlock Microbial Identification system.

are circular white, glossy, and 1.0–2.0 mm in diameter. They are facultative anaerobic and microaerophilic, homofermentative, and do not produce gas from glucose. Acids are produced from D-glucose, D-fructose, D-mannose, amygdalin, aesculin ferric citrate, salicin, cellobiose, maltose, melibiose, sucrose, raffinose, starch, *N*-acetyl-glucosamine but not from glycerol, erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl β-D-xylopyranoside, D-galactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl α-D-mannopyranoside, methyl α-D-glucopyranoside, lactose, trehalose, inulin, melezitose, glycogen, xylitol, turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, gluconate, 2-keto-gluconate, 5-keto-gluconate, arbutin or gentiobiose. In the API ZYM test system, cells are positive for D-glucose, D-fructose, D-mannose, *N*-acetylglucosamine, amygdalin, aesculin, maltose, cellobiose, melibiose, Sucrose, raffinose, starch, leucine arylamidase and α-glucosidase. The major fatty acids are $C_{18:1}\omega9c$, $C_{16:0}$, and summed feature 8 (comprising $C_{18:1}\omega7c$, and/or $C_{18:1}\omega6c$).

The type strain is c10Ua161M^T (=CECT 9755^T=DSM 108704^T), isolated from human urine. The DNA G+C content is 34.2 mol%.

The annotated genome data was deposited in GenBank under the accession number SDGL00000000. The 16S rRNA, *pheS* and *rpoA* nucleotide sequences data from *Lactobacillus mulieris* c10Ua161M^T isolate are available in the GenBank database under accession numbers MK775269, MN176183 and MN176184, respectively.

Funding information
This work received financial support from Applied Molecular Biosciences Unit UCIBIO and CBQF, which is financed by national funds from FCT/MCTES (UID/Multi/04378/2019, UID/Multi/50016/2013). Lígia Pimentel, Sofia Salsinha, João Botelho and Magdalena Księzarek were supported by fellowships from FCT (SFRH/BPD/119785/2016, SFRH/BD/136857/2018, SFRH/BD/104095/2014 and SFRH/BD/132497/2017, respectively).

Acknowledgements
We thank Helena Ramos and Paulo Pinto (Hospital Geral de Santo António, Porto, Portugal) for their technical support with MALDI-TOF MS-based isolate identification, and the bioMérieux (Portugal, Lda) for equipment and material provided.

Author contributions
J.R. performed the phenotypic, biochemical and genomic methodologies, phylogenetic analysis, submission of genomes to public databases, and wrote the original draft preparation. JB contributed to ***in silico*** genomic comparisons and genome annotation. M.K. contributed to strains isolation, phenotypic and biochemical characterization. SUP performed MALDI-TOF MS analysis, and contributed to data curation

and reviewing the manuscript. MM and JAC contributed to genome annotation. LLP, SS, LMR-A, and MP performed the chemotaxonomic methodologies. TGR contributed to genomic and phylogenetic analysis, writing, reviewing and editing of the manuscript. LP contributed to reviewing the manuscript, supervision, project administration and funding.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Hammes WP, Hertel C. Genus I. *Lactobacillus*. In: Vos P, Garrity G, Jones D, Krieg NR and Ludwig W (editors). *Bergey's Manual of Systematic Bacteriology*, 3, 2nd ed. New York: Springer; 2009. pp. 465–510.
2. Giraffa G, Chanishvili N, Widyastuti Y. Importance of lactobacilli in food and feed biotechnology. *Res Microbiol* 2010;161:480–487.
3. Salvetti E, O'Toole PW. The genomic basis of lactobacilli as health-promoting organisms. *Microbiol Spectr* 2017;5.
4. Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards, *et al*. Statement on the update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 9: suitability of taxonomic units notified to EFSA until September 2019. *EFSA Journal*;2019:5555.
5. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T *et al*. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* 2017;41:S27–S48.
6. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K *et al*. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *mBio* 2014;5:e01283–14.
7. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ *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;52:871–876.
8. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V *et al*. Refseq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 2018;46:D851–D860.
9. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
10. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425.
11. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16:111–120.
12. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39:783–791.
13. Naser SM, Dawyndt P, Hoste B, Gevers D, Vandemeulebroecke K *et al*. Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. *Int J Syst Evol Microbiol* 2007;57:2777–2789.
14. Salvetti E, Harris HMB, Felis GE, O'Toole PW. Comparative genomics of the genus *Lactobacillus* reveals robust phylogroups that provide the basis for reclassification. *Appl Environ Microbiol* 2018;84:e00993–18.
15. Jung MY, Lee SH, Lee M, Song JH, Chang JY *et al*. Isolated from scallion kimchi. *Int J Syst Evol Microbiol* 2017;67:4936–4942.
16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
17. 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;8:12–24.
18. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:60.
19. Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M *et al*. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* 2014;64:384–391.
20. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V *et al*. Complete genome sequence of DSM 30083, the type strain (U5/41^T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci* 2014;9:2.
21. Tohno M, Kitahara M, Uegaki R, Irisawa T, Ohkuma M *et al*. *Lactobacillus hokkaidonensis* sp. nov., isolated from subarctic timothy grass (Phleum pratense L.) silage. *Int J Syst Evol Microbiol* 2013;63:2526–2531.
22. Kuykendall LD, Roy MA, O'NEILL JJ, Devine TE. Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of *Bradyrhizobium japonicum*. *Int J Syst Bacteriol* 1988;38:358–361.
23. Miller LT. A single derivatization method for bacterial fatty acid methyl esters including hydroxy acids. *J Clin Microbiol* 1982;16:584–586.
24. Morita H, Shimazu M, Shiono H, Toh H, Nakajima F *et al*. *Lactobacillus equicursoris* sp. nov., isolated from the faeces of a thoroughbred racehorse. *Int J Syst Evol Microbiol* 2010;60:109–112.
25. Lawson PA, Wacher C, Hansson I, Falsen E, Collins MD *et al*. *Lactobacillus psittaci* sp. nov., isolated from a hyacinth macaw (*Anodorhynchus hyacinthinus*). *Int J Syst Evol Microbiol* 2001;51:967–970.

The status of the species *Lactobacillus fornicalis* Dicks *et al.* 2000. Request for an opinion

Teresa Gonçalves Ribeiro, Joana Rocha, Magdalena Ksiezarek, Svetlana Ugarcina Perovic, Filipa Grosso and Luísa Peixe*

UCIBIO-REQUIMTE. Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal;

International Journal of Systematic and Evolutionary Microbiology 2020 Jun;70(6):3953-3954.

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

REQUESTS FOR AN OPINION

Ribeiro *et al.*, *Int. J. Syst. Evol. Microbiol.* 2020;70:3953–3954

DOI 10.1099/ijsem.0.004259

MICROBIOLOGY SOCIETY

The status of the species *Lactobacillus fornicalis* Dicks *et al.* 2000. Request for an opinion

Teresa Gonçalves Ribeiro, Joana Rocha, Magdalena Ksiezarek, Svetlana Ugarcina Perovic, Filipa Grosso and Luísa Peixe*

Abstract

During a recent study on members of the genus *Lactobacillus* we realized that cultures of *Lactobacillus fornicalis* TV 1018^T (=DSM 13171^T=ATCC 700934^T) are no longer available from the online catalogue of the German Collection of Microorganisms and Cell Cultures GmbH, being displayed as *Lactobacillus plantarum* at the American Type Culture Collection. Based on data currently available, the organism deposited as ATCC 700934^T is a member of the species *Lactobacillus plantarum* subs. *plantarum*. Therefore, the type strain of *Lactobacillus fornicalis* cannot be included in any further scientific comparative study. This matter is referred to the Judicial Commission, asking for an opinion on the status of the species.

The species *Lactobacillus fornicalis* was proposed by Dicks *et al.* [1], based on the description of four strains (TV 1018^T, TV 1010, TG 1013 and TV 1045) isolated from the posterior fornix fluid of the human vagina. The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of strain TV 1018^T is Y18654, and the type strain TV 1018^T was deposited in two culture collections, namely the German Collection of Microorganisms and Cell Cultures GmbH (DSM 13171^T) and the American Type Culture Collection (ATCC 700934^T). Moreover, strain TV 1018^T is also available in the Culture Collection University of Gothenburg (CCUG 43621^T), The Biological Resource Center of Institut Pasteur (CIP 106679^T) and in the Japan Collection of Microorganisms (JCM 12512^T).

During characterization of *Lactobacillus mulieris* sp. nov., a new species of the *Lactobacilllus delbrueckii* group isolated from a urine sample [2], the type strain *Lactobacillus fornicalis* TV 1018^T was required for comparative taxonomic studies and the availability of the original type strain (TV 1018^T) was checked in different culture collections (DSM, ATCC, CIP and tJCM).

It was found that strains DSM 13171^T and TV 1010 (=DSM 13172) deposited as *L. fornicalis* are no longer listed on the DSM online catalogue, and a communication from DSM staff stated that those strains did not match the properties of *L. fornicalis*, being identified as members of the *Lactobacillus plantarum* group by automated ribotyping. The type strain of *L. fornicalis* is displayed as *Lactobacillus plantarum* in the ATCC (www.lgcstandards-atcc.org/products/all/700934.aspx?geo_country=ro) and in the CCUG (www.ccug.se/strain?id=43621) online catalogues and as *Lactobacillus* sp. in the JCM (www.jcm.riken.jp/cgi-bin/jcm/jcm_number?JCM=12512) online catalogue. Subsequent correspondence with Dr L.M. Dicks on 8 January 2020 to obtain an authentic culture was unsuccessful.

Further, the available 16S rRNA gene sequence of CIP 106679^T (=DSM 13171^T=ATCC 700934^T; https://catalogue-crbip.pasteur.fr/resultatRecherche.xhtml) is not identical to the sequence Y18654 deposited by Dicks *et al.* [1] (89.4% sequence similarity; 143 differences; comparison based on 1380 nt). In fact, a comparison of the published 16S rRNA gene sequences of type strains of members of the genus *Lactobacillus* allowed the sequence of the strain preserved as type strain of *L. fornicalis* in the CIP to be clearly allocated to *Lactobacillus plantarum* subsp. *plantarum* JCM 1149^T, with 0 nt divergence out of 1380 to the respective 16S rRNA sequence.

The authentic strain described by Dicks *et al.* [1] showing the 16S rRNA gene sequence deposited under Y18654 seems to have been lost, since it cannot be found either in public or private collections. In accordance with Rule 18g of the International Code of Nomenclature of Prokaryotes [3], we are referring to the Judicial Commission the unavailability of strain TV 1018^T and, therefore, its unsuitability as the type strain of *L. fornicalis*. Rule 18c of the International Code of Nomenclature of Prokaryotes

Author affiliations: [1]UCIBIO-REQUIMTE. Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Portugal.
***Correspondence:** Luísa Peixe, lpeixe@ff.up.pt
Keywords: *Lactobacillus fornicalis*; *Lactobacillus plantarum*; TV 1018T.
Abbreviations: ATCC, American Type Culture Collection; CCUG, Culture Collection University of Gothenburg; CIP, The Biological Resource Center of Institut Pasteur; DSM, German Collection of Microorganisms and Cell Cultures GmbH; JCM, Japan Collection of Microorganisms.

[3] states that 'if a strain on which the original description was based cannot be found, a neotype strain may be proposed'. Unfortunately, all the existing strains of *L. fornicalis* are not present in major culture collections. Since, according to Rule 30 of the International Code of Nomenclature of Prokaryotes [3], the name of a species is not validly published if a culture of a type strain is not readily available in at least one of the established culture collections, the status of the whole species *L. fornicalis* is also referred to the Judicial Commission.

On the basis of the presented evidence, we suggest the rejection of the species name, at least until a neotype strain will be proposed and confirmed as belonging to the species *L. fornicalis*.

Funding information
This work received financial support from the Applied Molecular Biosciences Unit UCIBIO, which is financed by national funds from FCT/MCTES (UIDB/Multi/04378/2019, UIDB/Multi/04378/2020). M.K. was supported by FCT (Fundação para a Ciência e Tecnologia) (SFRH/BD/132497/2017). S.U.P. was supported as a researcher from the project NORTH-01–0145-FEDER-000024, and F.G. was supported through FCT under the Transitional Norm DL57/2016/CP1346/CT0034.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Dicks LM, Silvester M, Lawson PA, Collins MD. *Lactobacillus fornicalis* sp. nov., isolated from the posterior fornix of the human vagina. *Int J Syst Evol Microbiol* 2000;50 Pt 3:1253–1258.
2. Rocha J, Botelho J, Ksiezarek M, Perovic SU, Machado M *et al*. *Lactobacillus mulieris* sp. nov., a new species of *Lactobacillus delbrueckii* group. *Int J Syst Evol Microbiol* 2020;70:1522–1527.
3. Parker CH, Tindall BJ, Garrity GM. International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* 2019;69:S1–S111.

Effectiveness Of MALDI-TOF Mass Spectrometry For Bacterial Identification In The Urogenital Microbiome Studies

Svetlana Ugarcina Perovic*[1], Magdalena Ksiezarek*[2,3], Teresa Gonçalves Ribeiro[2,3], Joana Rocha[4], Elisabete Alves Cappelli[2,3], Márcia Sousa[2,3], Filipa Grosso[2,3], Ângela Novais[2,3], Luisa Peixe[2,3]#

*SUP and MK contributed equally to this work.

[1]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

[2]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[3]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

[4]UCGenomics/GenomePT, Laboratório de Sequenciação e Genómica Funcional da Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal

Manuscript in preparation

Abstract

Background

Reliable high-throughput identification of bacterial species is critical for human microbiome studies and their interpretation, both in health and disease. MALDI-TOF mass spectrometry (MS) is a commonly employed identification method in culturomics microbiome characterization, however, in the context of female urogenital microbiome (FUM), its accuracy has not been yet evaluated. In this study, we assessed the effectiveness of MALDI-TOF MS in the identification of bacterial species isolated from FUM.

Methods

We selected 719 FUM bacterial isolates, representing abundant and/or frequent species in 46 midstream urine samples of 30 women. Isolates were identified by sequencing of 16S rRNA gene and/or appropriate species-specific marker genes (e.g. *pheS*, *cpn60*, *dltS*, *rpoB*) and by MALDI-TOF VITEK® MS (bioMérieux), with prior direct colony deposition and/or on-slide extraction.

Results

The isolates collection represented 77 species (32 genera) comprising mostly *Lactobacillus* spp. (23.5% of total isolates), *Streptococcus agalactiae* (15.7%), *Corynebacterium* spp. (15.2%), and *Gardnerella* spp. (13.3%), including putative novel species. Using MALDI-TOF MS, we were able to identify 89.4% of isolates (58 species of 23 genera). The unassigned identification of the remaining isolates (10.6%; 19 species of 9 genera) was mostly due to lack of spectra in the database from recently described species. Notably, most of isolates could be identified only with on-slide extraction, e.g., *Klebsiella pneumoniae* and *Escherichia coli*. We observed MALDI-TOF MS misidentifications for 14 bacterial species, mostly from the *Lactobacillus* (e.g., *L. mulieris*), *Gardnerella* (e.g., *G. leopoldii*) and *Corynebacterium* genus (e.g., *C. pyruviciproducens*). Most of these cases were related to lack of representative reference spectra in the database.

Conclusion

Overall, we demonstrated a high-level performance of Vitek® MS however the accurate and robust identification of certain prevalent and/or FUM abundant bacterial species (particularly *Lactobacillaceae* and *Gardnerella* members) still requires other procedures. Noteworthy, accuracy of Vitek® MS can be greatly improved by extending spectral reference databases.

Background

Besides next-generation sequencing technologies, high-throughput culture-dependent characterization of the human microbiome has been enabling significant progress in investigating the composition and diversity of human microbiome (1–3). In 2014, Hilt et al. (4) uncovered the presence of viable bacteria in the adult female bladder using both 16S rRNA gene amplicon sequencing and advanced urine culture technique. Since then, various culturomics approaches, e.g., enhanced quantitative urine culture, have been used to describe diversity of urogenital microbiome profiling by isolation and identification of both pathogenic and commensal bacteria (5–12). In this microbiome diversity characterization, a reliable high-throughput bacterial identification to the species level is critically required for determining the key members of the microbiome under healthy and unhealthy conditions.

Currently, MALDI-TOF mass spectrometry (MS) is the most frequently employed isolates' identification method in characterization of urogenital microbiome (5, 6, 10, 11, 13). The application of MALDI-TOF MS for the accurate and fast species identification has been reported for well-characterized bacteria isolated from diverse environments (14, 15), including successful bacterial typing (16). MALDI-TOF MS has been undoubtedly useful for high-throughput identification, but this technique still has a limited reference database and low discriminatory power for some relevant bacterial species. Therefore, an identification of certain bacterial species by MALDI-TOF MS may be still problematic (17–22).

Previous female urogenital microbiome (FUM) studies employing MALDI-TOF MS isolates identification already reported a high bacterial species diversity within urogenital microbiome (6, 10, 11). However, since 2019, several studies have contributed to important improvements in taxonomic classification of potentially relevant FUM members, suggesting that previously reported FUM diversity may be still underestimated. For instance, taxonomic reassessment of *Gardnerella* genus (23) demonstrated the existence of at least four distinct species instead of unique previously known *Gardnerella vaginalis;* or reorganization of *Lactobacillaceae* family (24) and descriptions of novel species [*Lactobacillus mulieris* (25), *Limosilactobacillus urinaemulieris* and *Limosilactobacillus portuensis* (26)]. Considering these taxonomic advances, further studies are needed to clarify FUM members identification by MALDI-TOF MS at species level.

Aim of this study was to assess the effectiveness of MALDI-TOF MS (VITEK MS, bioMerieux) in the identification of bacterial species from FUM that were previously reliably identified by genotypic markers. Here we evaluated a collection of over 700 bacterial isolates, representing abundant and prevalent species selected from our large FUM culture collection (46 midstream

urine samples of 30 women). To the best of our knowledge, this work presents the first study evaluating the MALDI-TOF MS performance in the (urogenital) microbiome characterization.

Methods

Bacterial isolates.

The selected FUM isolates (n=719) occurring with a high relative abundance and/or frequency were obtained from 46 first-morning midstream urine samples of 31 women (20 healthy, asymptomatic women; 7 women with overactive bladder; 3 women with urinary tract infection) in the course of Female Urogenital Microbiome project at Faculty of Pharmacy, University of Porto, Portugal (2016-2020) (**Table 1**). The urine samples were cultured onto the Columbia agar with 5% sheep blood plate (BAP, Biogerm, Portugal) and chromogenic agar plate (CAP, HiCrome UTI, HiMedia, India) supplemented as previously described (27, 28). BAPs and CAPs were incubated under aerobic and microaerophilic conditions (GENbox MICROAER, bioMérieux, France), at 37 °C, for up to 48 h. Additionally, BAPs were incubated under anaerobic conditions (GENbox ANAER, bioMérieux, France), at 37 °C for 48 h. Each morphologically distinct colony type was selected and reisolated on Columbia Agar with 5% Sheep Blood agar plates (bioMérieux, France), and stored at −80 °C in Tryptic Soy Broth (Liofilchem) supplemented with 20 % (v/v) glycerol.

Isolate preparation for MALDI-TOF MS analysis.

Bacterial isolates freshly grown on Columbia agar with 5% sheep blood plate were used for MALDI-TOF MS analysis. Firstly, we performed the direct colony technique applying the pure colony as a thin film on a spot of the target slide (disposable 48 wells target slide, bioMérieux) and after the addition of 1 µl CHCA (alpha-cyano-4-hydroxycinnamic acid) matrix (bioMérieux, France) allowing it to completely dry at room temperature. In the case of an unsuccessful identification, the on slide formic acid extraction technique was applied adding 0.5 µl of 25% formic acid (bioMérieux, France) on the bacterial spot and allowing it to completely dry before the addition of 1 µl CHCA matrix. Applied slide preparation techniques for slide preparation for bacterial isolates of different species are given in **Table 1.**

MALDI-TOF MS analysis.

Measurement was performed on a VITEK MS system (bioMérieux, France) using the manufacturer's suggested settings using automated collecting spectra. A standard (*Escherichia coli* ATCC 8739) was included to calibrate the instrument and validate the run.

Captured spectra were analyzed using VITEK MS automation control with Myla software. The library used for identification of the mass spectra was Knowledge Base database of the *in vitro* diagnostic (IVD) system version 3.2.0. Confident genus- and species-level identification were defined as a genus and species result scoring 99.9 %, respectively. Challenging isolates i.e., scoring below this threshold, not identified, multi-identified (with different identification outcomes) or misidentified isolates were reanalyzed another two times to check the repeatability of these results.

Genotypic-based isolate identification.

DNA of selected isolates was extracted mostly through alkaline lysis (SDS + NaOH) or alternatively InstaGene matrix (Bio-Rad) according to manufacturer recommendations in case of insufficient DNA amount or quality. The isolates identification was performed by sequencing of 16S rRNA gene and/or appropriate species-specific gene markers (*pheS*, *cpn60*, *dltS*, *rpoB*, *rpoD*, and*sodA*, *recN*, *malB* and Kp50233) or specific PCR and direct amplicon detection (*dltS, sodA, malB, Kp50233*). Detailed information on genetic markers used for bacterial species identification is given in **Table 1.** For the reliable identification, we performed phylogenetic analysis of the bacterial genera with large number of sequenced isolates and often questionable identification i.e., *pheS* phylogenetic analysis for *Lactobacillaceae*, *cpn60* for *Gardnerella* spp. and *rpoB* for *Corynebacterium* spp. (Perovic SU in preparation). Additionally, Sanger sequencing of sequenced isolates were released to NCBI database and their accession numbers can be found in Ksiezarek et al (13) and Perovic SU in preparation.

Results

Accuracy of identification of isolates from female urogenital microbiome by VITEK MS

By genotyping, we identified a total of 77 species (32 genera) from our selected collection (n=719 isolates), comprising mostly *Lactobacillus* spp. (23.5% of total isolates), *Streptococcus agalactiae* (15.7%), *Corynebacterium* spp. (15.2%), *Gardnerella* spp. (13.3%) including putative novel species, *Escherichia coli* (10.7%), *Enterococcus faecalis* (6.5%) and *Limosilactobacillus* spp. (5.4%). Using MALDI-TOF MS, we were able to identify 58 species (23 genera) (**Table 2**) which correspond to 89.4% (n=643 of total 719) of the isolates from the same collection. Besides the well-known putative pathogens (e.g., *Klebsiella pneumoniae, Escherichia coli, Enterococcus faecalis* and *Streptococcus agalactiae*) also *Lactobacillus*

crispatus, Lactobacillus delbrueckii and *Lactobacillus iners* were correctly identified by MALDI-TOF MS. The remaining 10.6% (76/719) of the isolates could not be identified by this methodology (mostly novel species which does not yet have a reference spectrum in the database e.g., recently described *Limosilactobacillus urinaemulieris* and *Limosilactobacillus portuensis* (26).

Of all isolates identified by MALDI-TOF MS, species-level identification was accurate for 69.7% (448/643) of isolates, while for 30.3% of isolates (195/643) species were identified incorrectly. Consideringly only isolates identified at species-level, 36.6% (164/448) of them could be correctly identified after direct colony deposition (e.g., *Citrobacter koseri, Acinetobacter pittii, Enterococcus faecalis*). Notably, we observed a higher successful rate using on-slide protein extraction, with an additional 63.4% (284/448) of isolates obtaining correct species-level identification. *Corynebacterium* spp., *Staphylococcus* spp., *Lactobacillus* spp., *Klebsiella pneumoniae* were among species that required an improved slide preparation method for effective identification (**Table 2**).

Misidentification of isolates from female urogenital microbiome by VITEK MS

In the cases of inaccurate identification at species-level (n=195 isolates) we observed two main patterns i.e., complete and partial species-level misidentification.

Complete misidentification of 148 isolates was observed mostly for *Lactobacillus* species (*Lactobacillus paragasseri* and *Lactobacillus mulieris*) and *Gardnerella* species (*Gardnerella* genomospecies 3, *Gardnerella leopoldii* and *Gardnerella swidsinskii*), followed by *Streptococcus urinalis, Alloscardovia omnicolens, Citrobacter portucalensis* and several species of *Corynebacterium* (**Table 3**). Noteworthy, we noticed that this type of inaccurate identification mostly results from the lack of reference spectra in the database for recently described species e.g., recently described *Lactobacillus mulieris* (25).

Less often we observed partial species-level misidentification (47 isolates), which results from a low discriminatory power of reference spectra. In our collection this applied to five species (*Corynebacterium aurimucosum, Lacticaseibacillus paracasei, Lacticaseibacillus rhamnosus, Lactobacillus gasseri* and *Moraxella osloensis*) that could not be identified to the single species by MALDI-TOF MS (**Table 3**).

Occasionally, a combination of both misidentification patterns could be observed with split identification of *Gardnerella vaginalis* with *Bifidobacterium* spp. by MALDI-TOF MS, which finally was neither of those but one of novel recently re-assessed *Gardnerella* species (23).

Discussion

We evaluated the performance and accuracy of MALDI-TOF MS (VITEK MS) in identifying FUM bacterial members to the species level.

The MALDI-TOF MS provided an accurate species identification of over 60% of the FUM isolates tested. This level of accuracy is slightly lower compared to previous reports focusing on clinically relevant species (>80%) (29–32). Indeed, clinical isolates mostly belong to species with reference spectra well provided in databases, while microbiome may contain a large number of environmental species with poor database representation. Additionally, as a consequence of evolving improvements in bacterial taxonomy, namely for *Gardnerella* (23) and *Lactobacillaceae* species (24, 26), these prevalent and abundant FUM members are still lacking their more discriminative reference spectra to be included in databases.

Our MALDI-TOF MS analysis demonstrated an accurate species-level identification for species often associated with urinary tract pathologies, e.g., *Klebsiella pneumoniae*, *Escherichia coli* and *Enterococcus faecalis*, as well as diverse *Staphylococcus* and *Corynebacterium* species. We also observed that protein extraction increases the overall yield of valid results.

Species frequently detected in the urogenital microbiome both under healthy and unhealthy conditions, such as *Lactobacillus* and *Gardnerella* species were often misidentified and/or unidentified. Although previous studies (33, 34) suggested that the MALDI-TOF MS is able to reliably distinguish between the *Lactobacillus* species, our findings including recent taxonomic improvements indicated that only some *Lactobacillus* species may be precisely identified by this method (*Lactobacillus crispatus*, *Lactobacillus delbrueckii* and *Lactobacillus iners*). For instance, *Lactobacillus jensenii* can be identified correctly, but it may be often misidentified with recently described *Lactobacillus mulieris* (25), that does not yet have a reference spectrum in the database. We also identified lack of spectra for novel *Limosilactobacillus* genus, often present in FUM [including *Limosilactobacillus urinaemulieris* and *Limosilactobacillus portuensis* (26)]. In our study, *Gardnerella* was another challenging genus for which MALDI-TOF MS seems insufficient method to accurately identify species. Recently published report on FUM *Gardnerella* spp. (35) reported similar observations to ours, although typing was performed by a different MALDI-TOF MS system (Bruker Daltonik). As a prevalent FUM member (5, 6, 10, 11, 13), reliable taxonomic classification of species belonging to *Gardnerella* may be essential for further understanding of FUM role.

Overall, we observed that the reasons for species misidentification include lack of their reference spectra in database (database mostly clinically relevant species oriented) or insufficient discrimination at the species level for genera that are suffering great taxonomic rearrangements, some of which are relevant for FUM.

The main strength of this study was testing a large and diverse FUM isolates collection from both healthy and unhealthy women. Also, a strict MALDI-TOF MS protocol with repeatability testing i.e., challenging isolates were reanalyzed in duplicate to discard technical problems in sample preparation, and to confirm repeatability. The main limitation of this study is use of only one commercially available MALDI-TOF MS systems for bacterial identification. We used Vitek MS (bioMérieux), one of two of the most widely used MALDI-TOF MS systems, while the second one - Bruker Biotyper (Bruker Daltonics) - is also commonly used in FUM studies (4, 5). However, they have been shown to have similar performances for bacterial identification of primarily clinically relevant species (36–40).
Further work should be performed in order to standardize the MALDI-TOF MS protocols for identification of isolates obtained by culturomics in microbiome studies, and especially to update reference databases with newly discovered species.

Conclusions

The MALDI-TOF MS is a valuable but not always sufficient method for the identification of FUM members to the species level. This method must be used in combination with genotypic techniques, with high power of discrimination, to achieve unambiguous results for identification of important FUM members, particularly *Lactobacillaceae* and *Gardnerella* species. Future culturomics research coupled with accurate and reliable high-throughput isolate identification is undoubtedly a promising approach to characterization of urogenital microbiome.

Declarations

Ethics approval and consent to participate

No approval was required for this study.

Availability of data and materials

All data included in this work is available within the manuscript or in complementary publications i.e., Ksiezarek et al 2021, Perovic SU in preparation.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal, and by the UCIBIO – Applied Molecular Biosciences Unit, Laboratory of Microbiology, Department of Biological Sciences, REQUIMTE, Faculty of Pharmacy, University of Porto, 4050-313 Porto, which are financed by national funds from FCT (Fundação para a Ciência e a Tecnologia, I.P.) (UIDB/MULTI/04378/2020 and UIDP/04378/2020). SUP was supported as researcher from the project NORTH-01-0145-FEDER-000024, MK and MS by an FCT PhD grant (SFRH/BD/132497/2017 and SFRH/BD/05038/2020, respectively), and JR by a postdoctoral fellowship from ICETA (UID/MULTI/04378/2013). FG and ÂN is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0034 and DL57/2016/CP1346/CT0032, respectively].

Authors' contributions

SUP, MK, FG and LP designed the study and supervised participant recruitment. SUP, MK and JR processed the samples and collected the data. SUP, MK, JR, MS, EC and TGR performed the isolates' identification. TGR supervised MS and EC. SUP conducted the community data analysis and visualization. SUP and MK interpreted the data and wrote the manuscript. TGR, FG and LP revised the article. All authors read and approved the manuscript.

Acknowledgements

We thank Helena Ramos and Paulo Pinto (Hospital Geral de Santo António, Porto, Portugal) for their technical support with MALDI-TOF MS-based microbial identification. The bioMérieux (Portugal, Lda.) provided equipment and material for MALDI-TOF MS analysis, and had no role in the study design, data collection and analysis, decision to publish, or writing of the manuscript.

REFERENCES

1. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. 7604. Nature 533:543–546.
2. Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerce J, Mailhe M, Ricaboni D, Bilen M, Dangui Nieko NPM, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahão J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier P-E, Levasseur A, Raoult D. 2016. Culture of previously uncultured members of the human gut microbiota by culturomics. 12. Nature Microbiology 1:1–8.
3. Neugent ML, Hulyalkar NV, Nguyen VH, Zimmern PE, Nisco NJD. 2020. Advances in Understanding the Human Urinary Microbiome and Its Potential Role in Urinary Tract Infection. mBio 11.
4. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. Journal of Clinical Microbiology 52:871–876.
5. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. mBio 5:e01283-01214.
6. Coorevits L, Heytens S, Boelens J, Claeys G. 2017. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. Eur J Clin Microbiol Infect Dis 36:635–639.
7. Curtiss N, Balachandran A, Krska L, Peppiatt-Wildman C, Wildman S, Duckett J. 2017. A case controlled study examining the bladder microbiome in women with Overactive Bladder (OAB) and healthy controls. Eur J Obstet Gynecol Reprod Biol 214:31–35.
8. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, Hilt EE, Price TK, Wolfe AJ, Lawley TD. 2018. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. 1. Nature Communications 9:1557.
9. Gill K, Kang R, Sathiananthamoorthy S, Khasriya R, Malone-Lee J. 2018. A blinded observational cohort study of the microbiological ecology associated with pyuria and overactive bladder symptoms. Int Urogynecol J 29:1493–1500.

10. Sathiananthamoorthy S, Malone-Lee J, Gill K, Tymon A, Nguyen TK, Gurung S, Collins L, Kupelian AS, Swamy S, Khasriya R, Spratt DA, Rohn JL. 2019. Reassessment of Routine Midstream Culture in Diagnosis of Urinary Tract Infection. J Clin Microbiol 57.
11. Price TK, Hilt EE, Thomas-White K, Mueller ER, Wolfe AJ, Brubaker L. 2020. The urobiome of continent adult women: a cross-sectional study. BJOG 127:193–201.
12. Putonti C, Shapiro JW, Ene A, Tsibere O, Wolfe AJ. 2020. Comparative Genomic Study of *Lactobacillus jensenii* and the Newly Defined *Lactobacillus mulieris* Species Identifies Species-Specific Functionality. mSphere 5.
13. Ksiezarek M, Ugarcina-Perovic S, Rocha J, Grosso F, Peixe L. 2021. Long-term stability of the urogenital microbiota of asymptomatic European women. BMC Microbiol 21:64.
14. Florio W, Baldeschi L, Rizzato C, Tavanti A, Ghelardi E, Lupetti A. 2020. Detection of Antibiotic-Resistance by MALDI-TOF Mass Spectrometry: An Expanding Area. Frontiers in Cellular and Infection Microbiology 10:616.
15. Hou T-Y, Chiang-Ni C, Teng S-H. 2019. Current status of MALDI-TOF mass spectrometry in clinical microbiology. Journal of Food and Drug Analysis 27:404–414.
16. Novais Â, Sousa C, de Dios Caballero J, Fernandez-Olmos A, Lopes J, Ramos H, Coque TM, Cantón R, Peixe L. 2014. MALDI-TOF mass spectrometry as a tool for the discrimination of high-risk *Escherichia coli* clones from phylogenetic groups B2 (ST131) and D (ST69, ST405, ST393). Eur J Clin Microbiol Infect Dis 33:1391–1399.
17. Marín M, Cercenado E, Sánchez-Carrillo C, Ruiz A, Gómez González Á, Rodríguez-Sánchez B, Bouza E. 2017. Accurate Differentiation of *Streptococcus pneumoniae* from other Species within the *Streptococcus mitis* Group by Peak Analysis Using MALDI-TOF MS. Frontiers in Microbiology 8:698.
18. Slotved H-C, Facklam RR, Fuursted K. 2017. Assessment of a novel bile solubility test and MALDI-TOF for the differentiation of *Streptococcus pneumoniae* from other mitis group streptococci. Sci Rep 7:7167.
19. Elbehiry A, Marzouk E, Hamada M, Al-Dubaib M, Alyamani E, Moussa IM, AlRowaidhan A, Hemeg HA. 2017. Application of MALDI-TOF MS fingerprinting as a quick tool for identification and clustering of foodborne pathogens isolated from food products. New Microbiol 40:269–278.
20. Homem de Mello de Souza HAP, Dalla-Costa LM, Vicenzi FJ, Camargo de Souza D, Riedi CA, Filho NAR, Pilonetto M. 2014. MALDI-TOF: a useful tool for laboratory identification of uncommon glucose non-fermenting Gram-negative bacteria associated with cystic fibrosis. J Med Microbiol 63:1148–1153.

21. Rodríguez-Sánchez B, Alcalá L, Marín M, Ruiz A, Alonso E, Bouza E. 2016. Evaluation of MALDI-TOF MS (Matrix-Assisted Laser Desorption-Ionization Time-of-Flight Mass Spectrometry) for routine identification of anaerobic bacteria. Anaerobe 42:101–107.
22. Veloo ACM, Jean-Pierre H, Justesen US, Morris T, Urban E, Wybo I, Kostrzewa M, Friedrich AW, On Behalf Of The Enria Workgroup null. 2018. An overview of the data obtained during the validation of an optimized MALDI-TOF MS Biotyper database for the identification of anaerobic bacteria. Data Brief 18:1484–1496.
23. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P. 2019. Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. Int J Syst Evol Microbiol 69:679–687.
24. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. Int J Syst Evol Microbiol 2782–2858.
25. Rocha J, Botelho J, Ksiezarek M, Perovic SU, Machado M, Carriço JA, Pimentel LL, Salsinha S, Rodríguez-Alcalá LM, Pintado M, Ribeiro TG, Peixe L. 2020. *Lactobacillus mulieris* sp. nov., a new species of *Lactobacillus delbrueckii* group. Int J Syst Evol Microbiol https://doi.org/10.1099/ijsem.0.003901.
26. Ksiezarek M, Ribeiro TG, Rocha J, Grosso F, Perovic SU, Peixe L. 2021. *Limosilactobacillus urinaemulieris* sp. nov. and *Limosilactobacillus portuensis* sp. nov. isolated from urine of healthy women. International Journal of Systematic and Evolutionary Microbiology 71:004726.
27. Alves P, Castro J, Sousa C, Cereija TB, Cerca N. 2014. *Gardnerella vaginalis* outcompetes 29 other bacterial species isolated from patients with bacterial vaginosis, using in an in vitro biofilm formation model. J Infect Dis 210:593–596.
28. Man JCD, Rogosa M, Sharpe ME. 1960. A Medium for the Cultivation of Lactobacilli. Journal of Applied Bacteriology 23:130–135.
29. Bizzini A, Greub G. 2010. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. Clin Microbiol Infect 16:1614–1619.

30. Rychert J, Burnham C-AD, Bythrow M, Garner OB, Ginocchio CC, Jennemann R, Lewinski MA, Manji R, Mochon AB, Procop GW, Richter SS, Sercia L, Westblade LF, Ferraro MJ, Branda JA. 2013. Multicenter evaluation of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system for identification of Gram-positive aerobic bacteria. J Clin Microbiol 51:2225–2231.
31. Ge M-C, Kuo A-J, Liu K-L, Wen Y-H, Chia J-H, Chang P-Y, Lee M-H, Wu T-L, Chang S-C, Lu J-J. 2017. Routine identification of microorganisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Success rate, economic analysis, and clinical outcome. J Microbiol Immunol Infect 50:662–668.
32. Kassim A, Pflüger V, Premji Z, Daubenberger C, Revathi G. 2017. Comparison of biomarker based Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) and conventional methods in the identification of clinically relevant bacteria and yeast. BMC Microbiol 17:128.
33. Anderson DMG, Ablonczy Z, Koutalos Y, Spraggins J, Crouch RK, Caprioli RM, Schey KL. 2014. High resolution MALDI imaging mass spectrometry of retinal tissue lipids. J Am Soc Mass Spectrom 25:1394–1403.
34. Schulthess B, Bloemberg GV, Zbinden R, Böttger EC, Hombach M. 2014. Evaluation of the Bruker MALDI Biotyper for identification of Gram-positive rods: development of a diagnostic algorithm for the clinical laboratory. J Clin Microbiol 52:1089–1097.
35. Putonti C, Thomas-White K, Crum E, Hilt EE, Price TK, Wolfe AJ. 2021. Genome Investigation of Urinary *Gardnerella* Strains and Their Relationship to Isolates of the Vaginal Microbiota. mSphere 6.
36. Florio W, Tavanti A, Ghelardi E, Lupetti A. 2018. MALDI-TOF MS Applications to the Detection of Antifungal Resistance: State of the Art and Future Perspectives. Front Microbiol 9:2577.
37. Pandey A, Jain R, Sharma A, Dhakar K, Kaira GS, Rahi P, Dhyani A, Pandey N, Adhikari P, Shouche YS. 2019. 16S rRNA gene sequencing and MALDI-TOF mass spectrometry based comparative assessment and bioprospection of psychrotolerant bacteria isolated from high altitudes under mountain ecosystem. SN Appl Sci 1:278.
38. Lévesque S, Dufresne PJ, Soualhine H, Domingo M-C, Bekal S, Lefebvre B, Tremblay C. 2015. A Side by Side Comparison of Bruker Biotyper and VITEK MS: Utility of MALDI-TOF MS Technology for Microorganism Identification in a Public Health Reference Laboratory. PLoS One 10:e0144878.

39. de Alegría Puig CR, Torres MF, Marfil-Pérez E, Ferández MIR, Del Río MC, Balbín JA, Martínez-Martínez L. 2019. Comparison between Vitek MS, Bruker Biotyper, Vitek2, and API20E for differentiation of species of the genus Raoultella. Eur J Clin Microbiol Infect Dis 38:467–470.
40. Cao Y, Wang L, Ma P, Fan W, Gu B, Ju S. 2018. Accuracy of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Identification of Mycobacteria: a systematic review and meta-analysis. Sci Rep 8:4131.

Table 1. Isolates information, MALDI-TOF MS slide preparation technique and genetic markers used for bacterial species identification.

Genus/Species	Total number of isolates	Number of isolates from healthy women	women with OAB	women with UTI	pregnant women	MALDI_TOF MS slide preparation technique	Sequencing/ Amplicon detection	Target gene	PCR primers' sequences (5'–3')	Reference
Acinetobacter	**1**	**1**				Direct colony	Sequencing	*rpoB*	TAYCGYAAAGAYT TGAAAGAAG CGBGCRTGCATYT TGTCRT	La Scola et al., 2006
Acinetobacter pittii	1	1								
Citrobacter	**2**	**1**	**1**			Direct colony	Sequencing	*recN*	ATTGCCATTGATG CTCTCGG ANCGAGTCGGCCT GATCGT	Ribeiro et al., 2015
Citrobacter koseri	1	1								
Citrobacter portucalensis	1		1							
Corynebacterium	**109**	**87**	**22**			Direct colony On-slide extraction	Sequencing	*rpoB*	CNTCBCACTAYGG NCGNATG GAVCGNGCGTGRA TCTTYTC	Perovic et al. in preparation
Corynebacterium afermentans	3	3						*rpoB*	CGWATGAACATYG GBCAGGT TCCATYTCRCCRA ARCGCTG	Khamis et al., 2005
Corynebacterium amycolatum	6	6								
Corynebacterium aurimucosum	23	17	6							
Corynebacterium coyleae	7	4	3							
Corynebacterium fournieri	1	1								
Corynebacterium freneyi	2		2							
Corynebacterium glucuronolyticum	1	1								
Corynebacterium imitans	2		2							
Corynebacterium jeikeium	5	4	1							
Corynebacterium kroppenstedtii	1	1								

Corynebacterium minutissimum	1		1							
Corynebacterium pyruviciproducens	6	6								
Corynebacterium riegelii	1	1								
Corynebacterium simulans	8	7	1							
Corynebacterium striatum	3	3								
Corynebacterium tuberculostearicum	20	17	3							
Corynebacterium tuscaniense	2	2								
Corynebacterium urealyticum	1		1							
Corynebacterium sp. nov. 1	1	1								
Corynebacterium sp. nov. 2	3	3								
Corynebacterium sp. nov. 3	4	4								
Corynebacterium sp. nov. 4	1	1								
Corynebacterium sp. nov. 5	5	5								
Corynebacterium sp. nov. 6	2		2							
Enterococcus faecalis	**47**	**35**	**7**	**4**	**1**	Direct colony	Amplicon detection	*sodA*	ACTTATGTGACTA ACTTAACC TAATGGTGAATCT TGGTTTGG	Jackson et al., 2004
Escherichia coli	77	**14**	**58**	**5**		Direct colony On-slide extraction	Amplicon detection	*malB*	GACCTCGGTTTAG TTCACAGA CACACGCTGACGC TGACCA	Wang et al., 1996
Gardnerella	**96**	**57**	**39**			Direct colony On-slide extraction	Sequencing	*cpn60*	GAIIIIGCIGGIGAY GGIACIACIAC YKIYKITCICCRAAI	Hill et al., 2006
Gardnerella genomospecies 3	44	24	20						CCIGGIGCYTT	
Gardnerella leopoldii	6	2	4						GAIIIIGCIGGYGAC GGYACSACSAC CGRCGRTCRCCGA	
Gardnerella swidsinskii	16	16							AGCCSGGIGCCTT	
Gardnerella vaginalis	30	15	15							
Klebsiella pneumoniae sensu stricto	**5**		**4**	**1**		Direct colony On-slide extraction	Amplicon detection	Kp50233	GCTCTGGGAGATA GACCGCA	Bialek-Davenet et al., 2014

Lactobacillus	**169**	**144**	**12**	**11**	**2**	Direct colony On-slide extraction	Sequencing	*pheS*	CAYCCNGCHCGYGAYATGC GGRTGRACCATVCCNGCHCC	Naser et al., 2005
Lactobacillus crispatus	41	32	7	1	1					
Lactobacillus delbrueckii	2	2								
Lactobacillus gasseri	25	25								
Lactobacillus iners	19	11	1	7						
Lactobacillus jensenii	19	17		2						
Lactobacillus mulieris	27	27								
Lactobacillus paragasseri	36	30	4	1	1					
Limosilactobacillus	**39**	**39**								
Limosilactobacillus mucosae	1	1								
Limosilactobacillus portuensis	22	22								
Limosilactobacillus urinaemulieris	16	16								
Lacticaseibacillus	**9**	**3**	**6**							
Lacticaseibacillus paracasei	6		6							
Lacticaseibacillus rhamnosus	3	3								
Pseudomonas	**1**		**1**			Direct colony	Sequencing	*rpoD*	ATYGAAATCGCCAARCG CGGTTGATKTCCTTGA	Mulet et al. 2009
Pseudomonas luteola	1		1							
Staphylococcus	**5**	**5**				Direct colony On-slide extraction	Sequencing	*rpoB*	CAATTCATGGACCAAGC GCIACITGITCCATACCTGT	Mellmann et al., 2006
Staphylococcus condimenti	1	1								
Staphylococcus haemolyticus	1	1								
Staphylococcus lugdunensis	2	2								
Staphylococcus petrasii	1	1								
Streptococcus agalactiae	**113**	**112**		**1**		Direct colony	Amplicon detection	*dltS*	AGGAATACCAGGCGATGAACCGAT TGCTCTAATTCTCCCCTTATGGC	Poyart et al., 2007
Various	**46**	**36**	**3**	**7**		Direct colony On-slide extraction	Sequencing	*rrs*	AGAGTTTGATCHTGGYTYAGA ACGGYTACCTTGTTACGACTTC	Heritier et al., 2014
Actinomyces neuii	5	3		2						
Alloscardovia omnicolens	2	2								

Atopobium minutum	1	1								
Bacillus circulans	1			1						
Collinsella aerofaciens	1	1								
Dialister micraerophilus	2	2								
Dialister propionicifaciens	1	1								
Facklamia hominis	1	1								
Finegoldia magna	1			1						
Fusobacterium nucleatum	1	1								
Helcobacillus massiliensis	1		1							
Kocuria rhizophila	1	1								
Moraxella osloensis	5	3		2						
Oligella urethralis	1	1								
Peptoniphilus coxii	1	1								
Peptoniphilus duerdenii	1	1								
Prevotella brunnea	1		1							
Prevotella corporis	1	1								
Propionimicrobium lymphophilum	1	1								
Rothia dentocariosa	1	1								
Staphylococcus epidermidis	1			1						
Staphylococcus saprophyticus	1	1								
Stenotrophomonas maltophilia	1	1								
Streptococcus urinalis	12	12								
Weeksella virosa	1		1							

Abbreviations: *cpn60*, 60 kDa chaperonin protein; *dltS*, histidine kinase specific to Group B *Streptococcus*; *malB*, maltose operon protein; OAB, overactive bladder; *pheS*, phenylalanyl-tRNA synthetase alpha subunit; *recN*, DNA repair protein; *rrs*, 16S rRNA; *rpoB*, RNA polymerase beta subunit; *rpoD*, RNA polymerase sigma factor; *sodA*, superoxide dismutase; UTI, urinary tract infection.

Table 2. Accuracy of bacterial species identification using MALDI-TOF MS.

Species	Total number of isolates	Number of isolates			
		with correct species identification	with correct genus identification (species misidentification)	with species and genus misidentification	without identification
Acinetobacter pittii	1	1			
Actinomyces neuii	5	5			
*Alloscardovia omnicolens**	2			1	1
*Atopobium minutum**	1				1
Bacillus circulans	1	1			
Citrobacter koseri	1	1			
*Citrobacter portucalensis**	1		1		
Collinsella aerofaciens	1	1			
Corynebacterium afermentans	3	2			1
Corynebacterium amycolatum	6	6			
Corynebacterium aurimucosum	23	13	8		2
Corynebacterium coyleae	7	6	1		
Corynebacterium fournieri	1				1
Corynebacterium freneyi	2	2			
Corynebacterium glucuronolyticum	1	1			
Corynebacterium imitans	2	2			
Corynebacterium jeikeium	5	5			
Corynebacterium kroppenstedtii	1	1			
*Corynebacterium minutissimum**	1				1
*Corynebacterium pyruviciproducens**	6	1	2		3
Corynebacterium riegelii	1	1			
Corynebacterium simulans	8	7	1		
Corynebacterium striatum	3	3			
Corynebacterium tuberculostearicum	20	20			
*Corynebacterium tuscaniense**	2				2
Corynebacterium urealyticum	1	1			
*Corynebacterium sp. nov. 1**	1		1		
*Corynebacterium sp. nov. 2**	3				3
*Corynebacterium sp. nov. 3**	4		4		
*Corynebacterium sp. nov. 4**	1		1		
*Corynebacterium sp. nov. 5**	5				5
*Corynebacterium sp. nov. 6**	2				2
Dialister micraerophilus	2	2			
*Dialister propionicifaciens**	1				1
Enterococcus faecalis	47	47			
Escherichia coli	77	77			

Facklamia hominis	1	1			
Finegoldia magna	1	1			
Fusobacterium nucleatum	1	1			
Gardnerella genomospecies 3*	44		44		
*Gardnerella leopoldii**	6		2	4	
*Gardnerella swidsinskii**	16		10	6	
Gardnerella vaginalis	30	30			
*Helcobacillus massiliensis**	1				1
Klebsiella pneumoniae	5	5			
Kocuria rhizophila	1	1			
Lacticaseibacillus paracasei	6		6		
Lacticaseibacillus rhamnosus	3		3		
Lactobacillus crispatus	41	41			
Lactobacillus delbrueckii	2	2			
Lactobacillus gasseri	25		25		
Lactobacillus iners	19	19			
Lactobacillus jensenii	19	19			
*Lactobacillus mulieris**	27		25		2
*Lactobacillus paragasseri**	36		36		
*Limosilactobacillus mucosae**	1				1
*Limosilactobacillus portuensis**	22				22
*Limosilactobacillus urinaemulieris**	16				16
Moraxella osloensis	5			5	
Oligella urethralis	1	1			
*Peptoniphilus coxii**	1				1
*Peptoniphilus duerdenii**	1				1
*Prevotella brunnea**	1				1
*Prevotella corporis**	1				1
*Propionimicrobium lymphophilum**	1				1
Pseudomonas luteola	1	1			
Rothia dentocariosa	1	1			
*Staphylococcus condimenti**	1				1
Staphylococcus epidermidis	1	1			
Staphylococcus haemolyticus	1	1			
Staphylococcus lugdunensis	2	1			1
Staphylococcus petrasii	1				1
Staphylococcus saprophyticus	1	1			
Stenotrophomonas maltophilia	1	1			
Streptococcus agalactiae	113	113			
*Streptococcus urinalis**	12		9		3
Weeksella virosa	1	1			
Total number of isolates	**719**	**448**	**179**	**16**	**76**
*Species lacking reference spectra	**217**				

Table 3. Summary of misidentifications by MALDI-TOF MS.

MALDI-TOF MS identification (% of confidence)	Number of isolates	Species
Citrobacter freundii (99.9)	1	*Citrobacter portucalensis**
Corynebacterium tuberculostearicum (50.0), aurimucosum (50.0)	8	*Corynebacteriym aurimucosum*
	4	*Corynebacterium sp. nov. 3**
Corynebacterium freneyi (99.9)	1	*Corynebacterium coyleae*
Corynebacterium tuberculostearicum (99.9)	1	*Corynebacterium minutissimum**
Corynebacterium glucuronolyticum (99.9)	2	*Corynebacterium pyruviciproducens**
Corynebacterium simulans (49.9), striatum (50.0)	1	*Corynebacterium simulans*
Corynebacterium kroppenstedtii (99.9)	1	*Corynebacterium sp. nov. 1**
Corynebacterium jeikeium (99.9)	1	*Corynebacterium sp. nov. 4**
Gardnerella vaginalis (50.0), Bifidobacterium spp. (50.0)	2	*Gardnerella leopoldii**
	6	*Gardnerella swidsinskii**
	1	*Alloscardovia omnicolens**
Gardnerella vaginalis (99.9)	4	*Gardnerella leopoldii**
	44	*Gardnerella* genomospecies 3*
	10	*Gardnerella swidsinskii**
Lacticaseibacillus paracasei (33.3), rhamnosus (33.3), casei (33.3)**	6	*Lacticaseibacillus paracasei*
	3	*Lacticaseibacillus rhamnosus*
Lactobacillus acidophilus/gasseri (99.9)**	25	*Lactobacillus gasseri*
	36	*Lactobacillus paragasseri**
Lactobacillus jensenii (99.9)	25	*Lactobacillus mulieris**
Moraxella osloensis (50.0), Enhydrobacter aerosaccus (50.0)**	5	*Moraxella osloensis*
Streptococcus agalactiae (99.9)	6	*Streptococcus urinalis**
Streptococcus equi spp zooepidemicus/agalactiae/equi spp equi (33.3/33.3/33.3)	3	*Streptococcus urinalis**

*Species lacking reference spectra
**Closely related species without single-species reference spectra

3.4. Infra species differentiation of Female Urinary Tract Microbiome members

Pangenome analysis of *Prevotella corporis* and *Prevotella brunnea* reveals features associated with their adaptation and virulence

The darkest place is under the candlestick - healthy urogenital tract as a source of worldwide disseminated Extraintestinal Pathogenic *Escherichia coli* lineages

Phylogenomic analysis of a highly virulent *Escherichia coli* ST83 lineage with potential animal-human transmission

Limosilactobacillus spp. diversity and relevant features of urogenital species

Pangenome analysis of *Prevotella corporis* and *Prevotella brunnea* reveals features associated with their adaptation and virulence

Magdalena Ksiezarek[1,2], Teresa Gonçalves Ribeiro[1,2], Svetlana Ugarcina Perovic[3], Tiago Antunes-Lopes[4,5,6], Filipa Grosso[1,2], Luisa Peixe[1,2]

[1]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[2]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

[3] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

[4]Department of Urology, Hospital de S. João, Porto, Portugal

[5]Faculty of Medicine, University of Porto, Portugal

[6]i3S – Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal

Manuscript in preparation

ABSTRACT

Objectives

To identify features contributing to adaptation and pathogenicity of novel *P. brunnea* and its closely related *P. corporis*.

Methods

P. corporis and *P. brunnea* strains isolated from the urogenital microbiome of asymptomatic and overactive bladder syndrome (OAB) women were characterized and genomic comparison with publicly available genomes was performed.

Results

The pangenome of *P. corporis* and *P. brunnea* had a total of 3429 gene clusters of which 1904 have known function. Both species possessed an array of putative virulence factors that could facilitate bacterial survival and competition (e.g., phospholipases, OmpA) and some virulence-associated traits seemed to be species specific (e.g., clostripain-like cysteine protease, sialic acid synthase). Two *P. corporis* and two *P. brunnea* strains, including the one from OAB woman, shared some genomic features (e.g., conjugal transfer proteins, arsenate resistance) and were predicted as human pathogens. All genomes encoded the *tetQ* gene and two *P. corporis* genomes encoded *cfxA* β-lactamase.

Conclusions

Our findings suggest differences in adaptation and virulence-related genomic features between *P. corporis* and *P. brunnea,* as well as between commensal and disease-associated strains.

INTRODUCTION

Since the description of the genus *Prevotella*[1], more than 50 species have been validly published[2,3]. Over the past decade, culture-independent (DNA sequencing based) studies have been reporting numerous *Prevotella* and closely related *Bacteroides* species as common members of the human microbiome[4].

Prevotella species associated with human microbiome are considered commensal bacteria due to their extensive presence in the healthy human body (e.g., oral cavity, gastrointestinal, genital and urinary tract), yet some species or specific strains have been associated with infections, inflammatory disorders or unhealthy status[2,5,6], suggesting that certain strains exhibit pathogenic properties. For instance, specific strains of *P. intermedia* and *P. nigrescens*, frequently found members of subgingival microbiome, were involved in the development of chronic periodontitis[7,8], and *Prevotella copri* has also been linked to

inflammatory diseases, although commonly detected in the gut microbiome of healthy individuals[2,9,10].

Currently, the urinary and urogenital microbiome is the less explored human microbial community due to long-lasting dogma that healthy urinary tract does not contain bacteria[11]. Still, *Prevotella* seems to be a relevant genus in this niche based on available studies. Most studies analyzing urine (voided/catheterized) by DNA-based approaches, reported *Prevotella* as a frequent member of the urogenital microbiome[2,12–18]. To date, *P. bivia*, *P. buccae*, *P. buccalis*, *P. copri*, *P. corporis, P. disiens, P. histicola, P. melaninogenica, P. oris* and *P. timonensis* constitute the repertoire of human urogenital microbiome[13,17,18], with *P. bivia* and *P. disiens* being reported most often[14–16,18].

Studying the urogenital microbiome of asymptomatic women and women diagnosed with overactive bladder syndrome, we identified the newly described *P. brunnea*[19] (this study), a closely related species of *P. corporis*. For both species, scarce knowledge is available, and their genomic properties have not been described yet.

In this study, whole-genome features, antimicrobial susceptibility, and biochemical patterns of *P. corporis* and *P. brunnea* strains isolated from the urogenital microbiome of asymptomatic and overactive bladder syndrome women were characterized. Further genomic comparison with publicly available genomes was performed to identify features contributing to adaptation and pathogenicity of those species.

MATERIAL AND METHODS

Bacterial isolation

Strains c24Ua_14AN and c17Ua_125AN were isolated within the ongoing study on female urogenital microbiome. First-morning voided urine samples were collected and processed based on culturomics approach[18]. Strains c24Ua_14AN and c17Ua_125AN were recovered from mixed culture from Columbia agar with 5% Sheep Blood (bioMérieux) after sampling 100 µl of urine incubated in anaerobic atmosphere, 37 °C for 48 hours. Strain c24Ua_14AN was isolated in April 2018 from the urine of a 27-year old asymptomatic woman, while the strain c17Ua_125AN was isolated in October 2017 from the urine of a 38-year old woman diagnosed with overactive bladder syndrome, both from microbiome composed of diverse species. Both strains were initially subjected to identification by VITEK matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry (bioMérieux), but

could not be identified at species, nor at genus level. Further, strains were identified as *Prevotella* spp. by amplification and sequencing of the 16S rRNA gene[20].

DNA isolation and sequencing

The total DNA was extracted with DNeasy Blood & Tissue Kit (Qiagen, Germany), and whole-genome sequencing was performed by Illumina HiSeq 2x150 nt. Reads were trimmed by Trimmomatic version 0.36[21] and quality checked using FastQC[22]. *De novo* assembly was performed by SPADes version 3.12.0[23] with quality of assembly assessed by QUAST[24]. Genome size of the strain c24Ua_14AN was 2 951 478 bp, assembled in 105 contigs, and with a GC content of 43.4%, while the genome size of the strain c17Ua_125AN was 2 749 488 bp, assembled in 252 contigs, and with a GC content of 43.7%. The genome was annotated with NCBI Prokaryotic Genome Annotation Pipeline[25]. The annotated draft genomes of strains c24Ua_14AN and c17Ua_125AN has been deposited in DDBJ/ENA/GenBank under the accession numbers JABUXP000000000 and VFFG00000000, respectively (BioSample SAMN15159719 and BioSample SAMN12023332, respectively; BioProject PRJNA548360).

Genomes collection

Assembled sequences of 3 *Prevotella corporis* and 2 *Prevotella brunnea* stored in the National Center for Biotechnology Information (NCBI) genome assembly database as of 24 September 2020 were collected. Two collected genomes represented the type strain of *P. corporis* (DSM 18810=JCM 8529), thus one of them was excluded from further analysis (Assembly ASM61336v1). In this study we focused only on isolate genomes available, and we did not include data derived from metagenomic studies. Additionally, we included our two newly deposited *Prevotella* genomes i.e., strains c24Ua_14AN and c17Ua_125AN.

To confirm the species of *Prevotella*, we calculated the ANI values based on BLAST+ (ANIb) using JSpecies[26] Web Server. An ANI cutoff value of 95% was used as species delimitation criteria[27]. The ANI values confirmed that strains c24Ua_14AN, MJR7716, and DSM_18810^T belonged to the species *P. corporis*, while strains c17Ua_125AN, Marseille_P4334, and A2672^T belonged to the species *P. brunnea* as they showed ANI values ≥ 95% (**Supplementary Table S1**).

Therefore, 6 draft genomes were subjected to further analysis (**Table 1**). Metadata was also collected for each genome from NCBI BioSample entries (**Table 1**).

Pangenome analysis

Pangenome analysis was performed with anvi'o[28] v6.2. Prodigal was used for protein-coding genes prediction[29]. Predicted genes were grouped into Clusters of Orthologous Groups (COGs) using anvi'o script with MUSCLE[30] for sequence alignment. Hierarchical clustering was performed with default Euclidean distance and Ward algorithm. Pangenome was divided into collections classified as follows: core genome represented by gene clusters present in 100% of genomes; remaining genes were classified as accessory genome including unique genes (gene clusters representing singletons, present in just one unique genome). Additional layers of metadata (incorporated in **Fig. 1**) were applied to pangenome analysis in anvi'o which allowed us to investigate functional associations among groups of genomes and identify functions enriched according to species or prediction of human pathogen. Output of functional enrichment[31] provided by anvi'o also included enrichment score and p value. Individual genes with no functional COG annotation were searched by NCBI Protein Blast database (BLASTp)[32]. Pangenome plot was generated by anvi'o and edited in Inkscape[33]. Figure representing COG distribution was created in R[34] version 3.6.2 ggplot2[35] package version 3.3.2. In case of multiple COG categories representing single gene call, for the purpose of Table 3 and Figure 2 we followed the first category as the most significant hit. Complete results including all COG categories are available in **Supplementary Table S2**.

Pathogenic potential, mobilome, and antibiotic resistance genes

Draft genomes were subjected to human pathogen prediction online tool PathogenFinder[36] 1.1 available on the Center for Genomic Epidemiology server (https://www.genomicepidemiology.org/). ICEfinder[37] online tool was used to detect putative integrative and conjugative elements (ICEs) and integrative mobilizable elements (IMEs). PlasmidFinder[38] 2.1 available on the Center for Genomic Epidemiology server was used to screen for potential presence of plasmids. Phaster[39] was used to search for putative prophages sequences. Screening of genes presumed to confer antibiotic resistance was searched in the Comprehensive Antibiotic Resistance Database (CARD[40]; version 3.1.0) using Resistance Gene Identifier (RGI; version 5.1.1), with criteria for perfect and strict hits only and high-quality coverage, and also in the ResFinder[41] 4.1 available on the Center for Genomic Epidemiology server, with criteria of 90% as ID threshold, and minimum length of 60%. Findings were compared with the available phenotypic susceptibility test results.

Antimicrobial susceptibility testing

The two strains (c24Ua_14AN and c17Ua_125AN) isolated in our laboratory were further subjected to phenotypic antimicrobial susceptibility testing. Minimal Inhibitory Concentrations (MICs) were determined by agar dilution method according to Clinical and Laboratory Standards Institute (CLSI)[42] recommendation for anaerobic bacteria. Brucella agar (LIOFILCHEM) supplemented with 5 µg/ml hemin (SIGMA), 1 µg/ml Vitamin K (Merck KGaA, Germany) and 5% v/v sheep blood (ThermoFisher) was used. Agar plates were prepared for 8 antibiotics (Penicillin G, Amoxicillin with Clavulanic acid, Cefoxitin, Imipenem, Clindamycin, Metronidazole, Chloramphenicol and Tetracycline) in concentrations 0.125, 0.25, 0.5, 1.0, 2.0, 4.0, 8.0, 16.0, 32.0 and 64.0 µg/ml. Suspensions equivalent to 0.5 McFarland of each strain were directly applied on the agar plates and incubated at 37°C for 48 hours under anaerobic atmospheric conditions. Results were interpreted according to the CLSI recommendations (https://clsi.org/).

Biochemical characterization

The two strains (c24Ua_14AN and c17Ua_125AN) were also subjected to biochemical characterization by Rapid ID 32 A and API 50 CH (bioMérieux). Complete reaction panels were performed for 72 h culture from Columbia Agar with 5% Sheep Blood (bioMérieux), according to instructions provided by the manufacturer.

RESULTS

Genome features

The draft genomes of available *P. corporis* and *P. brunnea* strains were nearly equal regarding genome size and number of predicted genes per species (**Table 1**). Average genome size for *P. corporis* (n=3) was 2.87 Mb with mean of 2513 genes, and for *P. brunnea* strains (n=3) was 2.81 Mb with 2458 predicted genes. None of these species had a representative complete genome available. Both species were isolated from different human sources: *P. corporis* exclusively from urogenital tract, and *P. brunnea* from wound, feces and urine (**Table 1**). *P. corporis* was isolated in the United Kingdom, USA and Portugal, while *P. brunnea* was so far isolated from European countries (Germany, France, and Portugal) (**Table 1**). Genome summaries and features are listed in **Table 1**.

Pangenome characterization

The pangenome of *P. corporis* and *P. brunnea* comprised a total of 14 912 genes, classified to 3 429 gene clusters, of which 1904 clusters have known function. Pangenome of *P. corporis* and *P. brunnea* strains visualized by anvi'o is presented in **Figure 1**. The core genome contained 10570 genes (1532 gene clusters), and the accessory genome contained 4342 genes (1897 gene clusters), of which 810 were unique genes (795 gene clusters) present in a single genome (**Supplementary Table S2**). The majority of core gene clusters had known COG function (82%; 1252/1532), which were related to housekeeping functions such as translation, ribosomal structure and biogenesis (10.1%; 1066/10570 genes), cell wall/membrane/envelope biogenesis (7.7%; 817/10570 genes) or replication, recombination and repair (6.8%; 718/10570 genes). Besides essential genes required for basic cellular functions, within the core we also observed some gene calls that could contribute to *P. corporis* and *P. brunnea* survival skills and virulence e.g., opacity proteins (COG3637) or membrane associated OmpA (COG2885), TonB (COG0810), or TolC (COG1538) proteins. We also found COGs related to metabolism and/or virulence conserved among these species e.g., various phospholipases, likely belonging to type A2 and patatin-like proteins or ABC-type bacteriocin exporters that could act as defense mechanism.

Contrary to core, in the accessory genome most of the gene clusters have unknown COG function (66%; 1245/1897). Among the 4342 accessory genes detected, 1579 could be classified into specific COG category (36,4%), with most of them belonging to replication, recombination and repair (4.9%; 211 genes), mobilome: prophages, transposons (3.8%; 164 genes) or transcription (3.3%; 142 genes). **Supplementary Table S2** contains all genes found in 6 draft genomes, with their gene clusters, COG category classification, putative function, and amino acid sequences.

Using ICEfinder we also detected one to two putative ICEs (mostly type T4SS-type ICEs) and two to five putative IMEs, independently of species. Multiple integrases and relaxases were detected on putative ICEs (**Supplementary Table S3**) and conjugation related proteins, of which most were members of Type IV Secretion System (T4SS), including VirB proteins (e.g., VirB3, reported for all *P. brunnea* strains, and *P. corporis* strain DSM 18810^T). We also found other VirB proteins (e.g., VirB4, VirB9) belonging to T4SS in all strains analyzed, however due to genomes incompleteness we were not able to identify all proteins from the complex.

None prophage sequences (intact, incomplete or questionable) per genome were identified in *P. brunnea* strains. However, two and one incomplete prophage sequences were identified in

P. corporis strain c24Ua_14AN and DSM 18810^T, respectively. Interestingly, in core genome of *P. brunnea* and *P. corporis* we found phage-related lysozyme (muramidase) which is phage-derived gene that lost its toxicity and now plays an essential role in core cellular functions[43]. Plasmid replicon sequences were not detected in any of the strains when queried against the PlasmidFinder 2.1 databases, which is likely due to database insufficiency.

The draft genomes were further subjected to PathogenFinder 1.1 which suggested that *P. corporis* strain c24Ua_14AN isolated from a healthy host, and *P. brunnea* strain A2672 were not predicted as human pathogens (probability=0.474 and 0.447, respectively; **Table 2**), although the latter was isolated from a wound of a patient as part of a mixed infection also involving *Staphylococcus aureus*, *Actinotignum schaalii*, *Enterobacter cloacae* complex and *Klebsiella pneumoniae*[19]. Contrary, *P. corporis* strains DSM 18810^T and MJR7716, and *P. brunnea* strains Marseille P4334 and c17Ua_125AN were estimated to be pathogenic in humans (probability=0.811, 0.757, and 0.765, 0.713, respectively; **Table 2**). Of note, strain c17Ua_125AN was isolated from a diverse urogenital microbiome of woman diagnosed with overactive bladder, a syndrome where microbiome dysbiosis has been suggested to be involved in its etiology[44]. According to PathogenFinder data, multiple conjugal transfer proteins (e.g., TraE, TraF, TraG, TraJ, TraM, TraN), and/or two-component system response regulator rteB, Exc protein involved in excision of conjugative transposons or tyrosine type site-specific recombinase have been detected in these potential pathogenic strains (**Table 2**). We identified 54 COG functions specific to these 4 potentially pathogenic strains however, only the DNA-binding transcriptional regulator AlpA, belonging to helix-turn-helix domain containing DNA-binding protein, was found to be significantly functionally enriched in this group ($p<0.05$) (**Supplementary Table S4**). Functional enrichment analysis was also performed for strains isolated from diseased host based on available metadata. Three functions were significantly associated with these strains ($p<0.05$) namely Arsenite efflux pump ArsB, ACR3 family (COG0798), DNA modification methylase belonging to Conserved Protein Domain Family YhdJ (COG0863) and DNA-binding transcriptional regulator, MarR family (COG1846) (**Supplementary Table S5**).

Genomic characteristics specific to *Prevotella corporis* and *Prevotella brunnea*

We identified gene calls specific to each species, with *P. corporis* represented by 861 genes (283 clusters) and *P. brunnea* represented by 894 genes (286 clusters) (**Figure 1**). Although the genomes collection is very small, several COG functions were found to be significantly

enriched ($p < 0.05$) for each species, with 51 hits being identified for *P. corporis* and 19 hits for *P. brunnea* (**Figure 2)**. Complete list of all significantly enriched COG functions for these species is presented in **Table 3**. Of note, these findings should be validated on a larger genome collection.

P. corporis enriched COG classes were mainly associated with energy production and conversion (C) [e.g., L-lactate utilization protein LutB (COG1139), Polyferredoxin (COG0348), Fe-S oxidoreductase (COG0247)], defense mechanisms (V) [e.g., Alkyl hydroperoxide reductase subunit AhpC (peroxiredoxin) (COG0450), Alkyl hydroperoxide reductase subunit AhpF (COG3634)], or amino acid transport and metabolism (E). Additionally, gene related to cell motility such as *flgK*, coding for flagellar hook-associated protein FlgK (COG1256) was detected only in *P. corporis*. Most of these specific proteins likely represent fitness factors e.g., L-lactate utilization protein or Polyferredoxin, however, some could be characterized as putative virulence factors e.g., DNA-binding transcriptional regulator, XRE family (COG3655) or predicted metalloprotease (COG2321) which by NCBI BLASTp is annotated as neutral zinc metallopeptidase. *P. corporis* also possessed mobilome related prophage maintenance system killer protein (COG3654), also called Fic family protein or type II toxin-antitoxin system death-on-curing family toxin.

Each strain of *P. corporis* also contained unique amino acid sequences (from 142 to 191 singletons/strain), of which most did not have predicted function (**Supplementary Table S2**). Among those strain-specific proteins potentially involved in niche adaptation, survival and/or putative virulence we can include lysozyme family protein (COG3926) specific to DSM_18810^T strain; cold shock protein, CspA family (COG1278) specific to MJR7716 or toxin component of the Txe-Axe toxin-antitoxin module, Txe/YoeB family (COG4115) specific to c24Ua_14AN strain.

As per the other species, *P. brunnea* enriched COG classes (**Figure 2**) included inorganic ion transport and metabolism (P) [fluoride ion exporter CrcB/FEX, affects chromosome condensation (COG0239), periplasmic ferric-dicitrate binding protein FerR, regulates iron transport through sigma-19 (COG3712)], intracellular trafficking, secretion, and vesicular transport (U) such as the preprotein translocase subunit SecE (COG0690) or cell wall/membrane biogenesis (M). We also identified numerous putative virulence factors among *P. brunnea* specific proteins, e.g., Ca2+-binding protein, RTX toxin-related (COG2931) which by NCBI BLASTp is highly homologous with Peptidase_C11- Clostripain family protein of few other *Prevotella* species; or proteins involved in synthesis of sialic acid i.e., CMP-N-acetylneuraminic acid synthetase (COG1083), sialic acid synthase SpsE

(COG2089). We further observed that indeed all *3 P. brunnea* strains contained complete operon for sialic acid synthesis composed of *neuD, neuB, neuC* and *neuA* genes. Furthermore, uncharacterized N-terminal domain of lipid-A-disaccharide synthase (COG3952) and 4-amino-4-deoxy-L-arabinose (COG1807) are involved in lipopolysaccharide (LPS) modification. Another could be Fido, protein-threonine AMPylation domain (COG2184) which more specifically represents cell filamentation protein Fic.

Intra-species *P. brunnea* diversity was also observed with a range of 72 to 153 singletons/strain (**Supplementary Table S2**). We identified genes calls that could contribute to virulence and/or survival abilities e.g., unique amino acids sequence representing ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain (COG2274) present only in c17Ua_125AN strain, isolated from overactive bladder patient. Interestingly, based on BLASTp and reference proteins database search, this bacteriocin-related protein is highly similar to those found in *Lactobacillus* genomes (~96% amino acids identity with *L. crispatus*). Furthermore, we identified additional proteases specific to A2672^{T} and Marseille_P4334 strains.

A list of *P. corporis* and *P. brunnea* species and/or strains specific COG functions putatively associated with their pathogenicity is presented in **Table 4**.

Metabolism of carbohydrates

Analysis of the metabolic potential of *P. brunnea* strain c17Ua_125AN and *P. corporis* strain c24Ua_14AN provided insight into the biochemical reactions underlying the complex growth requirements of these fastidious organisms. Our genome-wide metabolic reconstruction analyses suggested that both *P. brunnea* and *P. corporis* can utilize a limited variety of carbohydrates, including D-glucose, D-maltose, gelatin, D-mannose, and starch. Additionally, we observed that these strains are lacking enzymes such as lactate dehydrogenase, mannitol phosphate dehydrogenase, sucrose phosphorylase, urease, and galactosidases, thus they would be unable to ferment lactose, mannitol, sucrose, α-galactosides (melibiose and raffinose), and to hydrolyze urea. Phenotypic assays confirmed that, as predicted, these *P. brunnea* and *P. corporis* strains ferments D-glucose, D-maltose, D-mannose and starch. Results of API Rapid ID 32 A and API 50 CH reaction panels are available in **Supplementary Table S6** and **S7**, respectively.

Antibiotic resistance

We applied ResFinder 4.1 and RGI by CARD to identify acquired antibiotic resistance genes among the *Prevotella* genomes. All strains harbored the *tetQ* gene, associated with resistance to tetracyclines, and the beta-lactamase gene *cfxA* was detected in 2 *P. corporis* strains (DSM 18810^T and MJR7716). In addition, we assessed antibiotic susceptibility of two urogenital strains (*P. corporis* c24Ua_14AN and *P. brunnea* c17Ua_125AN), to 8 clinically relevant antibiotics. Both strains were susceptible to all antibiotics tested, even tetracycline: penicillin G (MICs, 0.125 µg/ml), cefoxitin (MIC, 0.125 µg/ml), amoxicillin-clavulanic acid (MICs, 0.125 µg/ml), imipenem (MICs, 0.125 µg/ml), clindamycin (MICs, 0.125 µg/ml), metronidazole (MICs, 0.125 µg/ml), tetracycline (MICs, 2.0 µg/ml), and chloramphenicol (MICs, 0.125-0.5 µg/ml). These MICs corresponded, in most cases, to the lower values of MIC distributions, according to EUCAST collated data for *Prevotella*.

DISCUSSION

In this study, we report pangenome analysis of two poorly explored *Prevotella* species, namely *P. brunnea* and *P. corporis* isolated from human microbiome and/or from infections.

Overall, genome sizes and the number of genes predicted for *P. brunnea* and *P. corporis* were in the same range as previously observed regarding *Prevotella* genus[2,45].

Our pangenome analysis of *P. corporis* and *P. brunnea* suggests relatively high relatedness of these species with 44.7% of gene families representing core genome (corresponding to previously reported conserved functions for oral *Prevotella* species)[46] and 23.2% representing singletons. Although accessory genes were mostly without functional annotation, nearly 150 accessory genes were related to mobilome, which suggest that the variability observed among the strains most probably originates from horizontal gene transfer. Phage elements are capable of mediating genetic exchange between bacteria, including transfer of virulence and antimicrobial-resistance genes. Considering previous reports of common phages in *Prevotella* abundant microbiomes[47,48], we suspect that phages are underestimated in our study due to usage of draft genomes. Nevertheless, from our initial report *P. corporis* tends to carry more prophage sequences compared to *P. brunnea*. We also found mobilome analysis to be challenging using widely accessible web-based tools with insufficiencies in their databases, designed mostly for well-known Gram-positive or fast-growing Gram-negative bacteria (e.g., Enterobacterales).

We performed a comparative genomics of *P. corporis* and *P. brunnea* species and identified core genome related mostly to bacterial housekeeping functions. However, among conserved genes for both species we also identified putative virulence-associated features that could provide an advantage for bacterial survival and cellular attachment (**Table 4**). For instance, opacity proteins were previously shown to have an important role in the adherence and by phase-variation enable escape the immune system of genital pathogen *Neisseria gonorrhoeae*, thus contributing to bacterial persistence in genital tract during infection[49,50]. We also observed various conserved patatin-like phospholipases, that are often recognized as bacterial virulence factor due to their contribution to host cell invasion or escaping host defense mechanisms during the infection[51–53]. Furthermore, several outer membrane protein such as OmpA, TonB or TolC are conserved among those species and were previously reported as virulence and fitness feature in Gram-negative pathogens e.g., *Escherichia*, *Haemophilus, Salmonella*[54–57].

We further identified a set of *P. corporis* and *P. brunnea* species-specific genes that also could contribute to their pathogenicity. Notably, due to the very small collection of genomes available, our species-specific findings should be treated as preliminary and will require further validation on larger strains collection.

For instance, in *P. corporis* we found motility related proteins (e.g., FlgK) that might be associated with secretion systems and consequently, in secretion of virulence proteins[58,59], metalloproteases that are known to degrade extracellular matrix, thus interfere with cellular communication and processes, and may play a role in disease pathology[60–62] and proteins involved in defense mechanism that are responsible for protection against oxidative stress (e.g., alkyl hydroperoxide reductases), are involved in biofilm formation and contributing to bacterial survival and persistence[63,64]. Furthermore, we also found prophage maintenance system killer protein (Fic domain) which belongs to a toxins family (toxin-antitoxin system) with potential harmful effect for host cells[65,66]. Although in our study *P. corporis* was isolated from healthy urogenital microbiome, in the context of infection this species was previously reported from wound, cervical and rectal abscess, and pleuropulmonary infection[46,67–69].

Regarding *P. brunnea*, besides the initial isolation report (newly described in late 2019) from mixed bacterial wound infection[19], its possible associations with human infections are not reported yet in the literature. Interestingly, in this species we found proteins involved in synthesis of sialic acid, which might be relevant for hiding from host's immune response, thus bacterial virulence[70,71]. Moreover, in all *P. brunnea* strains we identified RTX toxin-related protease, which represents clostripain-like protein that could increase bacterial survival[72–74].

The clostripain-like proteins belong to C11 family cysteine proteases, and they are found in many commensal bacteria (mostly of Bacteroidetes and Firmicutes phylum)[74,75]. However, they are also involved in pathogenic processes such as host immune responses modulation, activation of bacterial toxins or adhesion to host epithelium[72–74,76]. The currently known proteases from C11 family are e.g., fragipain encoded by *Bacteroides fragilis*, clostripain encoded by *Clostridium histolyticum*, gingipains encoded by *Porphyromonas gingivalis* and interpain A from *Prevotella intermedia*[46,72–75]. To the best of our knowledge, interpain A is the only cysteine protease described to date from *Prevotella* species. Thus, here we describe putatively novel cysteine protease encoded by *P. brunnea*, with high amino acid similarity (95.6-96.4%) only to few deposited *Prevotella*-related protein sequences (*Prevotella buccalis*, *Prevotella bergensis*, *Prevotella colorans* and *Prevotella timonensis*). We also identified two additional gene calls involved in modification of lipid A (LPS), specific for *P. brunnea* strains, which could be of importance regarding survival and pathogenicity of this species[46,77–79]. For instance, lipid A modification with 4-amino-4-deoxy-L-arabinose can lead to reduction of LPS negative charge and thus, suppress binding of cationic molecules e.g., antimicrobial peptides[79,80]. *P. brunnea* also expressed a certain level of intra-species diversity with e.g., bacteriocin-related protein highly similar to these from *Lactobacillus* strains. They can play an important role in microbe-microbe competition and microbe-host interactions and contribute to adaptation and potential virulence.

In strains predicted as human pathogens, we found significant association of DNA-binding transcriptional regulator AlpA. AlpA activator is required for transfer/induction of phage-related mobile genetic elements in Gram-negative bacteria and is associated with bacterial pathogenicity (phage-inducible chromosomal islands)[81]. We also found that strains isolated from diseased host were significantly enriched in 3 gene clusters e.g., arsB, which seems to be very common among prokaryotes[82] or MarR family of proteins, detected for the first time in multidrug resistant *E. coli* strain, which may be involved in various processes such as control of metabolic pathways, stress responses, environment adaptation, virulence genes or antibiotics efflux pumps genes[83,84]. Interestingly, previous studies on *P. intermedia* and *P. nigrescens* also highlighted that diseased genomes cohort was enriched in genes related to arsenate resistance[7].

To the best of our knowledge, there have been no previous studies describing the antimicrobial susceptibility of *P. corporis* and *P. brunnea*. Our study demonstrated highly susceptible

antimicrobial phenotypes of *these urogenital strains*, although high resistance to clindamycin, and susceptible rates to imipenem, meropenem and amoxicillin-clavulanate were previously observed within this genus[85–87]. Here, we also reported sporadic presence of *cfxA* among *P. corporis* strains and presence of *tetQ* in both species. In the closely related *Bacteroides* genus, *tetQ* transfer is stimulated by low concentration of tetracycline and it is disseminated by conjugative transposons (CTns) i.e., CTnDOT, proven to be an ICE[88–91]. In the previous studies, *tetQ* was also often present together with *cfxA* in *Bacteroides* and other *Prevotella* strains[85]. Indeed, it was demonstrated that tetracycline mediates the transfer of mobilizable transposon Tn4555 containing *cfxA* gene[92,93]. Thus, the common consumption of broad-spectrum tetracycline antibiotics by humans and usage in agriculture and among farm animals likely contributes to continuous dissemination of these resistance genes. The identification of *tetQ* in isolates recovered from urine suggests that these species/strains may be one of the sources of tetracycline resistance genes among bacteria of the urogenital microbiome. Interestingly, despite presence of *tetQ* gene in our urogenital strains, we observed lower MICs to tetracycline (MICs of 2.0 μg/ml), compared to previously reported[94]. Discrepancy of phenotypic and genotypic resistance to tetracycline is occasionally reported[94] and require further studies.

In conclusion, although strict set of virulence factors of *Prevotella* species essential to develop disease in human have not been define, here we report a variety of genes (e.g., proteases, sialic acid synthetases, bacteriocin exporters, motility-related genes) of *P. corporis* and *P. brunnea* strains that may act advantageously on their fitness, survival, and bacterial competition abilities, as well as potentially constitute a thread for the host cells. We demonstrated that strains predicted as pathogens and isolated from diseased host share traits associated with adaptation and pathogenicity. Moreover, our data suggest that these *Prevotella* species may act as a source of antimicrobial resistance genes (e.g., *tetQ*) for other bacteria.
Further sequencing and characterization of strains from healthy and diseased hosts of these species would allow to verify and understand the relevance of genomic characteristics here identified.

Protologue

The Whole Genome Shotgun project of the isolate c17Ua_125AN has been deposited at DDBJ/ENA/GenBank under the accession VFFG00000000. The Whole Genome Shotgun project of the isolate c24Ua_14AN has been deposited at DDBJ/ENA/GenBank under the

accession JABUXP000000000. The versions described in this paper are versions VFFG01000000 and JABUXP01000000, respectively.

AUTHOR STATEMENTS

Funding information

This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is financed by national funds from FCT (UIDP/04378/2020 and UIDB/04378/2020). MK is supported by FCT [SFRH/BD/132497/2017]. FG is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0034]. SUP was supported from the project NORTH-01-0145-FEDER-000024.

Contributions

MK performed genomic extraction and analysis, pangenome analysis, antibiotic susceptibility testing and wrote the manuscript. TGR performed ANIb analysis, carbohydrates characterization, revised and edited the manuscript. SUP performed MALDI-TOF analysis and revised the manuscript. TA-L recruited an overactive bladder patient and provided the sample. FG contributed to manuscript revision. LP contributed to reviewing the manuscript, project design and administration, and funding.

Acknowledgements

We would like to thank bioMérieux (Portugal, Lda) for equipment and material provided. We would also like to acknowledge the medical staff of the Department of Urology (São João University Hospital, Porto) for their support in OAB patient recruitment and sample collection.

Conflicts of interest

The authors declare that there are no conflicts of interest.

ABBREVIATIONS

CLSI, Clinical and Laboratory Standard Institute; COG, Cluster of Orthologous Group; ICE, integrative and conjugative element; IME, integrative mobilizable element; LPS, lipopolysaccharide; MIC, minimal inhibitory concentration; OAB, overactive bladder.

REFERENCES

1 Shah H. N., Collins D. M. *Prevotella*, a new genus to include *Bacteroides melaninogenicus* and related species formerly classified in the genus *Bacteroides*. *Int J Syst Bacteriol* 1990;**40**(2):205–8. Doi: 10.1099/00207713-40-2-205.

2 Tett A., Pasolli E., Masetti G., Ercolini D., Segata N. *Prevotella* diversity, niches and interactions with the human host. *Nature Reviews Microbiology* 2021:1–15. Doi: 10.1038/s41579-021-00559-y.

3 Parte A. C., Sardà Carbasse J., Meier-Kolthoff J. P., Reimer L. C., Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;**70**(11):5607–12. Doi: 10.1099/ijsem.0.004332.

4 Wexler H. M. *Bacteroides*: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* 2007;**20**(4):593–621. Doi: 10.1128/CMR.00008-07.

5 Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. *Immunology* 2017;**151**(4):363–74.

6 Iljazovic A., Roy U., Gálvez E. J. C., Lesker T. R., Zhao B., Gronow A., et al. Perturbation of the gut microbiome by *Prevotella* spp. enhances host susceptibility to mucosal inflammation. *Mucosal Immunol* 2020. Doi: 10.1038/s41385-020-0296-4.

7 Zhang Y., Zhen M., Zhan Y., Song Y., Zhang Q., Wang J. Population-Genomic Insights into Variation in *Prevotella intermedia* and *Prevotella nigrescens* Isolates and Its Association with Periodontal Disease. *Front Cell Infect Microbiol* 2017;**7**. Doi: 10.3389/fcimb.2017.00409.

8 Jia L., Han N., Du J., Guo L., Luo Z., Liu Y. Pathogenesis of Important Virulence Factors of *Porphyromonas gingivalis* via Toll-Like Receptors. *Front Cell Infect Microbiol* 2019;**9**. Doi: 10.3389/fcimb.2019.00262.

9 Tett A., Huang K. D., Asnicar F., Fehlner-Peach H., Pasolli E., Karcher N., et al. The *Prevotella copri* Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host & Microbe* 2019;**26**(5):666-679.e7. Doi: 10.1016/j.chom.2019.08.018.

10 Drago L. *Prevotella Copri* and Microbiota in Rheumatoid Arthritis: Fully Convincing Evidence? *Journal of Clinical Medicine* 2019;**8**(11):1837. Doi: 10.3390/jcm8111837.

11 Wolfe A. J., 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;**50**(4):1376–83. Doi: 10.1128/JCM.05852-11.

12 Pearce M. M., Hilt E. E., Rosenfeld A. B., Zilliox M. J., Thomas-White K., Fok C., et al. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *MBio* 2014;**5**(4):e01283-01214. Doi: 10.1128/mBio.01283-14.

13 Morand A., Cornu F., Dufour J-C., Tsimaratos M., Lagier J-C., Raoult D. Human Bacterial Repertoire of the Urinary Tract: a Potential Paradigm Shift. *J Clin Microbiol* 2019;**57**(3). Doi: 10.1128/JCM.00675-18.

14 Thomas-White K., Forster S. C., Kumar N., Van Kuiken M., Putonti C., Stares M. D., et al. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nature Communications* 2018;**9**(1):1557. Doi: 10.1038/s41467-018-03968-5.

15 Hilt E. E., McKinley K., Pearce M. M., Rosenfeld A. B., Zilliox M. J., Mueller E. R., 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;**52**(3):871–6. Doi: 10.1128/JCM.02876-13.

16 Coorevits L., Heytens S., Boelens J., Claeys G. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. *Eur J Clin Microbiol Infect Dis* 2017;**36**(4):635–9. Doi: 10.1007/s10096-016-2839-x.

17 Dubourg G., Morand A., Mekhalif F., Godefroy R., Corthier A., Yacouba A., et al. Deciphering the Urinary Microbiota Repertoire by Culturomics Reveals Mostly Anaerobic Bacteria From the Gut. *Front Microbiol* 2020;**11**. Doi: 10.3389/fmicb.2020.513305.

18 Ksiezarek M., Ugarcina-Perovic S., Rocha J., Grosso F., Peixe L. Long-term stability of the urogenital microbiota of asymptomatic European women. *BMC Microbiol* 2021;**21**(1):64. Doi: 10.1186/s12866-021-02123-3.

19 Buhl M., Dunlap C., Marschal M. *Prevotella brunnea* sp. nov., isolated from a wound of a patient. *Int J Syst Evol Microbiol* 2019;**69**(12):3933–8. Doi: 10.1099/ijsem.0.003715.

20 Héritier C., Poirel L., Nordmann P. Genetic and biochemical characterization of a chromosome-encoded carbapenem-hydrolyzing ambler class D beta-lactamase from *Shewanella algae*. *Antimicrob Agents Chemother* 2004;**48**(5):1670–5. Doi: 10.1128/aac.48.5.1670-1675.2004.

21 Bolger A. M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**(15):2114–20. Doi: 10.1093/bioinformatics/btu170.

22 FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

23 Bankevich A., Nurk S., Antipov D., Gurevich A. A., Dvorkin M., Kulikov A. S., et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**(5):455–77. Doi: 10.1089/cmb.2012.0021.

24 Gurevich A., Saveliev V., Vyahhi N., Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**(8):1072–5. Doi: 10.1093/bioinformatics/btt086.

25 Haft D. H, DiCuccio M., Badretdin A., Brover V., Chetvernin V., O'Neill K., et al. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Research* 2018;**46**(D1):D851–60. Doi: 10.1093/nar/gkx1068.

26 Richter M., Rosselló-Móra R., Oliver Glöckner F., Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 2016;**32**(6):929–31. Doi: 10.1093/bioinformatics/btv681.

27 Meier-Kolthoff J. P., Auch A. F., Klenk H-P., Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;**14**:60. Doi: 10.1186/1471-2105-14-60.

28 Eren A. M., Kiefl E., Shaiber A., Veseli I., Miller S. E., Schechter M. S., et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 2021;**6**(1):3–6. Doi: 10.1038/s41564-020-00834-3.

29 Hyatt D., Chen G-L., LoCascio P. F., Land M. L., Larimer F. W., Hauser L. J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119. Doi: 10.1186/1471-2105-11-119.

30 Edgar R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004;**32**(5):1792–7. Doi: 10.1093/nar/gkh340.

31 Shaiber A., Willis A. D., Delmont T. O., Roux S., Chen L-X., Schmid A. C., et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol* 2020;**21**(1):292. Doi: 10.1186/s13059-020-02195-w.

32 Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. Basic local alignment search tool. *J Mol Biol* 1990;**215**(3):403–10. Doi: 10.1016/S0022-2836(05)80360-2.

33 Inkscape Project. Inkscape [Internet]. 2020. Available from: https://inkscape.org

34 R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.; 2018.

35 Wickham H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* 2016;(ISBN 978-3-319-24277-4).

36 Cosentino S., Voldby L. M., Møller Aarestrup F., Lund O. PathogenFinder--distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One* 2013;**8**(10):e77302. Doi: 10.1371/journal.pone.0077302.

37 Liu M., Li X., Xie Y., Bi D., Sun J., Li J., et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Research* 2019;**47**(D1):D660–5. Doi: 10.1093/nar/gky1123.

38 Carattoli A., Zankari E., García-Fernández A., Voldby L. 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;**58**(7):3895–903. Doi: 10.1128/AAC.02412-14.

39 Arndt D., Grant J. R., Marcu A., Sajed T., Pon A., Liang Y., et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;**44**(W1):W16-21. Doi: 10.1093/nar/gkw387.

40 Alcock B. P., Raphenya A. R., Lau Tammy T. Y., Tsang K. K., Bouchard M., Edalatmand A., et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;**48**(D1):D517–25. Doi: 10.1093/nar/gkz935.

41 Bortolaia V., Kaas R. S., Ruppe E., Roberts M. C., Schwarz S., Cattoir V., et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;**75**(12):3491–500. Doi: 10.1093/jac/dkaa345.

42 CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 27th ed. CLSI supplement M100. 2017.

43 Randich A. M., Kysela D. T., Morlot C., Brun Y. V. Origin of a Core Bacterial Gene via Co-option and Detoxification of a Phage Lysin. *Curr Biol* 2019;**29**(10):1634-1646.e6. Doi: 10.1016/j.cub.2019.04.032.

44 Antunes-Lopes T., Vale L., Coelho A. M., Silva C., Rieken M., Geavlete B., et al. The Role of Urinary Microbiota in Lower Urinary Tract Dysfunction: A Systematic Review. *European Urology Focus* 2020;**6**(2):361–9. Doi: 10.1016/j.euf.2018.09.011.

45 Gupta V. K., Chaudhari N. M., Iskepalli S., Dutta C. Divergences in gene repertoire among the reference *Prevotella* genomes derived from distinct body sites of human. *BMC Genomics* 2015;**16**(1). Doi: 10.1186/s12864-015-1350-6.

46 Ibrahim M., Subramanian A., Anishetty S. Comparative pan genome analysis of oral *Prevotella* species implicated in periodontitis. *Funct Integr Genomics* 2017;**17**(5):513–36. Doi: 10.1007/s10142-017-0550-3.

47 Devoto A. E., Santini J. M., Olm M. R., Anantharaman K., Munk P., Tung J., et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nature Microbiology* 2019;**4**(4):693–700. Doi: 10.1038/s41564-018-0338-9.

48 Edwards R. A. Prodigious *Prevotella* phages. *Nature Microbiology* 2019;**4**(4):550–1. Doi: 10.1038/s41564-019-0419-4.

49 Stein D. C., LeVan A., Hardy B., Wang L-C., Zimmerman L., Song W. Expression of Opacity Proteins Interferes with the Transmigration of *Neisseria gonorrhoeae* across Polarized Epithelial Cells. *PLOS ONE* 2015;**10**(8):e0134342. Doi: 10.1371/journal.pone.0134342.

50 Cole J. G., Fulcher N. B., Jerse A. E. Opacity Proteins Increase *Neisseria gonorrhoeae* Fitness in the Female Genital Tract Due to a Factor under Ovarian Control. *Infect Immun* 2010;**78**(4):1629–41. Doi: 10.1128/IAI.00996-09.

51 Bandana K., Jashandeep K., Jagdeep K. Phospholipases in Bacterial Virulence and Pathogenesis. *AIBM* 2018;**10**(5). Doi: 10.19080/AIBM.2018.10.555798.

52 Cui Z., Dang G., Song N., Cui Y., Li Z., Zang X., et al. Rv3091, An Extracellular Patatin-Like Phospholipase in *Mycobacterium tuberculosis*, Prolongs Intracellular Survival of Recombinant *Mycolicibacterium smegmatis* by Mediating Phagosomal Escape. *Front Microbiol* 2020;**11**:2204. Doi: 10.3389/fmicb.2020.532371.

53 Kuhle K., Flieger A. *Legionella* phospholipases implicated in virulence. *Curr Top Microbiol Immunol* 2013;**376**:175–209. Doi: 10.1007/82_2013_348.

54 Vila-Farrés X., Parra-Millán R., Sánchez-Encinales V., Varese M., Ayerbe-Algaba R., Bayó N., et al. Combating virulence of Gram-negative bacilli by OmpA inhibition. *Sci Rep* 2017;**7**(1):14683. Doi: 10.1038/s41598-017-14972-y.

55 Morton D. J., Hempel R. J., Seale T. W., Whitby P. W., Stull T. L. A functional *tonB* gene is required for both virulence and competitive fitness in a chinchilla model of *Haemophilus influenzae* otitis media. *BMC Research Notes* 2012;**5**(1):327. Doi: 10.1186/1756-0500-5-327.

56 Confer A. W., Ayalew S. The OmpA family of proteins: roles in bacterial pathogenesis and immunity. *Vet Microbiol* 2013;**163**(3–4):207–22. Doi: 10.1016/j.vetmic.2012.08.019.

57 Ricci V., Piddock L. J. V. Exploiting the Role of TolC in Pathogenicity: Identification of a Bacteriophage for Eradication of *Salmonella* Serovars from Poultry. *Applied and Environmental Microbiology* 2010;**76**(5):1704–6. Doi: 10.1128/AEM.02681-09.

58 Neal-McKinney J. M., Konkel M. E. The *Campylobacter jejuni* CiaC virulence protein is secreted from the flagellum and delivered to the cytosol of host cells. *Front Cell Infect Microbiol* 2012;**2**:31. Doi: 10.3389/fcimb.2012.00031.

59 Haiko J., Westerlund-Wikström B. The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel)* 2013;**2**(4):1242–67. Doi: 10.3390/biology2041242.

60 Guan S-M., Shu L., Fu S-M., Liu B., Xu X-L., Wu J-Z. *Prevotella intermedia* induces matrix metalloproteinase-9 expression in human periodontal ligament cells. *FEMS Microbiology Letters* 2008;**283**(1):47–53. Doi: 10.1111/j.1574-6968.2008.01140.x.

61 Wang Y., Song J., Marquez-Lago T. T., Leier A., Li C., Lithgow T., et al. Knowledge-transfer learning for prediction of matrix metalloprotease substrate-cleavage sites. *Scientific Reports* 2017;**7**(1):5755. Doi: 10.1038/s41598-017-06219-7.

62 Göõz M., Göõz P., Smolka A. J. Epithelial and bacterial metalloproteinases and their inhibitors in *H. pylori* infection of human gastric cells. *Am J Physiol Gastrointest Liver Physiol* 2001;**281**(3):G823-832. Doi: 10.1152/ajpgi.2001.281.3.G823.

63 Charoenlap N., Shen Z., McBee M. E., Muthupalani S., Wogan G. N., Fox J. G., et al. Alkyl Hydroperoxide Reductase Is Required for *Helicobacter cinaedi* Intestinal Colonization and Survival under Oxidative Stress in BALB/c and BALB/c Interleukin-10−/− Mice. *Infect Immun* 2012;**80**(3):921–8. Doi: 10.1128/IAI.05477-11.

64 Oh E., Jeon B. Role of Alkyl Hydroperoxide Reductase (AhpC) in the Biofilm Formation of *Campylobacter jejuni*. *PLoS One* 2014;**9**(1). Doi: 10.1371/journal.pone.0087312.

65 Altermann E., Russell W. M., Azcarate-Peril M. A., Barrangou R., Buck B. L., McAuliffe O., et al. Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc Natl Acad Sci U S A* 2005;**102**(11):3906–12. Doi: 10.1073/pnas.0409188102.

66 Lehnherr H., Maguin E., Jafri S., Yarmolinsky M. B. Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. *J Mol Biol* 1993;**233**(3):414–28. Doi: 10.1006/jmbi.1993.1521.

67 Alauzet C., Marchandin H., Lozniewski A. New insights into *Prevotella* diversity and medical microbiology. *Future Microbiol* 2010;**5**(11):1695–718. Doi: 10.2217/fmb.10.126.

68 Ciodaro F. C., Gazia F., Galletti B., Galletti F. Hyperbaric oxygen therapy in a case of cervical abscess extending to anterior mediastinum, with isolation of *Prevotella corporis*. BMJ Case Reports. Doi: 10.1136/bcr-2019-229873 2019.

69 Finegold S. M., Strong C. A., McTeague M., Marina M. The importance of black-pigmented gram-negative anaerobes in human infections. *FEMS Immunol Med Microbiol* 1993;**6**(2–3):77–82. Doi: 10.1111/j.1574-695X.1993.tb00306.x.

70 Severi E., Hood D. W., Thomas G. H. Sialic acid utilization by bacterial pathogens. *Microbiology (Reading)* 2007;**153**(Pt 9):2817–22. Doi: 10.1099/mic.0.2007/009480-0.

71 Sudhakara P., Sellamuthu I., Aruni A. W. Bacterial sialoglycosidases in Virulence and Pathogenesis. *Pathogens* 2019;**8**(1). Doi: 10.3390/pathogens8010039.

72 Potempa M., Potempa J., Kantyka T., Nguyen K-A., Wawrzonek K., Manandhar S. P., et al. Interpain A, a Cysteine Proteinase from *Prevotella intermedia*, Inhibits Complement by Degrading Complement Factor C3. *PLOS Pathogens* 2009;**5**(2):e1000316. Doi: 10.1371/journal.ppat.1000316.

73 Pierce J. V., Fellows J. D., Anderson D. E., Bernstein H. D. A clostripain-like protease plays a major role in generating the secretome of enterotoxigenic *Bacteroides fragilis*. *Molecular Microbiology* 2021;**115**(2):290–304. Doi: https://doi.org/10.1111/mmi.14616.

74 Roncase E. J., Moon C., Chatterjee S., González-Páez G. E., Craik C. S., O'Donoghue A. J., et al. Substrate Profiling and High Resolution Co-complex Crystal Structure of a Secreted C11 Protease Conserved across Commensal Bacteria. *ACS Chem Biol* 2017;**12**(6):1556–65. Doi: 10.1021/acschembio.7b00143.

75 González-Páez G. E., Roncase E. J., Wolan D. W. X-ray structure of an inactive zymogen clostripain-like protease from *Parabacteroides distasonis*. *Acta Crystallogr D Struct Biol* 2019;**75**(Pt 3):325–32. Doi: 10.1107/S2059798319000809.

76 Hočevar K., Potempa J., Turk B. Host cell-surface proteins as substrates of gingipains, the main proteases of *Porphyromonas gingivalis*. *Biol Chem* 2018;**399**(12):1353–61. Doi: 10.1515/hsz-2018-0215.

77 Bohl T. E., Shi K., Lee J. K., Aihara H. Crystal structure of lipid A disaccharide synthase LpxB from *Escherichia coli*. *Nature Communications* 2018;**9**(1):377. Doi: 10.1038/s41467-017-02712-9.

78 Metzger L. E., Raetz C. R. H. Purification and Characterization of the Lipid A Disaccharide Synthase (LpxB) from *Escherichia coli*: a Peripheral Membrane Protein. *Biochemistry* 2009;**48**(48):11559–71. Doi: 10.1021/bi901750f.

79 Trent M. S., Ribeiro A. A., Lin S., Cotter R. J., Raetz C. R. H. An Inner Membrane Enzyme in *Salmonella* and *Escherichia coli* That Transfers 4-Amino-4-deoxy-l-arabinose to

Lipid A: introduction in polymyxin-resistant mutants and role of a novel lipid-linked donor. *Journal of Biological Chemistry* 2001;**276**(46):43122–31. Doi: 10.1074/jbc.M106961200.

80 Tavares-Carreón F., Patel K. B., Valvano M. A. *Burkholderia cenocepacia* and *Salmonella enterica* ArnT proteins that transfer 4-amino-4-deoxy-l-arabinose to lipopolysaccharide share membrane topology and functional amino acids. *Sci Rep* 2015;**5**:10773. Doi: 10.1038/srep10773.

81 Fillol-Salom A., Martínez-Rubio R., Abdulrahman R. F., Chen J., Davies R., Penadés J. R. Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J* 2018;**12**(9):2114–28. Doi: 10.1038/s41396-018-0156-3.

82 Ben Fekih I., Zhang C., Li Y. P., Zhao Y., Alwathnani H. A., Saquib Q., et al. Distribution of Arsenic Resistance Genes in Prokaryotes. *Front Microbiol* 2018;**9**. Doi: 10.3389/fmicb.2018.02473.

83 Gao Y-R., Li D-F., Fleming J., Zhou Y-F., Liu Y., Deng J-Y., et al. Structural analysis of the regulatory mechanism of MarR protein Rv2887 in *M. tuberculosis*. *Scientific Reports* 2017;**7**(1):6471. Doi: 10.1038/s41598-017-01705-4.

84 Grove A. Regulation of Metabolic Pathways by MarR Family Transcription Factors. *Comput Struct Biotechnol J* 2017;**15**:366–71. Doi: 10.1016/j.csbj.2017.06.001.

85 Veloo A. C. M., Baas W. H., Haan F. J., Coco J., Rossen J. W. Prevalence of antimicrobial resistance genes in *Bacteroides* spp. and *Prevotella* spp. Dutch clinical isolates. *Clinical Microbiology and Infection* 2019;**25**(9):1156.e9-1156.e13. Doi: 10.1016/j.cmi.2019.02.017.

86 Sherrard L. J., Graham K. A., McGrath S. J., McIlreavey L., Hatch J., Muhlebach M. S., et al. Antibiotic resistance in *Prevotella* species isolated from patients with cystic fibrosis. *J Antimicrob Chemother* 2013;**68**(10):2369–74. Doi: 10.1093/jac/dkt191.

87 Cobo F., Rodríguez-Granger J., Pérez-Zapata I., Sampedro A., Aliaga L., Navarro-Marí J. M. Antimicrobial susceptibility and clinical findings of significant anaerobic bacteria in southern Spain. *Anaerobe* 2019;**59**:49–53. Doi: 10.1016/j.anaerobe.2019.05.007.

88 Whittle G., Shoemaker N. B., Salyers A. A. The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cell Mol Life Sci* 2002;**59**(12):2044–54. Doi: 10.1007/s000180200004.

89 Juhas M., van der Meer J. R., Gaillard M., Harding R. M, Hood D. W, Crook D. W. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 2009;**33**(2):376–93. Doi: 10.1111/j.1574-6976.2008.00136.x.

90 Tribble G. D., Garza J. J., Yeung V. L., Rigney T. W., Dao D-H. V., Rodrigues P. H., et al. Genetic Analysis of Mobile *tetQ* Elements in Oral *Prevotella* Species. *Anaerobe* 2010;**16**(6):604–9. Doi: 10.1016/j.anaerobe.2010.08.008.

91 Burrus V., Pavlovic G., Decaris B., Guédon G.. Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 2002;**46**(3):601–10. Doi: 10.1046/j.1365-2958.2002.03191.x.

92 Smith C. J., Parker A. C. Identification of a circular intermediate in the transfer and transposition of Tn4555, a mobilizable transposon from *Bacteroides* spp. *J Bacteriol* 1993;**175**(9):2682–91. Doi: 10.1128/jb.175.9.2682-2691.1993.

93 Parker A. C., Smith C. J. A Multicomponent System Is Required for Tetracycline-Induced Excision of Tn4555. *Journal of Bacteriology* 2004;**186**(2):438–44. Doi: 10.1128/JB.186.2.438-444.2004.

94 Sherrard L. J., Schaible B., Graham K. A., McGrath S. J., McIlreavey L., Hatch J., et al. Mechanisms of reduced susceptibility and genotypic prediction of antibiotic resistance in *Prevotella* isolated from cystic fibrosis (CF) and non-CF patients. *J Antimicrob Chemother* 2014;**69**(10):2690–8. Doi: 10.1093/jac/dku192.

Table 1. *P. corporis* and *P. brunnea* genomes used in comparative genomic analysis.

***Prevotella* species**	**Strain**	**Source**	**Location**	**Host status**	**Predicted human pathogen**	**Genome size (Mb)**	**Contigs ≥ 500 bp**	**N50**	**CDS**	**GC (%)**	**WGS Project accessions**
Prevotella corporis	DSM_18810^T	cervical swab	UK	D	yes	2.80	84	60,710	2415	44.12	AUME00000000
	MJR7716	vagina	USA	UN	yes	2.86	177	67,706	2483	43.96	LRQG00000000
	c24Ua_14AN	urine	Portugal	H	no	2.95	105	104,027	2641	43.8	JABUXP000000000
Prevotella brunnea	A2672^T	wound swab	Germany	D	no	2.59	114	41,853	2234	44.02	SDIK00000000
	Marseille P4334	stool	France	UN	yes	3.04	24	1,147,385	2623	43.74	UWTY00000000
	c17Ua_125AN	urine	Portugal	D	yes	2.79	118	41,787	2517	43.72	VFFG00000000

Legend: D, diseased; H, healthy; UN, Unknown.

Table 2. Predicted pathogenic proteins and pathogenicity score for the *P. corporis* and *P. brunnea* strains according to PathogenFinder tool.

Species	Strain	Number of matched pathogenic families	Identification of matched pathogenic protein	Predicted human pathogen
P. corporis	DSM 18810^T	17	TraE, TraF, TraH, TraM, TraN, TraO, TraP, TraQ, *rteB* gene product, tyrosine type site-specific recombinase, putative DNA methylase, hypothetical proteins	Yes
	MJR7716	20	TraE, TraF, TraG, TraI, TraJ, TraK, TraM, TraN, TraO, TraQ, conjugative transposon protein TraE, *rteB* gene product, tyrosine type site-specific recombinase, Exc protein, putative mobilization protein, hypothetical proteins	Yes
	c24Ua_14AN	5	TraG family protein, fumarate hydratase class I, ribosomal protein L15, ribosomal protein S14, conserved hypothetical protein	No
P. brunnea	A2672^T	3	TraK, ribosomal protein S14, *rteB* gene product	No
	Marseille P4334	24	TraE, TraF, TraG, TraH, TraI, TraJ, TraK, TraM, TraN, TraO, TraP, TraQ, *rteB* gene product, tyrosine type site-specific recombinase, Exc protein, putative mobilization protein, hypothetical proteins	Yes
	c17Ua_125AN	16	TraE, TraF, TraG, TraJ, TraK, TraN, *rteB* gene product, tyrosine type site-specific recombinase, Exc protein, putative mobilization protein, putative cytochrome c biogenesis protein, hypothetical proteins	Yes

Table 3. List of COG functions enriched per each species (only hits with p value < 0.05 are shown).

***Prevotella* species**	**COG FUNCTION**	**COG category**
Prevotella corporis	Predicted transcriptional regulator YdeE, contains AraC-type DNA-binding domain	K
	Aldo/keto reductase, related to diketogulonate reductase	Q
	Uncharacterized membrane protein YhaH, DUF805 family	S
	Flagellar hook-associated protein FlgK	N
	RNA recognition motif (RRM) domain	J
	Type II restriction/modification system, DNA methylase subunit YeeA	V
	Glycerol uptake facilitator and related aquaporins (Major Intrinsic Protein Family)	G
	Alkylated DNA nucleotide flippase Atl1, participates in nucleotide excision repair, Ada-like DNA-binding domain	K
	O-Methyltransferase involved in polyketide biosynthesis	Q
	Polysaccharide pyruvyl transferase family protein WcaK	M
	Superfamily II DNA or RNA helicase	L
	Alkyl hydroperoxide reductase subunit AhpC (peroxiredoxin)	V
	Acyl dehydratase	I
	Restriction endonuclease	V
	Wyosine [tRNA(Phe)-imidazoG37] synthetase, radical SAM superfamily	J
	Prolyl oligopeptidase PreP, S9A serine peptidase family	E
	Beta-galactosidase, beta subunit	G
	TPP-dependent indolepyruvate ferredoxin oxidoreductase, alpha subunit	C
	Divalent metal cation (Fe/Co/Zn/Cd) transporter	P
	L-lactate utilization protein LutB, contains a ferredoxin-type domain	C
	DNA-binding transcriptional regulator, XRE family	K
	Restriction endonuclease S subunit	V
	Peptide methionine sulfoxide reductase MsrA	O
	Metal-dependent amidase/aminoacylase/carboxypeptidase	R
	Predicted acetyltransferase, GNAT superfamily	R
	Predicted flavoprotein YhiN	R
	Uncharacterized protein, contains FMN-binding domain	R
	ATP-dependent Clp protease adapter protein ClpS	O
	Indole-3-glycerol phosphate synthase	E
	Co/Zn/Cd efflux system component	P
	Predicted metalloprotease	R
	ABC-type sugar transport system, ATPase component	G
	Aspartate/methionine/tyrosine aminotransferase	E
	Polyferredoxin	C
	Coenzyme F420-reducing hydrogenase, beta subunit	C
	Nicotinamidase-related amidase	H
	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)	T
	L-lactate utilization protein LutC, contains LUD domain	C
	O6-methylguanine-DNA--protein-cysteine methyltransferase	L

	D-alanyl-D-alanine dipeptidase	M
	Xanthine/uracil/vitamin C permease, AzgA family	F
	Uncharacterized membrane protein YccF, DUF307 family	S
	Alcohol dehydrogenase, class IV	C
	Predicted ATPase	R
	Leu/Phe-tRNA-protein transferase	O
	Alkyl hydroperoxide reductase subunit AhpF	V
	Prophage maintenance system killer protein	X
	Pyrroline-5-carboxylate reductase	E
	Type I site-specific restriction endonuclease, part of a restriction-modification system	V
	Transcriptional regulator of competence genes, TfoX/Sxy family	K
	Fe-S oxidoreductase	C
Prevotella brunnea	Preprotein translocase subunit SecE	U
	Uncharacterized N-terminal domain of lipid-A-disaccharide synthase	R
	O-antigen ligase	M
	DNA-binding transcriptional regulator, CsgD family	K
	DNA anti-recombination protein (rearrangement mutator) RmuC	L
	N-acetylmuramic acid 6-phosphate (MurNAc-6-P) etherase	M
	Phage-related protein	X
	Uncharacterized conserved protein, DUF488 family	S
	Ca2+-binding protein, RTX toxin-related	Q
	Fido, protein-threonine AMPylation domain	T
	Tryptophanase	E
	Fluoride ion exporter CrcB/FEX, affects chromosome condensation	P
	Uncharacterized protein YaaW, UPF0174 family	S
	CMP-N-acetylneuraminic acid synthetase	M
	Sialic acid synthase SpsE, contains C-terminal SAF domain	M
	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	G
	Anion-transporting ATPase, ArsA/GET3 family	P
	4-amino-4-deoxy-L-arabinose transferase or related glycosyltransferase of PMT family	M
	Periplasmic ferric-dicitrate binding protein FerR, regulates iron transport through sigma-19	P

The COG functions belong to corresponding COG categories: C - Energy production and conversion; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - Translation, ribosomal structure and biogenesis; K - Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defense mechanisms; X - mobilome: phages and transposons.

Table 4. Putative virulence-associated COG functions identified in pangenome of *P. corporis* and *P. brunnea*. A list presents species and/or strains specific genes that may contribute to bacterial adherence, host colonization and invasion. Genes/ gene clusters detected in particular strain are marked with x.

		P. corporis			***P. brunnea***		
COG function	**COG**	**DSM_18810^T**	**MJR7716**	**c24Ua_14AN**	**A2672^T**	**Marseille_P4334**	**c17Ua_125AN**
Cell wall/membrane/envelope							
Predicted phospholipase, patatin/cPLA2 family	COG4667, COG0729, COG1752	x	x	x	x	x	x
Opacity protein and related surface antigens	COG3637	x	x	x	x	x	x
Outer membrane protein OmpA	COG2885	x	x	x	x	x	x
Periplasmic protein TonB	COG0810	x	x	x	x	x	x
Outer membrane protein TolC	COG1538	x	x	x	x	x	x
CMP-N-acetylneuraminic acid synthetase	COG1083				x	x	x
Sialic acid synthase SpsE	COG2089				x	x	x
4-amino-4-deoxy-L-arabinose transferase	COG1807				x	x	x
Defense mechanisms							
ABC-type bacteriocin/lantibiotic exporters	COG2274	x	x	x	x	x	x
Prophage maintenance system killer protein	COG3654	x	x	x			
Txe-Axe toxin-antitoxin module, Txe/YoeB family	COG4115			x			
Lactobacillus related ABC-type exporters	COG2274						x
Miscellaneous							
Flagellar hook-associated protein FlgK	COG1256	x	x	x			
Ca2+-binding protein, RTX toxin-related	COG2931				x	x	x
Uncharacterized N-terminal domain of lipid-A-disaccharide synthase	COG3952				x	x	x
Lysozyme family protein	COG3926	x					

Figure 1. Pangenome of *Prevotella* strains belonging to *Prevotella corporis* (green inner rings) and *Prevotella brunnea* (pink inner rings) species, generated by anvi'o v6.2. Plot is presenting presence/absence of gene clusters. Samples are ordered based on gene cluster frequencies. Number of contributing genomes is presented in an additional ring. Outside blue-yellow ring represents COGs functionality annotation. External thin rings represent particular gene collections: red - core, orange - accessory, that contain singletons (violet), gene clusters specific to P. corporis (green) and gene clusters specific to P. brunnea (pink). Additional information such as host status, human pathogen prediction, clade or genomes related information such as total length, GC content, completion, number of genes, singleton gene clusters and number of gene clusters are represented by bars in the right top of the figure.

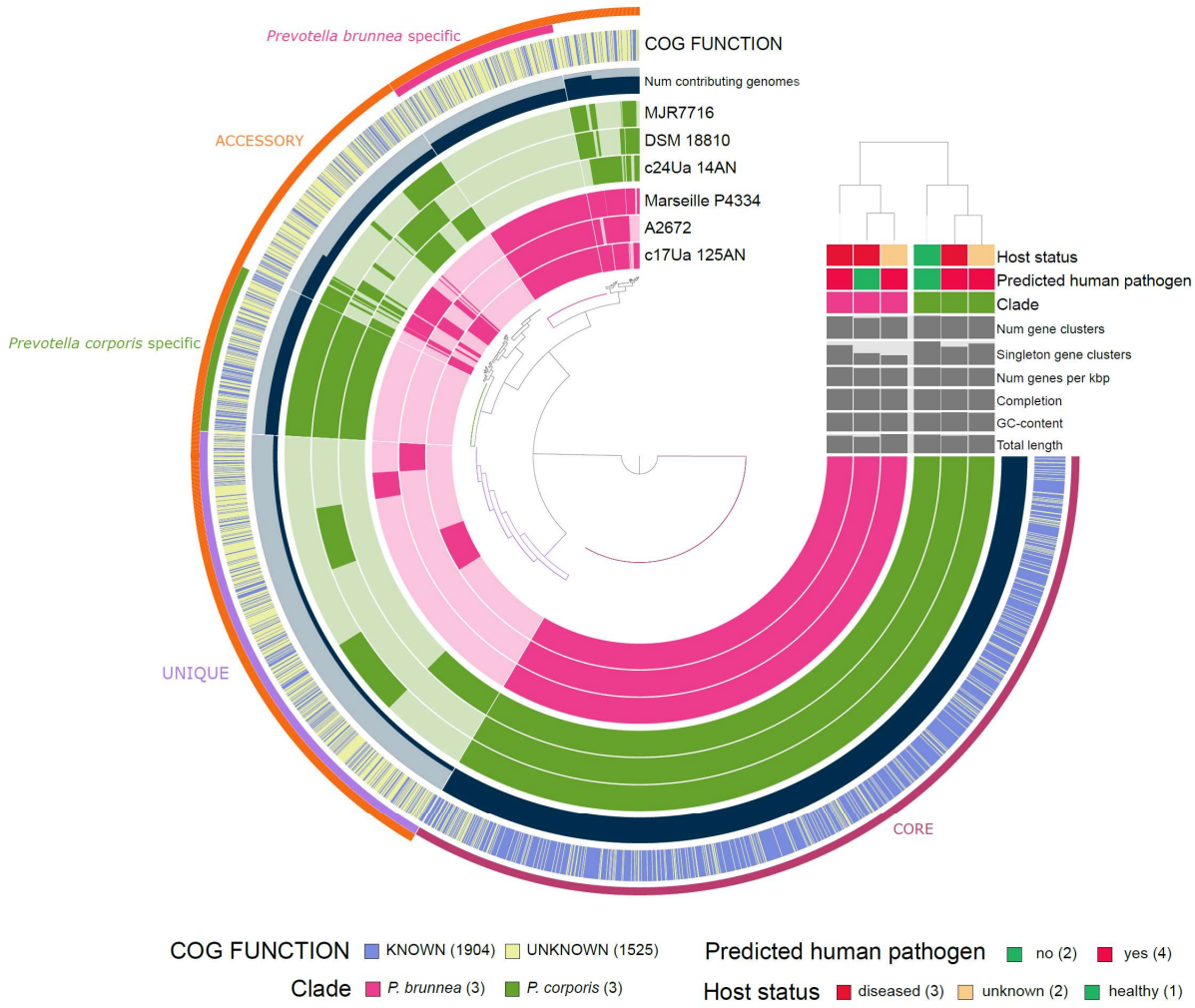
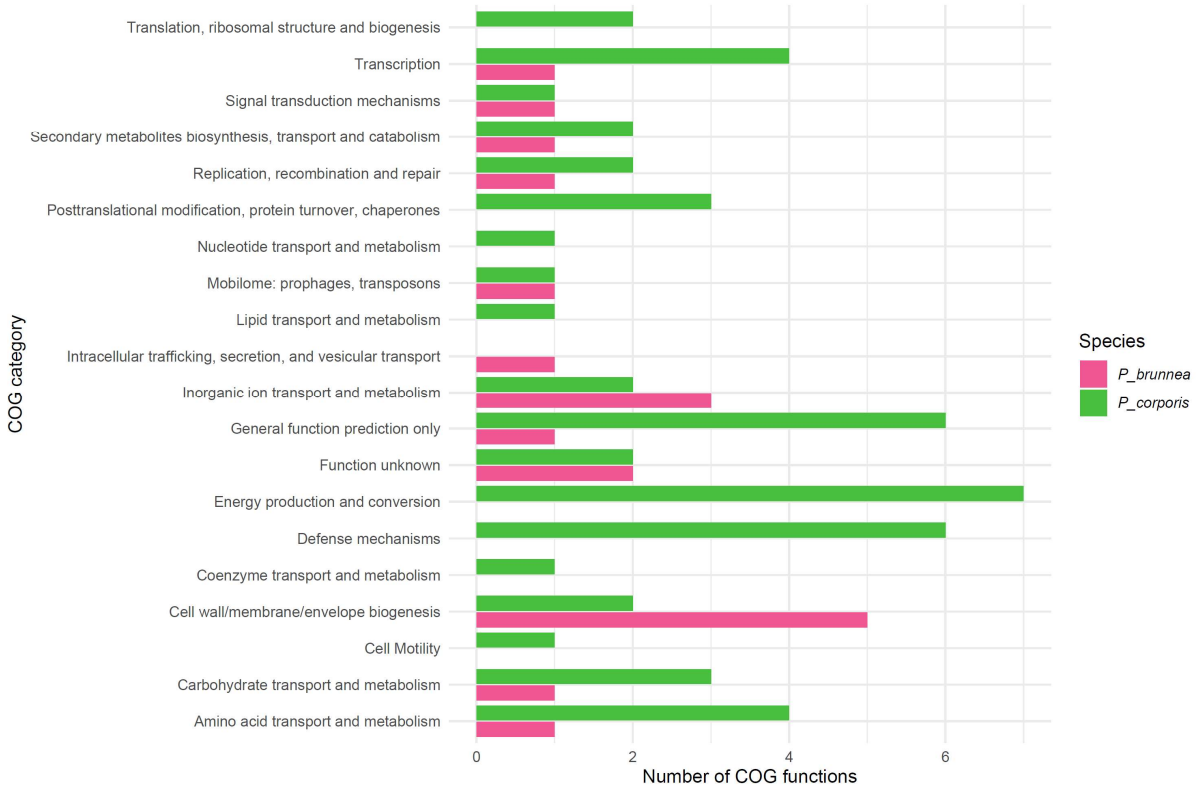


Figure 2. Distribution of COG enriched functions for *P. corporis* and *P. brunnea* grouped by COG categories.



Additional material associated with the manuscript entitled “Pangenome analysis of *Prevotella corporis* and *Prevotella brunnea* reveals features associated with their adaptation and virulence” is available through the following links:

Supplementary Tables S1-S7
https://docs.google.com/spreadsheets/d/1AdzDCKH8KccKUv5ZzMb49j7B40FzKuIE/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

The darkest place is under the candlestick - healthy urogenital tract as a source of worldwide disseminated Extraintestinal Pathogenic *Escherichia coli* lineages

Magdalena Ksiezarek[1,2], Angela Novais[1,2], *Luisa Peixe[1,2]

[1]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[2]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

Microorganisms 2022, 10, 27.

microorganisms

MDPI

Article

The Darkest Place Is under the Candlestick-Healthy Urogenital Tract as a Source of Worldwide Disseminated Extraintestinal Pathogenic *Escherichia coli* Lineages

Magdalena Ksiezarek [1,2], Ângela Novais [1,2] and Luísa Peixe [1,2,*]

[1] UCIBIO–Applied Molecular Biosciences Unit, REQUIMTE, Laboratory of Microbiology, Department of Biological Sciences, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; mag.ksiezarek@gmail.com (M.K.); angelasilvanovais@gmail.com (Â.N.)

[2] Associate Laboratory i4HB-Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

* Correspondence: lpeixe@ff.up.pt; Tel.: +351-220-428-580

Abstract: Since the discovery of the urinary microbiome, including the identification of *Escherichia coli* in healthy hosts, its involvement in UTI development has been a subject of high interest. We explored the population diversity and antimicrobial resistance of *E. coli* (n = 22) in the urogenital microbiome of ten asymptomatic women (representing 50% of the sample tested). We evaluated their genomic relationship with extraintestinal pathogenic *E. coli* (ExPEC) strains from healthy and diseased hosts, including the ST131 lineage. *E. coli* prevalence was higher in vaginal samples than in urine samples, and occasionally different lineages were observed in the same individual. Furthermore, B2 was the most frequent phylogenetic group, with the most strains classified as ExPEC. Resistance to antibiotics of therapeutic relevance (e.g., amoxicillin-clavulanate conferred by bla_{TEM-30}) was observed in ExPEC widespread lineages sequence types (ST) 127, ST131, and ST73 and ST95 clonal complexes. Phylogenomics of ST131 and other ExPEC lineages revealed close relatedness with strains from gastrointestinal tract and diseased host. These findings demonstrate that healthy urogenital microbiome is a source of potentially pathogenic and antibiotic resistant *E. coli* strains, including those causing UTI, e.g., ST131. Importantly, diverse *E. coli* lineages can be observed per individual and urogenital sample type which is relevant for future studies screening for this uropathogen.

Keywords: vaginal microbiome; voided urine; urinary tract microbiome; uropathogens; ExPEC; ST131

Citation: Ksiezarek, M.; Novais, Â.; Peixe, L. The Darkest Place Is under the Candlestick-Healthy Urogenital Tract as a Source of Worldwide Disseminated Extraintestinal Pathogenic *Escherichia coli* Lineages. *Microorganisms* **2022**, *10*, 27. https://doi.org/10.3390/microorganisms10010027

Academic Editor: Roberto Di Marco

Received: 28 November 2021
Accepted: 21 December 2021
Published: 24 December 2021

1. Introduction

Urinary tract infections (UTIs) are reported to be one of the most common infections worldwide, and occur more frequently in women than in men due to anatomical differences, and its incidence increases with age or sexual activity [1–4]. Although several bacterial species are reported as causative agents of UTIs (e.g., *Escherichia coli*, *Klebsiella* sp., *Proteus* sp., *Staphylococcus saprophyticus*), *E. coli* is responsible for the majority (up to 75%) of both uncomplicated and complicated UTIs [2,5–7].

Extraintestinal pathogenic *E. coli* (ExPEC) strains are most frequently responsible for these infections, with this pathotype associated with genetic characteristics that seem to favor the pathogenicity of particular strains, but also *E. coli* persistence in the urinary tract [8–10]. Indeed, most UTIs are caused by a subset of ExPEC strains [e.g., sequence type (ST) 131, ST95, ST69, ST73, ST127, ST12] that are highly disseminated across different continents and populations and represent great clinical challenges not only because they are often resistant to the main therapeutic choices but also because of their extended reservoirs (human and animal gastrointestinal tract and the environment) [11–16].

To date, multiple studies have evaluated putative human-related reservoirs of ExPEC strains causing UTI. The human gut or vaginal colonization with certain strains of *E. coli* have been considered risk factors for developing UTI [17–21]. Additionally, the existence

of intracellular bacterial communities or quiescent intracellular reservoirs that may silently persist within the epithelial cells of the urogenital tract also contribute to recurrent UTI (rUTI) development [7,22].

On the other hand, the collapse of the urine sterility dogma after the identification of the female urogenital microbiome (FUM) [23] and the urinary microbiome [24], raised the possibility of the urinary/urogenital tract being a reservoir of UTI pathogens [25,26]. In fact, studies analyzing FUM composition unveiled the presence of uropathogenic bacterial species, including *E. coli*, in the healthy urinary tract [27,28]. Moreover, a high inter-individual variability in the relative abundance of *E. coli* (varying from 10 to $>10^5$ CFU/mL) was reported in the healthy population [27,29,30]. Thus, since *E. coli* is the most frequent pathogen in causing UTI and because it is part of the healthy FUM, occasionally in high relative abundance, it is plausible that the urinary tract itself is a source of ExPEC lineages causing UTIs.

Until now, FUM characterization has focused mostly on the taxonomic profiling at the genus and species level, whereas characterization at the strain level is scarce. Thomas-White and co-workers [28] revealed the urinary and vaginal interconnection for potentially pathogenic *E. coli*. However, no data on virulence potential or antimicrobial resistance profiles of *E. coli* strains were provided. Strain level characterization from healthy hosts was only performed recently in a study from Garretto et al., that aimed at establishing an association between the genomic content of urinary strains and the presence of lower urinary tract symptoms [31]. The authors did not find any specific signatures on the microbiome composition, gene content or *E. coli* abundance that could predict UTI status, and only six urinary strains were obtained from asymptomatic women [31]. Thus, the nature and diversity of *E. coli* strains from the whole urogenital tract (urinary and vaginal) of healthy women are still practically unexplored.

In our study, we aimed to characterize the prevalence, strain diversity and genetic features (antimicrobial resistance and virulence) of *E. coli* isolated from both urine and vaginal samples of healthy asymptomatic women. Furthermore, we provided an in-depth phylogenomic analysis of the phylogenetic group B2 ExPEC strains.

2. Materials and Methods

2.1. Study Design and Sample Description

E. coli strains were isolated in the framework of a study on healthy Female Urogenital Microbiome conducted at Faculty of Pharmacy, University of Porto, Porto, Portugal (2016–2019) [30]. All women provided informed written consent for participation in the study. The study was developed according to the Helsinki Declaration principles and the protocol was submitted and approved by the Ethical Commission of Faculty of Pharmacy, University of Porto (Ethical Committee approval number 32-09-2017). This study included isolates obtained from 20 healthy women (aged 24–57) who were asymptomatic for any urinary disorder and that declared to be in good health condition, from which urogenital samples (30 voided urine and 30 vaginal swabs) were analyzed (10 women provided samples twice [30]). A list of donors, samples and isolates is presented in Supplementary Table S1.

2.2. Identification of E. coli Isolates

All 60 samples were analyzed based on a previously published extended culturomic protocol [30]. When colonies with a morphology compatible with *E. coli* were observed, an average of 3 isolates/samples (up to 7 colonies/sample; proportionally to the bacterial load) were isolated.

Isolates were preliminarily identified by MALDI-TOF MS (VITEK MS, bioMérieux, France), and further confirmed by the amplification of the *malB* gene [32]. A total of 50 *E. coli* isolates were identified and stored, 22 of which originated from the urine of healthy donors (1–7/sample) and the remaining 28 were detected from vaginal swabs of healthy donors (1–5/sample) (Supplementary Table S1).

2.3. Characterization of E. coli Population

The identification of *E. coli* phylogenetic groups and virulence gene profiling were performed on all *E. coli* isolates so as to discard overrepresented genotypes per sample type (urinary, vaginal). *E. coli* phylogenetic groups were identified according to the method proposed by Clermont et al. [33]. The presence of 41 virulence factors (VFs) associated with ExPEC pathotype including adhesins, toxins, siderophores, capsule type, protectins, invasins and miscellaneous was screened by PCR (Supplementary Table S2). Detection of *hlyA* operon duplication was performed by PCR, as previously described [34] (Supplementary Table S2). The strains were classified as putative ExPEC if fulfilled the following criteria: presence of >2 amongst *papAH* and/or *papC*, *sfa/focDE*, *afa/draBC*, *kpsM* II and *iutA* [35].

Based on this strain profiling strategy, we selected 10 representative urinary and 12 representative vaginal *E. coli* strains, which are described in detail in Table 1. To characterize the representative strains of the urogenital population, pulsed-field gel electrophoresis (PFGE) was used to identify identical strains detected in both urine and vaginal samples from the same donor. The standard protocol for PFGE adapted from Gautom et al. [36] was used, with 1.6% chromosomal grade agarose (SeaKem® Gold Agarose) for plugs preparation, 1.2% gel with pulse-field certified agarose (SeaKem® Gold Agarose) for PFGE run, genomic DNA digested with 20U *Xba* I and CHEF DNA Size Standard–Lambda Ladder, which was used as a size marker. The running conditions included an initial switch time of 2.2 s, a final switch time of 63.8 s and a run time of 19 h with 6 V/cm and an angle of 120.

This procedure narrowed down our sample to 16 unique strains representing unique genotypic profiles from the urogenital *E. coli* population of 10 healthy women. In vitro antibiotic susceptibility testing was evaluated using a disk diffusion for 21 antibiotics (diverse β-lactams, fluoroquinolones, aminoglycosides, tetracyclines, fosfomycin, nitrofurantoin, trimethoprim, sulfamethoxazole-trimethoprim) according to EUCAST guidelines (www.eucast.org, accessed on 12 December 2019).

2.4. Whole Genome Sequencing

Considering the high proportion of B2 *E. coli* strains (50%, 8/16) and the identification of typical ExPEC pathotype features amongst them, 8 B2 strains were characterized using whole genome sequencing (WGS) and comparative genomics approaches, as described below.

Genomic DNA (n = 4 urine, n = 4 vaginal) was extracted (Wizard Genomic DNA Purification Kit, PROMEGA) and sequenced by Illumina NovaSeq 2 × 150 nt. Reads were trimmed by Trimmomatic [37] version 0.39 and the quality of reads was checked by FastQC version 0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, accessed on 2 February 2021). De novo assembly was performed by SPAdes [38] version 3.13.0 and the quality of the assembly was checked by Quast [39] version 5.0.2. Annotations of the draft genomes were provided by the NCBI Prokaryotic Genome Annotation Pipeline, and additionally by Prokka [40] version 1.14.6.

The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the BioProject accession number PRJNA548360. Accession numbers to WGS submission and the BioSample for each strain are available in Supplementary Table S3.

Characterization of strains including serotype, *E. coli* phylogenetic group and multilocus sequence typing (MLST, Achtman scheme) was assessed using available in silico tools (http://www.genomicepidemiology.org/, accessed on 4 February 2021; http://clermontyping.iame-research.center/, accessed on 4 February 2021). The virulence profile of each strain previously defined by PCR screening was verified and extended to 49 putative virulence genes using in-house database (https://github.com/magksi/E.coli_VF_characterization, released on 20 December 2021) and ncbi-blast-2.8.1+ package (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/, accessed on 7 May 2019). Antimicrobial resistance genes were identified using ResFinder 4.1 (https://cge.cbs.dtu.dk/services/ResFinder/, accessed on 4 February 2021), plasmid replicon sequences by Plas-

midFinder 2.1 (https://cge.cbs.dtu.dk/services/PlasmidFinder/, accessed on 4 February 2021) and *fimH* typing using FimTyper 1.0 (https://cge.cbs.dtu.dk/services/FimTyper/, accessed on 4 February 2021).

Table 1. Phylogenetic origin and diversity of urinary (*n* = 10) and vaginal (*n* = 12) *E. coli* strains isolated from healthy women.

Donor	Origin	Strains	Phylogenetic Group	ExPEC	PFGE	WGS	MLST	Serotype
1	U	c1Ub_48	F	+	EC1	–	–	–
	VS	c1VSb_14	F	+	EC1	–	–	–
	U	c1Ub_56	B2	+	–	+	ST452	O81:H27
3	U	c3Ub_1	B2	+	EC2	+	ST681	O8:H10
	VS	c3VSb_22	B2	+	EC2	–	–	–
7	VS	c7VSa_62	D	–	–	–	–	–
10	U	c10Ua_105	F	+	EC3	–	–	–
	VS	c10VSa_39	F	+	EC3	–	–	–
11 *	U	c11Ua_88	D	–	–	–	–	–
	U	c11Ub_17	B2	+	EC4	+	ST1154 (ST73 complex)	O2:H1
	VS	c11VSb_7	B2	+	EC4	–	–	–
	VS	c11VSb_12	B2	+	EC5	+	ST998	O2:H6
12	VS	c12VSb_42	C	–	–	–	–	–
15	U	c15Ub_26	A	–	EC6	–	–	–
	VS	c15VSb_20	A	–	EC6	–	–	–
26	U	c26Ub_7	B2	+	EC7	+	ST131	O25:H4
	VS	c26VSb_8	B2	+	EC7	–	–	–
29 *	U	c29Ub_57	D	–	–	–	–	–
	VS	c29VSb_15	B2	+	EC8	+	ST140 (ST95 complex)	O2:H5
	VS	c29VSa_23	B2	–	EC9	+	ST569	O46:H31
31	U	c31Ua_56	E	–	–	–	–	–
	VS	c31VSa_9	B2	+	–	+	ST127	O6:H31

U-urine; VS- vaginal swab; ExPEC-extraintestinal pathogenic *E. coli*; PFGE-pulsed-field gel electrophoresis; WGS-whole genome sequencing; MLST-multilocus sequence typing. *-donors from whom *E. coli* was detected at 2 different sampling times; 'a' and 'b' in strain name denotes the sampling time.

2.5. Comparative Genomics of B2 E. coli Strains

We extracted 19,668 *E. coli* assemblies from the NCBI Assembly database (assessed at 14 July 2020), including 1266 complete genomes and 18,402 draft genomes (7432 scaffolds and 10,970 contigs). To enlarge the sample of urinary isolates from both healthy and diseased women, we included 66 *E. coli* assemblies of urinary origin, published in September 2020 by Garretto et al. [31] and 3 B2 *E. coli* genomes recovered from the urine of 3 women with a history of recurrent UTI from our collection (Supplementary Table S3).

A total of 19,737 assemblies were subjected to mlst 2.19.0 pipeline [41] (https://github.com/tseemann/mlst, accessed on 10 February 2021) to access STs of public genomes. Genomes (*n* = 1084) that represented the same STs as those detected in our collection were extracted. Only the genomes from humans and those with available metadata on host health status and isolation source were considered. Following these criteria, we grouped isolates according to host health status (healthy-H or diseased-D) and origin (urogenital

tract, gastrointestinal tract or other isolation source). Additionally, complete genomes of well-known reference strains for certain ST were included in the analysis.

2.6. Single Nucleotide Polymorphisms Analysis

The genomes were subjected to a snippy (version 4.4.0, https://github.com/tseemann/snippy, accessed on 12 February 2021), using snippy-multi and mapping to appropriate reference genome (stated in description of the figures). A core genome and whole genome SNPs alignments generated by snippy were used for Gubbins pipeline (Genealogies Unbiased By recomBinations In Nucleotide Sequences) [42] version 2.4.1 to remove recombinant regions. The SNP matrices from core genome and whole genome alignment for comparison were generated using snp-dist version 0.7.0 (https://github.com/tseemann/snp-dists, accessed on 12 February 2021). The resulting SNP-based alignments were used to reconstruct the phylogeny for each ST using RAxML, with the appropriate reference. The obtained phylogenetic trees were represented using the Interactive Tree of Life (iTOL, https://itol.embl.de, accessed on 15 February 2021).

2.7. Accessory Genome and Ordination-Based Analysis

A total of 520 ST131 *E. coli* genomes were annotated with prokka version 1.14.6, and a pangenome analysis was performed with Roary [43] version 3.13.0 with the default criteria for proteins identity of 95%. The accessory genome included genes that were present in <99% of the genomes. The data were reshaped and cleaned with tidyverse [44] package version 1.3.0. Further analyses, including non-metric multidimensional scaling (NMDS), and a calculation of the stress value was performed in R [45] version 4.0.3 with package vegan [46] version 2.5.7 and ggplot2 [47] version 3.3.5.

2.8. Statistical Analysis

Continuous variables were interpreted based on the descriptive statistics. Welch Two Sample *t*-test in R [45] version 3.6.2 was used to access significance of phylogenetic groups and VF distribution between the different sample types. Heatmap and clustering of the isolates based on VFs presence/absence was performed with Pheatmap package version 1.0.12 (https://github.com/raivokolde/pheatmap, accessed on 1 March 2021), with clustering based on Euclidean distance.

3. Results

3.1. Frequency and Diversity of E. coli in Urinary and Vaginal Microbiome

E. coli was identified in 30% of the urine samples (9/30) and 37% of the vaginal samples (11/30) (Supplementary Table S1). The bacterial load varied from 1×10 CFU/mL – 1×10^7 CFU/mL, with a median of 1×10 CFU/mL. Overall, *E. coli* was identified in 50% (10/20) of the total cohort of asymptomatic women.

A set of 10 urinary and 12 vaginal representative *E. coli* strains is presented in Table 1. All but one of the B2 strains were classified as ExPEC. Furthermore, two isolates belonging to phylogenetic group F (c1Ub_48 and c10Ua_105) were also considered ExPEC (Table 1). Interestingly, two donors carried different *E. coli* strains in either urine or vaginal samples and two donors had different *E. coli* strains in the same sample type.

Although there was no significant association between the phylogenetic groups' distribution or VF profiles and the origin (urine or vaginal) of the sample, *E. coli* phylogenetic group B2 were more frequently detected in vagina (58%) compared to the urinary isolates (40%). Other phylogenetic groups (F, D and A) were variably found in either urine and vagina samples, while C and E were detected in only one sample type (Tables 1 and 2).

Table 2. Detection and prevalence (%) of phylogenetic groups and 41 putative virulence genes among 22 unique *E. coli* strains per each sample type.

	Total (*n* = 22)	**Urinary Isolates (*n* = 10)**	**Vaginal Isolates (*n* = 12)**
Phylogenetic group			
A	2 (9)	1 (10)	1 (8)
B2	11 (50)	4 (40)	7 (58)
C	1 (5)	0	1 (8)
D	3 (14)	2 (20)	1 (8)
E	1 (5)	1 (10)	0
F	4 (18)	2 (20)	2 (17)
Adhesins			
fimH	20 (91)	9 (90)	11 (92)
papAH	3 (14)	2 (20)	1 (8)
papC	7 (32)	3 (30)	4 (33)
papEF	7 (32)	3 (30)	4 (33)
papG II, III	1 (5)	0	1 (8)
papG allele III	3 (14)	1 (10)	2 (17)
sfa/focD	6 (27)	2 (20)	4 (33)
sfaS	1 (5)	0	1 (8)
focG	6 (27)	2 (20)	4 (33)
afa/draBC	3 (14)	2 (20)	1 (8)
iha	7 (32)	4 (40)	3 (25)
matB	18 (82)	8 (80)	10 (83)
yfcV	15 (68)	6 (60)	9 (75)
Toxins			
hlyA	3 (14)	1 (10)	2 (17)
cnf1	3 (14)	1 (10)	2 (17)
sat	9 (41)	4 (40)	5 (42)
tsh	8 (36)	3 (30)	5 (42)
vat	5 (23)	1 (10)	4 (33)
tosA	4 (18)	2 (20)	2 (17)
Siderophores			
fyuA	14 (64)	8 (80)	6 (50)
iutA	8 (36)	5 (50)	3 (25)
iroN	8 (36)	3 (30)	5 (42)
ireA	3 (14)	1 (10)	2 (17)
Capsule			
kpsMT II	12 (55)	5 (50)	7 (58)
kpsMT III	2 (9)	1 (10)	1 (8)
kpsMT K1	5 (23)	1 (10)	4 (33)
kpsMT K5	6 (27)	3 (30)	3 (25)

Table 2. *Cont.*

	Total (*n* = 22)	Urinary Isolates (*n* = 10)	Vaginal Isolates (*n* = 12)
Protectins			
cvaC	2 (9)	1 (10)	1 (8)
traT	12 (55)	6 (60)	6 (50)
iss	2 (9)	1 (10)	1 (8)
Invasins			
ibeA	3 (14)	0	3 (25)
gimB	1 (5)	0	1 (8)
Miscellaneous			
usp	13 (59)	5 (50)	8 (67)
ompT	18 (82)	8 (80)	10 (83)
PAI *(malX)*	12 (55)	7 (70)	5 (42)
pafP	15 (68)	6 (60)	9 (75)
upaH	11 (50)	4 (40)	7 (58)

There was no isolate positive for *papG* allele II, *bmaE*, *gafD*, *rfc*. None of the isolates tested had two copies of the *hlyA* operon.

Certain VFs were highly prevalent in urinary and vaginal isolates (Table 2) such as adhesins *fimH* (90% and 92%, respectively), *matB* (80% and 83%, respectively) or *ompT* (80% and 83%, respectively). Several toxins and siderophores were detected in strains from both niches, while invasins were only detected in vaginal isolates. We also found that many urinary and vaginal strains possess *kpsMT* II type (50% and 58%, respectively).

3.2. Virulence Profile Characterization of Urogenital E. coli

A detailed characterization of virulence gene profiles was performed on representative isolates per donor, independently on the sample type. For this purpose, vaginal isolates that presented identical PFGE profiles to urinary isolates (Supplementary Figure S1, Table 1) from the same individual and phylogenetic group were discarded, leaving a total of 16 representative urogenital strains from healthy women.

A hierarchical clustering analysis was performed to assess strain similarity based on VF profiles (Figure 1). Isolates were grouped into the following three primary clusters: (i) representing strains from B2 phylogenetic group (*n* = 5) enriched in *hlyA*, *cnf*1, *tsh*, *vat* and *iroN*; (ii) strains from B2 (*n* = 3) and F (*n* = 2) phylogenetic group with *sat* and *iutA* (without adhesins and previously mentioned genes) and (iii) strains belonging to other phylogenetic groups (*n* = 6), less enriched in VF (without *yfcB*, *usp*, *upaH*, *pafP*). One strain belonging to phylogenetic group A did not possess any virulence gene screened (Figure 1).

3.3. Genomic Background and Antimicrobial Resistance of B2 E. coli Strains

The genetic features of the eight B2 ExPEC strains subjected to whole genome sequencing (WGS) are presented in Table 1, and characteristics of draft genomes assemblies and accession numbers are available in Supplementary Table S3. Remarkably, some of these B2 strains were identified as pandemic *E. coli* lineages such as ST131, ST127, clonal complex 73 or 95. Others were identified as ST452, ST569, ST681 and ST998 (Table 1). They represented various serogroups, including O2, O6, O8, O25, O46 and O81 (Table 1). Additionally, we performed an MLST analysis of the B2 strains from asymptomatic women identified in the previous study by Garretto et al. [31]. Remarkably, those strains also belonged to intercontinental STs including ST95 (*n* = 3), ST73 (*n* = 1), ST12 (*n* = 1) and ST1193 (*n* = 1).

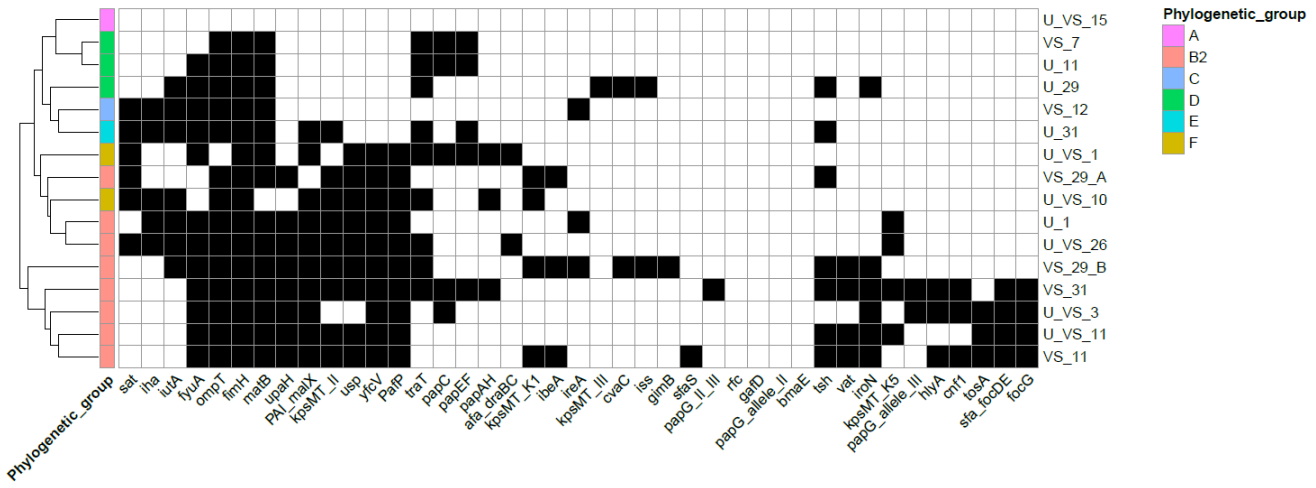


Figure 1. Heatmap representing presence/absence matrix of putative virulence genes among urogenital *E. coli* from different phylogenetic groups. The dendrogram in the left represents clustering of isolates based on Euclidean distance. Presence or absence of virulence genes is represented as black or white squares, respectively. The right hand side of the figure contains the list of unique *E. coli* profiles coded with U (for urinary) or VS (for vaginal swab) for origin and a number for identification of donor, according to Table 1.

We performed a core genome single nucleotide polymorphism (SNP) alignment of 14 B2 genomes isolated from asymptomatic women (8 from our study and 6 from Garretto et al. [31]), as shown in Figure 2. According to this phylogenetic tree, the genomes were substantially different and presented an average of ~28,000 SNP differences. The most related strains belonged to the ST95 clonal complex (UMB6611, c29VSb_15_M, UMB6713, UMB6454) which has ~180–4250 SNP differences, followed by the ST73 clonal complex (c11Ub_17_AN and UMB0939) which differed by 2353 SNP.

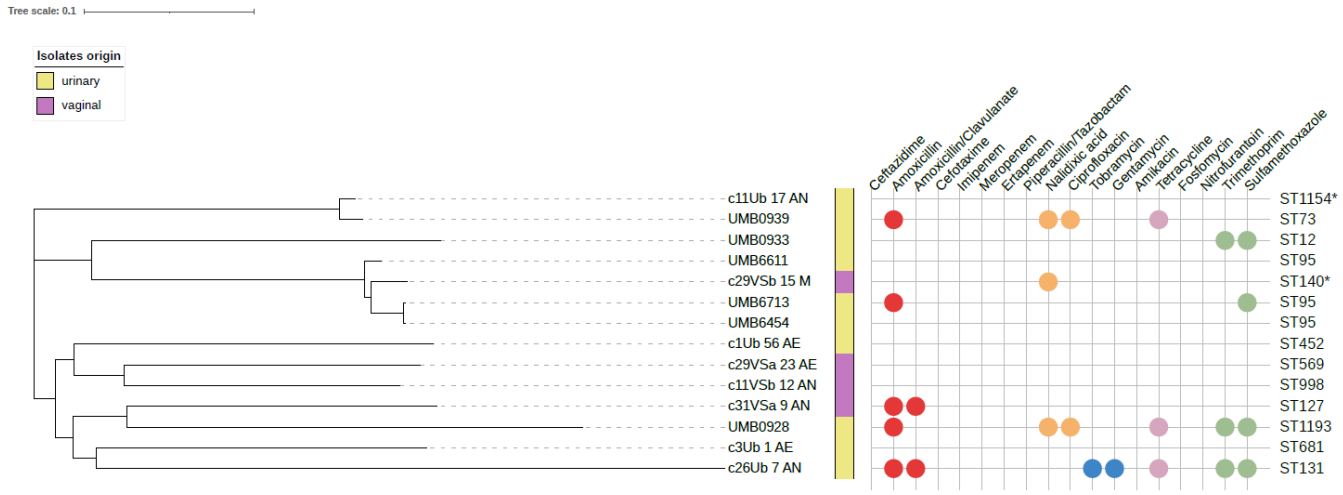


Figure 2. Phylogenetic tree representing core genome SNPs alignment of 14 urogenital *E. coli* strains isolated from asymptomatic women (8 from our study and 6 from Garretto et al.). The alignment was performed using *E. coli* strain UTI89 as a reference. The tree is unrooted. Colored balls represent the resistance phenotype/genotype to a given antibiotic (colors by antibiotic class). Resistance of the strains originated from this study was characterized by phenotypic and genotypic methods, while strains from Garretto et al. were characterized genotypically. Single locus variants (SLV) are represented with an asterisk (see Table 1).

Overall, 7/14 strains were resistant to at least one antibiotic, including antibiotics used to treat UTI (Figure 2). The ST131 (c26Ub_7_AN) and ST1193 (UMB0928) strains revealed multidrug resistance phenotypes (resistance to 7 and 6 antibiotics from 4 classes, respectively), including to critical antibiotics for UTI treatment (trimethoprim/sulfamethoxazole and amoxicillin-clavulanate) [48]. Additionally, ST12, ST73, ST95, ST140 (SLV 95) and ST127 strains were resistant to amoxicillin-clavulanate, nalidixic acid, ciprofloxacin, trimethoprim/sulfamethoxazole and/or tetracycline (Figure 2). The detection of the acquired resistance genes in our genomes was compatible with phenotypic characterization of resistance: (i) $bla_{\text{TEM-30}}$, in strains resistant to amoxicillin-clavulanic acid combination; (ii) *tet*(A) for tetracycline resistant strain; (iii) *aac*(3)-*IId*, *sul*1 or *dfrA*12 conferring resistance to aminoglycosides or trimethoprim (Supplementary Table S3). We also observed mutations in *gyrA* (S83L) for isolate c29VSb_15M, conferring resistance to fluoroquinolones (Figure 2, Supplementary Table S3).

Screening of plasmid replicon sequences in our eight isolates revealed that no known replicon sequences were detected for four of the strains (ST569, ST681, ST998, ST1154), whereas, in the other B2 strains, several replicon types were detected, mostly Col156 and IncFIB (AP001918) (Supplementary Table S3).

3.4. Phylogenomics of Urogenital ST131 E. coli from Healthy Urinary Microbiome

We performed a whole genome SNP phylogeny analysis in relation to the origin and health status of the ST131 *E. coli* strain isolated from asymptomatic woman comparatively with representative ST131 genomes from the NCBI public database.

In a total of 520 ST131 genomes for which metadata was available, most were obtained from human infections (D; n = 422 strains) and much less frequently from healthy carriers (H; n = 98 strains). They were identified mainly in wounds or pus, lung and sputum, blood or other body fluids (other; n = 242 strains), followed by strains isolated from the urinary (n = 179 strain) and gastrointestinal tract (n = 99 strains). While isolates obtained from the gastrointestinal tract originated mostly from healthy humans (H = 96, D = 3), isolates from the urinary tract were almost exclusively associated with urinary tract infections and only one genome (from this study) originated from healthy host (H = 1, D = 178). Isolates from other human niches were mostly isolated from the diseased host (H = 1, D = 241), e.g., bacteremia, sepsis, pneumonia.

The core genome phylogenetic tree was congruent with previous ST131 phylogenetic inferences and establishes the clustering of genomes in three clades corresponding to clade A (n = 67), clade B (n = 50) and clade C (n = 406) (Figure 3). Our ST131 strain from healthy woman (c26Ub_7_AN) clustered in clade B isolates despite carrying *fimH*30 and showed similarity to other clade B genomes from gastrointestinal tract colonizers (1111 SNPs) or infections (1157 SNPs). Interestingly, to date, this genome represents the unique ST131 from the urogenital tract of a healthy individual.

We then examined strains similarity according to their accessory genome (Figure 4). The NMDS ordination based on a 28,542 accessory gene matrix demonstrated three different clusters that are likely to correspond to clades A, B and C (from the right to the left, respectively; Figure 4), irrespective of host health status or the isolate's origin. The urinary strain c26Ub_7_AN isolated from healthy woman was identified in the proximity of two clusters.

Overall, genomes from different niches and disease statuses were closely related and randomly distributed in the phylogeny, often with less than 50 SNPs among them. The SNPs matrix for the ST131 core-genome-based phylogenetic tree is available in Supplementary Table S4.

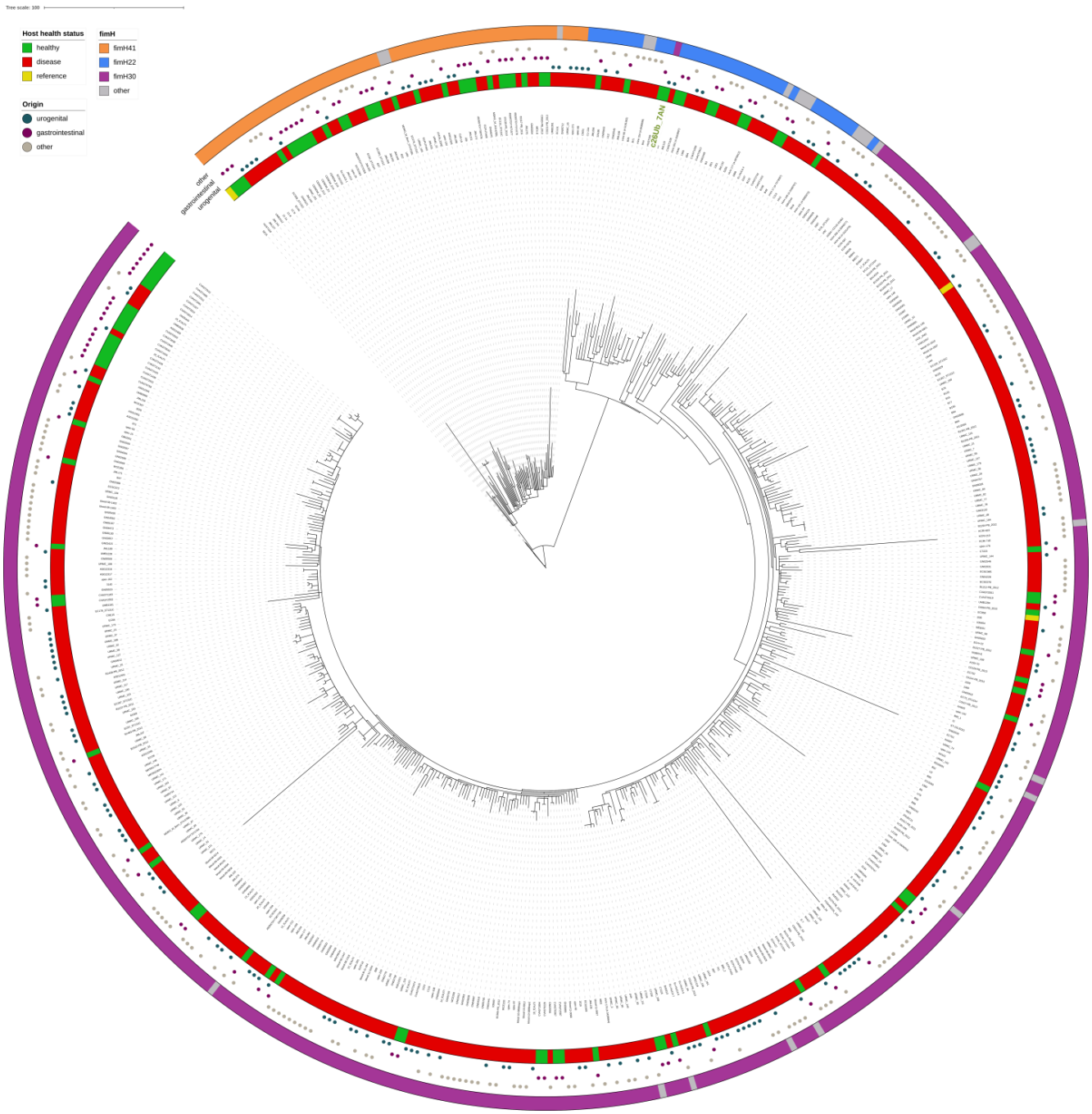


Figure 3. Core genome SNPs phylogenetic tree of 520 ST131 *E. coli* genomes of human origin. The phylogenetic tree includes 3 ST131 reference strains, i.e., SE15, JJ1886, EC958. The alignment was performed using *E. coli* strain EC958 as a reference and *E. coli* SE15 was used as an outgroup. The metadata including host health status, strain origin and *fimH* type are incorporated in the tree, as explained in the legend. The identifier of the strain from our urogenital collection is marked in green.

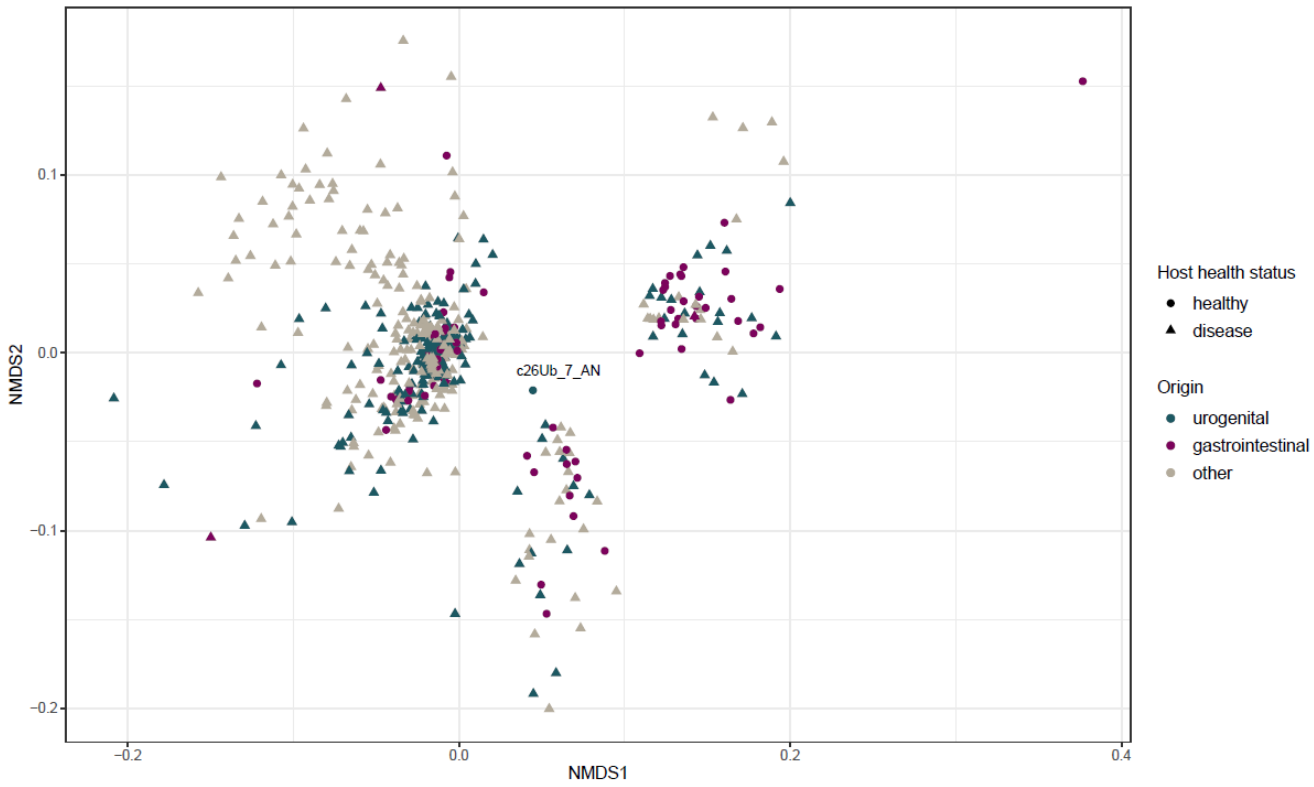


Figure 4. Non-metric multidimensional scaling (NMDS) based on 28,542 genes matrix (accessory genome) extracted from 520 ST131 *E. coli* genomes. Metadata layers of origin and host health status are incorporated in the figure. Stress value for this ordination is 0.18.

3.5. Phylogenomic Analysis of Other B2 E. coli Strains

A whole genome comparison was performed between our B2 non-ST131 isolates and the available genomes from the same ST on public databases. In agreement with the available literature, ST95 (n = 272) including ST140 (SLV 95) and ST73 (n = 217) including ST1154 (SLV 73) are the second and third most represented clonal groups (Supplementary Figures S2 and S3, respectively). SNP-based phylogenetic trees containing our ST140 (ST95 clonal complex), and ST1154 (ST73 clonal complex) urogenital *E. coli* isolates also evidenced a high similarity between strains from different host status and origin, often showing <1000 SNPs differences. The same was observed for the clone belonging to ST127 (n = 66), which was closely related to genomes from isolates causing UTI and/or isolated from other niche (<300 SNP differences, Supplementary Figure S4).

The remaining four STs (ST452, ST569, ST681, ST998) were much less represented in the NCBI database (uncommon STs). However, a comparative genomic analysis also showed, a high similarity (390–9452 SNPs) to genomes deposited from other sources (Figure 5). Interestingly, the corresponding strains were isolated in different continents.

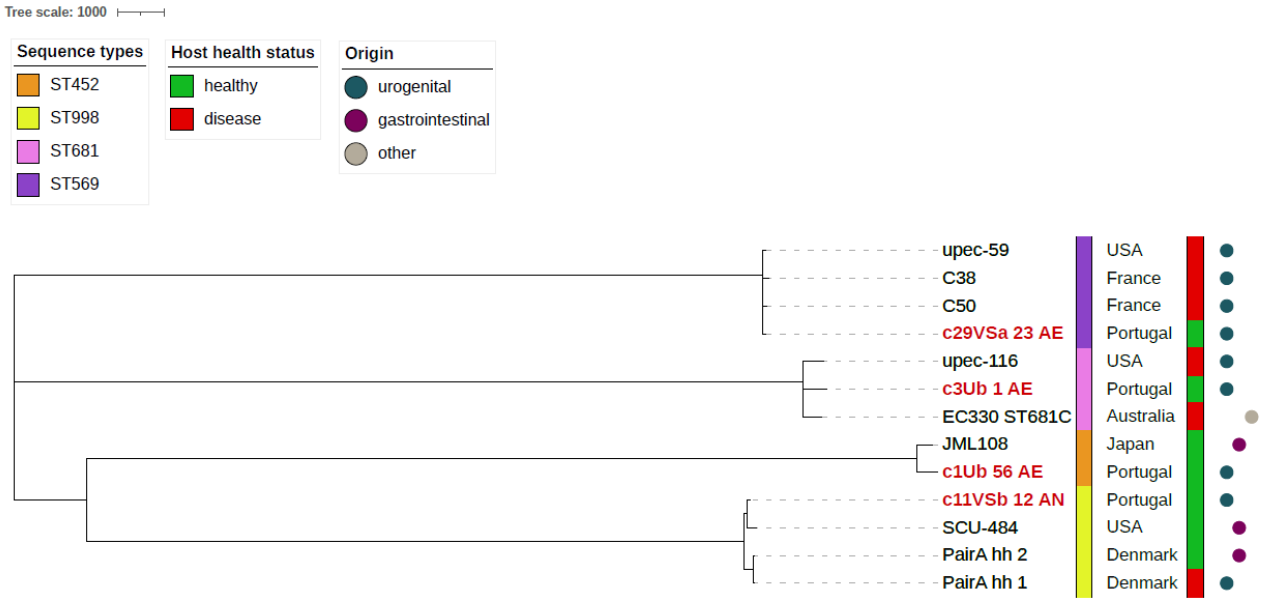


Figure 5. Whole genome SNPs phylogenetic tree of 4 infrequent *E. coli* STs (ST452, ST569, ST681, ST998) including 4 from our urogenital collection (marked in red) and 9 genomes retrieved from NCBI database. The alignment was performed using *E. coli* strain 536 as a reference. The tree is unrooted. Information on ST, health status of the host and origin of the strains is incorporated in the tree.

4. Discussion

In this study, we provided a detailed strain-level analysis of the *E. coli* isolates identified in the female urogenital microbiome, revealing that the urogenital tract of healthy women is an additional reservoir of pandemic ExPEC clones with the potential to cause UTI. Furthermore, we found an ST131 strain inhabiting the urogenital tract of healthy woman with a similar genetic background to those causing UTI or colonizing the gastrointestinal tract of humans.

E. coli was often found both in the urine (30%) and vagina (37%) of healthy women, the latter revealing a prevalence of slightly higher than that reported in previous studies [20,49]. Our data further supports that there is an interconnection between these two niches [28] in the same individual since the urine and vagina often share identical strains. However, in some samples, we also found variability in the number and type of *E. coli* strains identified in each of those niches, highlighting possible limitations in studies evaluating only one of the two locations for a risk assessment of UTI acquisition. Furthermore, our data also highlights that strain level characterization is relevant to accurately evaluate the contribution of these niches as reservoirs of strains with the potential to contribute to the persistence and/or cause UTI.

We described a higher *E. coli* diversity in the urogenital microbiome of healthy women (phylogenetic groups A, B2, D, E, F) than that reported in the study from Garretto et al., where only B2 *E. coli* strains (n = 6) were identified in asymptomatic women [31].

We demonstrated that strains from healthy urinary microbiome belong to serogroups known to be commonly involved in UTI (e.g., O2, O6, O25) [13,50,51]. Remarkably, the identification of ExPEC (n = 7 B2; n = 2 F) strains in healthy women was also observed, enriched in putative virulence factors known to favor colonization and/or the invasion of epithelial cells and the development of UTI [6]. We found that the virulence gene profiles detected in the ExPEC isolates from healthy women were identical to those described in clinical isolates [16]. All other non-ExPEC strains, except one phylogroup A *E. coli*, also possessed several adhesins, siderophores and other putative virulence genes, likely supporting their adaptation and survival in the urogenital tract. Thus, our study provides further evidence that the urogenital tract carries strains with pathogenic potential.

Remarkably, half of our B2 strains from asymptomatic women belonged to worldwide disseminated *E. coli* lineages (ST73 complex, ST95 complex, ST131, ST127), that have caused infections in humans [13,15,17,18]. Our phylogenomic analysis revealed close relatedness with publicly available genomes regardless of human origin (host status or infection site), further demonstrating the wide circulation of well-adapted and potentially pathogenic clones between different individuals and an interconnection between the urogenital and the gastrointestinal tract, which seems to be the most probable source of these strains.

The identification of strains, including ST131, in the urogenital tract of healthy women with phenotypes resistant to antibiotic classes used to treat UTI (amoxicillin-clavulanate, ciprofloxacin, trimethoprim/sulfamethoxazole; Figure 2) represents an additional concern due to the risk of treatment failure. Moreover, antibiotic-resistant strains in the microbiome constitute a reservoir of transferable antimicrobial resistance genes that can be shared with other strains/species by horizontal gene transfer [52,53]. Thus, the possibility of enriching antibiotic-resistant bacterial species (e.g., *E. coli*, *Citrobacter koseri*, *Klebsiella pneumoniae*) [29,30] in the urogenital microbiome can have direct implications on human health in future.

The identification of ST131 *E. coli* in the urine of healthy woman challenges the current understanding of the ecology of this pandemic clone. We demonstrate, for the first time, that the healthy urogenital microbiome can itself act as a reservoir of putative pathogenic and antibiotic resistant ST131 strains. We further highlighted that closely related isolates from different clades are identified as colonizers of the gastrointestinal or urogenital tracts of healthy people and can cause disease [54,55], and that the accessory genome structures those clades irrespective of niche origin. Interestingly, our ST131 strain that was isolated from a healthy woman (c26Ub_7) belongs to clade B, despite having *fimH*30. The detection of the *fimH*30 allele in clade B isolates is uncommon, but these recombination events have been occasionally observed in other collections [11,21]. Available data on the prevalence of ST131 subclades is largely biased by human clinical isolates, and clade C in particular [56–58]. However, there is evidence that non-human sources are underestimated and there is a specialization for each clade, including a foodborne origin for clade B-*fimH*22 [59] and wastewater for clade A-*fimH*41 [58], suggesting those as additional possible sources of acquisition.

Besides ST131, other intercontinental and/or emerging B2 *E. coli* clones were found in the urogenital microbiome of healthy women (i.e., ST95 complex, ST73 complex, ST127, ST12 and ST1193) [21,60,61]. Overall, these data suggest that the healthy urogenital microbiome is a source of particularly widespread (e.g., food, domestic animals, environment) *E. coli* lineages that are frequently responsible for UTI [62,63].

Much less data are available for uncommon and antibiotic susceptible *E. coli* clones (ST452, ST569, ST681, ST998) detected in the urogenital microbiome of healthy women, since they are poorly represented in public databases. Nevertheless, they have been previously reported as animal colonizers or from human infections in different continents (Figure 5), often associated with antibiotic resistance. For instance, *mcr*-1 positive ST452 or ESBL producing ST998 *E. coli* strains were previously isolated from animals, including domestic animals, i.e., dog or human UTI in five continents [64–68]. Similarly, ST681 strains have been detected worldwide in animals (wild boars, non-human primates) and among isolates causing human ExPEC infections, including UTI [69–72]. Furthermore, besides reports on UTI caused by ST569, this lineage could have possible food and environmental reservoirs, as, in the USA, ST569 *E. coli* was found in meat [73] and it was detected in wastewater in South Africa [74]. Mbanga et al. also demonstrated that a strain isolated from wastewater clustered together with previously isolated clinical ST998 *E. coli* isolated from a UTI patient in the same area [65].

Nevertheless, whether those strains possess the ability to persistently colonize the human urogenital tract, or if they are only transient microbiome members remains to be evaluated. Furthermore, one of the most fundamental questions is still unanswered, specifically, the circumstances that would trigger the development of infection by those colonizing strains [75].

The limitations of this study include the small sample size and the absence of fecal samples from the same individuals to assess concomitant gastrointestinal colonization; however, the in silico comparative genomic analysis performed showed a close relatedness of urinary and gastrointestinal strains. Using voided urine was also considered as a limitation, due to the possibility of collecting microbes that reside in the vaginal environment, but we demonstrated that a voided urine sample can provide distinct strains to those found in vagina. Furthermore, insights from this type of sample are valuable, as it is used in current UTI diagnostic practices.

5. Conclusions

In this study, we demonstrated that *E. coli* strains identified in the urogenital microbiome of healthy women frequently belong to international, pandemic, and occasionally antibiotic resistant B2 clones. Our data support the role of the urogenital tract as a source of ExPEC strains prone to cause UTI, which most probably originate from the gastrointestinal tract. Furthermore, we demonstrated that diverse *E. coli* lineages can be observed per individual and urogenital sample type, which should be considered in future studies focusing on *E. coli* screening. These findings are a hallmark to further understand the ecology of these clones and their distribution in human host reservoirs.

The influence of microbial communities, the host and environmental features in triggering the transition from colonization to infection is still unknown, but this subject deserves further investigation.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/microorganisms10010027/s1, Figure S1: Pulsed-field Gel Electrophoresis (PFGE) of selected *E. coli* isolates. Figure S2: Whole genome SNPs phylogenetic tree of 272 ST95 and ST140 (SLV 95) *E. coli* genomes of human origin. Figure S3: Whole genome SNPs phylogenetic tree of 217 ST73 and ST1154 (SLV 73) *E. coli* genomes of human origin.; Figure S4: Whole genome SNPs phylogenetic tree of 66 ST127 *E. coli* genomes of human origin. Table S1: List of donors, samples and isolates subjected to analysis in this study. Table S2: List of primers used for detection of ExPEC virulence genes. Table S3: Assemblies statistics and accession numbers per each novel strain used in this study (BioProject accession number PRJNA548360). Table S4: SNPs matrix for ST131 core genome.

Author Contributions: M.K., conceptualization, investigation, methodology, formal analysis, data curation, visualization, writing —original draft preparation; Â.N., conceptualization, validation, supervision, writing—review and editing; L.P., supervision, funding acquisition, project administration, resources, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by national funds from FCT-Fundação para a Ciência e a Tecnologia, I.P., in the scope of the project UIDP/04378/2020 and UIDB/04378/2020 of the Research Unit on Applied Molecular Biosciences-UCIBIO and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy-i4HB. MK is supported by FCT [SFRH/BD/132497/2017]. ÂN is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0032].

Institutional Review Board Statement: The study was developed according to the Helsinki Declaration principles and the protocol was approved by the Ethical Commission of Faculty of Pharmacy, University of Porto (Ethical Committee approval number 32-09-2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The Whole Genome Shotgun project of the isolates c1Ub_56_AE, c3Ub_1_AE, c11Ub_17_AN, c11VSb_12_AN, c13Ua_2_AN, c14Ub_22_AE, c26Ub_7_AN, c29VSa_23_AE, c29VSb_15_M, c31VSa_9_AN, c32Ub_19_M have been deposited at DDBJ/ENA/GenBank under the accession numbers JAECYT000000000, JAECYS000000000, JAECYR000000000, JAECYQ000000000, JAECYP000000000, JAECYO000000000, JAECYN000000000, JAECYM000000000, JAECYL000000000, JAECYK000000000, JAECYJ000000000, respectively.

Acknowledgments: We would like to thank Svetlana Ugarcina Perovic for performing the MALDI-TOF MS analysis and bioMérieux (Portugal, Lda) for MALDI-TOF MS equipment and material provided. We would also like to acknowledge Svetlana Ugarcina Perovic, Joana Rocha and Filipa Grosso for their fundamental contribution to Female Urogenital Microbiome Project. Additionally, we would like to acknowledge the support of HEALTH-UNORTE I&D&I project (NORTE-01-0145-FEDER-000039) cofinanced by European Regional Development Fund (ERDF), through the NORTE 2020 (Programa Operacional Regional do Norte 2014/2020).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Negus, M.; Phillips, C.; Hindley, R. Recurrent Urinary Tract Infections: A Critical Review of the Currently Available Treatment Options. *Obstet. Gynaecol.* **2020**, *22*, 115–121. [CrossRef]
2. Öztürk, R.; Murt, A. Epidemiology of Urological Infections: A Global Burden. *World J. Urol.* **2020**, *38*, 2669–2679. [CrossRef] [PubMed]
3. Subashchandrabose, S.; Mobley, H.L.T. Virulence and Fitness Determinants of Uropathogenic *Escherichia coli*. In *Urinary Tract Infections: Molecular Pathogenesis and Clinical Management*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015; pp. 235–261. [CrossRef]
4. Foxman, B. The Epidemiology of Urinary Tract Infection. *Nat. Rev. Urol.* **2010**, *7*, 653–660. [CrossRef]
5. Medina, M.; Castillo-Pino, E. An Introduction to the Epidemiology and Burden of Urinary Tract Infections. *Ther. Adv. Urol.* **2019**, *11*, 1756287219832172. [CrossRef]
6. Flores-Mireles, A.L.; Walker, J.N.; Caparon, M.; Hultgren, S.J. Urinary Tract Infections: Epidemiology, Mechanisms of Infection and Treatment Options. *Nat. Rev. Microbiol.* **2015**, *13*, 269–284. [CrossRef] [PubMed]
7. Tamadonfar, K.O.; Omattage, N.S.; Spaulding, C.N.; Hultgren, S.J. Reaching the End of the Line: Urinary Tract Infections. *Microbiol. Spectr.* **2019**, *7*, 3. [CrossRef]
8. Russo, T.A.; Johnson, J.R. Proposal for a New Inclusive Designation for Extraintestinal Pathogenic Isolates of *Escherichia coli*: ExPEC. *J. Infect. Dis.* **2000**, *181*, 1753–1754. [CrossRef]
9. Manges, A.R.; Geum, H.M.; Guo, A.; Edens, T.J.; Fibke, C.D.; Pitout, J.D.D. Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. *Clin. Microbiol. Rev.* **2019**, *32*, e00135-18. [CrossRef]
10. Brzuszkiewicz, E.; Brüggemann, H.; Liesegang, H.; Emmerth, M.; Olschläger, T.; Nagy, G.; Albermann, K.; Wagner, C.; Buchrieser, C.; Emody, L.; et al. How to Become a Uropathogen: Comparative Genomic Analysis of Extraintestinal Pathogenic *Escherichia coli* Strains. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12879–12884. [CrossRef]
11. Matsumura, Y.; Pitout, J.D.D.; Peirano, G.; DeVinney, R.; Noguchi, T.; Yamamoto, M.; Gomi, R.; Matsuda, T.; Nakano, S.; Nagao, M.; et al. Rapid Identification of Different *Escherichia coli* Sequence Type 131 Clades. *Antimicrob. Agents Chemother.* **2017**, *61*, e00179-17. [CrossRef]
12. Stoesser, N.; Sheppard, A.E.; Pankhurst, L.; De Maio, N.; Moore, C.E.; Sebra, R.; Turner, P.; Anson, L.W.; Kasarskis, A.; Batty, E.M.; et al. Evolutionary History of the Global Emergence of the *Escherichia coli* Epidemic Clone ST131. *MBio* **2016**, *7*, e02162. [CrossRef]
13. Ciesielczuk, H.; Jenkins, C.; Chattaway, M.; Doumith, M.; Hope, R.; Woodford, N.; Wareham, D.W. Trends in ExPEC Serogroups in the UK and Their Significance. *Eur. J. Clin. Microbiol. Infect. Dis.* **2016**, *35*, 1661–1666. [CrossRef] [PubMed]
14. Kallonen, T.; Brodrick, H.J.; Harris, S.R.; Corander, J.; Brown, N.M.; Martin, V.; Peacock, S.J.; Parkhill, J. Systematic Longitudinal Survey of Invasive *Escherichia coli* in England Demonstrates a Stable Population Structure Only Transiently Disturbed by the Emergence of ST131. *Genome Res.* **2017**, *27*, 1437–1449. [CrossRef] [PubMed]
15. Fibke, C.D.; Croxen, M.A.; Geum, H.M.; Glass, M.; Wong, E.; Avery, B.P.; Daignault, D.; Mulvey, M.R.; Reid-Smith, R.J.; Parmley, E.J.; et al. Genomic Epidemiology of Major Extraintestinal Pathogenic *Escherichia coli* Lineages Causing Urinary Tract Infections in Young Women Across Canada. *Open Forum Infect. Dis.* **2019**, *6*, ofz431. [CrossRef] [PubMed]
16. Flament-Simon, S.-C.; Nicolas-Chanoine, M.-H.; García, V.; Duprilot, M.; Mayer, N.; Alonso, M.P.; García-Meniño, I.; Blanco, J.E.; Blanco, M.; Blanco, J. Clonal Structure, Virulence Factor-Encoding Genes and Antibiotic Resistance of *Escherichia coli*, Causing Urinary Tract Infections and Other Extraintestinal Infections in Humans in Spain and France during 2016. *Antibiotics* **2020**, *9*, 161. [CrossRef]
17. Johnson, J.R.; Davis, G.; Clabots, C.; Johnston, B.D.; Porter, S.; DebRoy, C.; Pomputius, W.; Ender, P.T.; Cooperstock, M.; Slater, B.S.; et al. Household Clustering of *Escherichia coli* Sequence Type 131 Clinical and Fecal Isolates According to Whole Genome Sequence Analysis. In *Open Forum Infectious Diseases*; Oxford University Press: Oxford, UK, 2016; Volume 3.
18. Banerjee, R.; Johnson, J.R. A New Clone Sweeps Clean: The Enigmatic Emergence of *Escherichia coli* Sequence Type 131. *Antimicrob. Agents Chemother.* **2014**, *58*, 4997–5004. [CrossRef] [PubMed]
19. Lewis, A.L.; Gilbert, N.M. Roles of the Vagina and the Vaginal Microbiota in Urinary Tract Infection: Evidence from Clinical Correlations and Experimental Models. *GMS Infect. Dis.* **2020**, *8*, Doc02. [CrossRef] [PubMed]
20. O'Brien, V.P.; Gilbert, N.M.; Lebratti, T.; Agarwal, K.; Foster, L.; Shin, H.; Lewis, A.L. Low-Dose Inoculation of *Escherichia coli* Achieves Robust Vaginal Colonization and Results in Ascending Infection Accompanied by Severe Uterine Inflammation in Mice. *PLoS ONE* **2019**, *14*, e0219941. [CrossRef]

21. Johnson, J.R.; Russo, T.A. Molecular Epidemiology of Extraintestinal Pathogenic *Escherichia coli*. *EcoSal Plus* **2018**, *8*. [CrossRef]
22. Brannon, J.R.; Dunigan, T.L.; Beebout, C.J.; Ross, T.; Wiebe, M.A.; Reynolds, W.S.; Hadjifrangiskou, M. Invasion of Vaginal Epithelial Cells by Uropathogenic *Escherichia coli*. *Nat. Commun.* **2020**, *11*, 2803. [CrossRef]
23. Siddiqui, H.; Nederbragt, A.J.; Lagesen, K.; Jeansson, S.L.; Jakobsen, K.S. Assessing Diversity of the Female Urine Microbiota by High Throughput Sequencing of 16S RDNA Amplicons. *BMC Microbiol.* **2011**, *11*, 244. [CrossRef] [PubMed]
24. Wolfe, A.J.; Toh, E.; Shibata, N.; Rong, R.; Kenton, K.; Fitzgerald, M.; Mueller, E.R.; Schreckenberger, P.; Dong, Q.; Nelson, D.E.; et al. Evidence of Uncultivated Bacteria in the Adult Female Bladder. *J. Clin. Microbiol.* **2012**, *50*, 1376–1383. [CrossRef] [PubMed]
25. Neugent, M.L.; Hulyalkar, N.V.; Nguyen, V.H.; Zimmern, P.E.; Nisco, N.J.D. Advances in Understanding the Human Urinary Microbiome and Its Potential Role in Urinary Tract Infection. *MBio* **2020**, *11*, e00218-20. [CrossRef]
26. Jones-Freeman, B.; Chonwerawong, M.; Marcelino, V.R.; Deshpande, A.V.; Forster, S.C.; Starkey, M.R. The Microbiome and Host Mucosal Interactions in Urinary Tract Diseases. *Mucosal Immunol.* **2021**, *14*, 779–792. [CrossRef] [PubMed]
27. Price, T.K.; Dune, T.; Hilt, E.E.; Thomas-White, K.J.; Kliethermes, S.; Brincat, C.; Brubaker, L.; Wolfe, A.J.; Mueller, E.R.; Schreckenberger, P.C. The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. *J. Clin. Microbiol.* **2016**, *54*, 1216–1222. [CrossRef]
28. Thomas-White, K.; Forster, S.C.; Kumar, N.; Van Kuiken, M.; Putonti, C.; Stares, M.D.; Hilt, E.E.; Price, T.K.; Wolfe, A.J.; Lawley, T.D. Culturing of Female Bladder Bacteria Reveals an Interconnected Urogenital Microbiota. *Nat. Commun.* **2018**, *9*, 1557. [CrossRef]
29. Coorevits, L.; Heytens, S.; Boelens, J.; Claeys, G. The Resident Microflora of Voided Midstream Urine of Healthy Controls: Standard versus Expanded Urine Culture Protocols. *Eur. J. Clin. Microbiol. Infect. Dis.* **2017**, *36*, 635–639. [CrossRef]
30. Ksiezarek, M.; Ugarcina-Perovic, S.; Rocha, J.; Grosso, F.; Peixe, L. Long-Term Stability of the Urogenital Microbiota of Asymptomatic European Women. *BMC Microbiol.* **2021**, *21*, 64. [CrossRef]
31. Garretto, A.; Miller-Ensminger, T.; Ene, A.; Merchant, Z.; Shah, A.; Gerodias, A.; Biancofiori, A.; Canchola, S.; Canchola, S.; Castillo, E.; et al. Genomic Survey of *E. coli* From the Bladders of Women With and Without Lower Urinary Tract Symptoms. *Front. Microbiol.* **2020**, *11*, 2094. [CrossRef]
32. Wang, R.F.; Cao, W.W.; Cerniglia, C.E. PCR Detection and Quantitation of Predominant Anaerobic Bacteria in Human and Animal Fecal Samples. *Appl. Environ. Microbiol.* **1996**, *62*, 1242–1247. [CrossRef]
33. Clermont, O.; Christenson, J.K.; Denamur, E.; Gordon, D.M. The Clermont *Escherichia coli* Phylo-Typing Method Revisited: Improvement of Specificity and Detection of New Phylo-Groups. *Environ. Microbiol. Rep.* **2013**, *5*, 58–65. [CrossRef] [PubMed]
34. Ksiezarek, M.; Novais, Â.; Felga, H.; Mendes, F.; Escobar, M.; Peixe, L. Phylogenomic Analysis of a Highly Virulent *Escherichia coli* ST83 Lineage with Potential Animal-Human Transmission. *Microb. Pathog.* **2021**, *155*, 104920. [CrossRef] [PubMed]
35. Johnson, J.R.; Murray, A.C.; Gajewski, A.; Sullivan, M.; Snippes, P.; Kuskowski, M.A.; Smith, K.E. Isolation and Molecular Characterization of Nalidixic Acid-Resistant Extraintestinal Pathogenic *Escherichia coli* from Retail Chicken Products. *Antimicrob. Agents Chemother.* **2003**, *47*, 2161–2168. [CrossRef] [PubMed]
36. Gautom, R.K. Rapid Pulsed-Field Gel Electrophoresis Protocol for Typing of *Escherichia coli* O157:H7 and Other Gram-Negative Organisms in 1 Day. *J. Clin. Microbiol.* **1997**, *35*, 2977–2980. [CrossRef]
37. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
38. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
39. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
40. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
41. Jolley, K.A.; Maiden, M.C. BIGSdb: Scalable Analysis of Bacterial Genome Variation at the Population Level. *BMC Bioinform.* **2010**, *11*, 595. [CrossRef]
42. Croucher, N.J.; Page, A.J.; Connor, T.R.; Delaney, A.J.; Keane, J.A.; Bentley, S.D.; Parkhill, J.; Harris, S.R. Rapid Phylogenetic Analysis of Large Samples of Recombinant Bacterial Whole Genome Sequences Using Gubbins. *Nucleic Acids Res.* **2015**, *43*, e15. [CrossRef]
43. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef] [PubMed]
44. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018; Available online: https://www.R-project.org/ (accessed on 10 April 2020).
46. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. Vegan: Community Ecology Package. 2019. Available online: https://CRAN.R-project.org/package=vegan (accessed on 10 April 2020).
47. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.

48. Magiorakos, A.-P.; Srinivasan, A.; Carey, R.B.; Carmeli, Y.; Falagas, M.E.; Giske, C.G.; Harbarth, S.; Hindler, J.F.; Kahlmeter, G.; Olsson-Liljequist, B.; et al. Multidrug-Resistant, Extensively Drug-Resistant and Pandrug-Resistant Bacteria: An International Expert Proposal for Interim Standard Definitions for Acquired Resistance. *Clin. Microbiol. Infect.* **2012**, *18*, 268–281. [CrossRef]
49. Salinas, A.M.; Osorio, V.G.; Pacha-Herrera, D.; Vivanco, J.S.; Trueba, A.F.; Machado, A. Vaginal Microbiota Evaluation and Prevalence of Key Pathogens in Ecuadorian Women: An Epidemiologic Analysis. *Sci. Rep.* **2020**, *10*, 18358. [CrossRef]
50. Johnson, J.R.; Stell, A.L. Extended Virulence Genotypes of *Escherichia coli* Strains from Patients with Urosepsis in Relation to Phylogeny and Host Compromise. *J. Infect. Dis.* **2000**, *181*, 261–272. [CrossRef] [PubMed]
51. Sarkar, S.; Ulett, G.C.; Totsika, M.; Phan, M.-D.; Schembri, M.A. Role of Capsule and O Antigen in the Virulence of Uropathogenic *Escherichia coli*. *PLoS ONE* **2014**, *9*, e94786. [CrossRef]
52. Liu, L.; Chen, X.; Skogerbø, G.; Zhang, P.; Chen, R.; He, S.; Huang, D.-W. The Human Microbiome: A Hot Spot of Microbial Horizontal Gene Transfer. *Genomics* **2012**, *100*, 265–270. [CrossRef]
53. Langille, M.G.I.; Meehan, C.J.; Beiko, R.G. Human Microbiome: A Genetic Bazaar for Microbes? *Curr. Biol.* **2012**, *22*, R20–R22. [CrossRef]
54. Nielsen, K.L.; Dynesen, P.; Larsen, P.; Frimodt-Møller, N. Faecal *Escherichia coli* from Patients with *E. coli* Urinary Tract Infection and Healthy Controls Who Have Never Had a Urinary Tract Infection. *J. Med. Microbiol.* **2014**, *63*, 582–589. [CrossRef]
55. Moreno, E.; Andreu, A.; Perez, T.; Sabate, M.; Johnson, J.R.; Prats, G. Relationship between *Escherichia coli* Strains Causing Urinary Tract Infection in Women and the Dominant Faecal Flora of the Same Hosts. *Epidemiol. Infect.* **2006**, *134*, 1015–1023. [CrossRef]
56. Petty, N.K.; Ben Zakour, N.L.; Stanton-Cook, M.; Skippington, E.; Totsika, M.; Forde, B.M.; Phan, M.-D.; Gomes Moriel, D.; Peters, K.M.; Davies, M.; et al. Global Dissemination of a Multidrug Resistant *Escherichia coli* Clone. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5694–5699. [CrossRef]
57. Peirano, G.; van der Bij, A.K.; Freeman, J.L.; Poirel, L.; Nordmann, P.; Costello, M.; Tchesnokova, V.L.; Pitout, J.D.D. Characteristics of *Escherichia coli* Sequence Type 131 Isolates That Produce Extended-Spectrum β-Lactamases: Global Distribution of the H30-Rx Sublineage. *Antimicrob. Agents Chemother.* **2014**, *58*, 3762–3767. [CrossRef]
58. Finn, T.J.; Scriver, L.; Lam, L.; Duong, M.; Peirano, G.; Lynch, T.; Dong, T.; Pitout, J.D.D.; DeVinney, R. A Comprehensive Account of *Escherichia coli* Sequence Type 131 in Wastewater Reveals an Abundance of Fluoroquinolone-Resistant Clade A Strains. *Appl. Environ. Microbiol.* **2020**, *86*, e01913-19. [CrossRef] [PubMed]
59. Liu, C.M.; Stegger, M.; Aziz, M.; Johnson, T.J.; Waits, K.; Nordstrom, L.; Gauld, L.; Weaver, B.; Rolland, D.; Statham, S.; et al. *Escherichia coli* ST131-*H22* as a Foodborne Uropathogen. *MBio* **2018**, *9*, e00470-18. [CrossRef] [PubMed]
60. Valenza, G.; Werner, M.; Eisenberger, D.; Nickel, S.; Lehner-Reindl, V.; Höller, C.; Bogdan, C. First Report of the New Emerging Global Clone ST1193 among Clinical Isolates of Extended-Spectrum β-Lactamase (ESBL)-Producing *Escherichia coli* from Germany. *J. Glob. Antimicrob. Resist.* **2019**, *17*, 305–308. [CrossRef] [PubMed]
61. Ding, Y.; Zhang, J.; Yao, K.; Gao, W.; Wang, Y. Molecular Characteristics of the New Emerging Global Clone ST1193 among Clinical Isolates of *Escherichia coli* from Neonatal Invasive Infections in China. *Eur. J. Clin. Microbiol. Infect. Dis.* **2021**, *40*, 833–840. [CrossRef] [PubMed]
62. Riley, L.W. Extraintestinal Foodborne Pathogens. *Ann. Rev. Food Sci. Technol.* **2020**, *11*, 275–294. [CrossRef] [PubMed]
63. Manges, A.R.; Johnson, J.R. Reservoirs of Extraintestinal Pathogenic *Escherichia coli*. *Microbiol. Spectr.* **2015**, *3*. [CrossRef] [PubMed]
64. Delgado-Blas, J.F.; Ovejero, C.M.; Abadia-Patiño, L.; Gonzalez-Zorn, B. Coexistence of *Mcr-1* and *BlaNDM-1* in *Escherichia coli* from Venezuela. *Antimicrob. Agents Chemother.* **2016**, *60*, 6356–6358. [CrossRef]
65. Mbelle, N.M.; Feldman, C.; Osei Sekyere, J.; Maningi, N.E.; Modipane, L.; Essack, S.Y. The Resistome, Mobilome, Virulome and Phylogenomics of Multidrug-Resistant *Escherichia coli* Clinical Isolates from Pretoria, South Africa. *Sci. Rep.* **2019**, *9*, 16457. [CrossRef]
66. Bozcal, E.; Eldem, V.; Aydemir, S.; Skurnik, M. The Relationship between Phylogenetic Classification, Virulence and Antibiotic Resistance of Extraintestinal Pathogenic *Escherichia coli* in İzmir Province, Turkey. *PeerJ* **2018**, *6*, e5470. [CrossRef]
67. Alghoribi, M.F.; Gibreel, T.M.; Farnham, G.; Al Johani, S.M.; Balkhy, H.H.; Upton, M. Antibiotic-Resistant ST38, ST131 and ST405 Strains Are the Leading Uropathogenic *Escherichia coli* Clones in Riyadh, Saudi Arabia. *J. Antimicrob. Chemother.* **2015**, *70*, 2757–2762. [CrossRef]
68. Yasugi, M.; Hatoya, S.; Motooka, D.; Matsumoto, Y.; Shimamura, S.; Tani, H.; Furuya, M.; Mie, K.; Miyake, M.; Nakamura, S.; et al. Whole-Genome Analyses of Extended-Spectrum or AmpC β-Lactamase-Producing *Escherichia coli* Isolates from Companion Dogs in Japan. *PLoS ONE* **2021**, *16*, e0246482.
69. Alonso, C.A.; González-Barrio, D.; Ruiz-Fons, F.; Ruiz-Ripa, L.; Torres, C. High Frequency of B2 Phylogroup among Non-Clonally Related Fecal *Escherichia coli* Isolates from Wild Boars, Including the Lineage ST131. *FEMS Microbiol. Ecol.* **2017**, *93*, fix016. [CrossRef] [PubMed]
70. Foster-Nyarko, E.; Alikhan, N.-F.; Ravi, A.; Thilliez, G.; Thomson, N.M.; Baker, D.; Kay, G.; Cramer, J.D.; O'Grady, J.; Antonio, M.; et al. Genomic Diversity of *Escherichia coli* Isolates from Non-Human Primates in the Gambia. *Microb. Genom.* **2020**, *6*, mgen000428. [CrossRef]
71. Hertz, F.B.; Nielsen, J.B.; Schønning, K.; Littauer, P.; Knudsen, J.D.; Løbner-Olesen, A.; Frimodt-Møller, N. Population Structure of Drug-Susceptible, -Resistant and ESBL-Producing *Escherichia coli* from Community-Acquired Urinary Tract Infections. *BMC Microbiol.* **2016**, *16*, 63.

72. Banerjee, R.; Johnston, B.; Lohse, C.; Chattopadhyay, S.; Tchesnokova, V.; Sokurenko, E.V.; Johnson, J.R. The Clonal Distribution and Diversity of Extraintestinal *Escherichia coli* Isolates Vary According to Patient Characteristics. *Antimicrob. Agents Chemother.* **2013**, *57*, 5912–5917. [CrossRef] [PubMed]
73. Yamaji, R.; Friedman, C.R.; Rubin, J.; Suh, J.; Thys, E.; McDermott, P.; Hung-Fan, M.; Riley, L.W. A Population-Based Surveillance Study of Shared Genotypes of *Escherichia coli* Isolates from Retail Meat and Suspected Cases of Urinary Tract Infections. *Msphere* **2018**, *3*, e00179-18. [CrossRef]
74. Mbanga, J.; Amoako, D.G.; Abia, A.L.K.; Allam, M.; Ismail, A.; Essack, S.Y. Genomic Insights of Multidrug-Resistant *Escherichia coli* From Wastewater Sources and Their Association With Clinical Pathogens in South Africa. *Front. Vet. Sci.* **2021**, *8*, 636715. [CrossRef]
75. Price, L.B.; Hungate, B.A.; Koch, B.J.; Davis, G.S.; Liu, C.M. Colonizing Opportunistic Pathogens (COPs): The Beasts in All of Us. *PLoS Pathog.* **2017**, *13*, e1006369. [CrossRef]

Additional material associated with the manuscript entitled “The darkest place is under the candlestick - healthy urogenital tract as a source of worldwide disseminated Extraintestinal Pathogenic *Escherichia coli* lineages” is available through the following links:

Supplementary Tables and Figures:
https://www.mdpi.com/2076-2607/10/1/27#supplementary

Phylogenomic analysis of a highly virulent *Escherichia coli* ST83 lineage with potential animal-human transmission

Magdalena Ksiezarek[1], Angela Novais[1], Helena Felga[2], Fatima Mendes[2], Marta Escobar[2], *Luisa Peixe[1]

[1]UCIBIO-REQUIMTE. Laboratory of Microbiology, Faculty of Pharmacy, University of Porto, 4050, Porto, Portugal
[2]Clínica dos Gatos. Veterinary Clinic, 4100, Porto, Portugal

Microbial Pathogenesis 2021 Jun;155:104920.

Microbial Pathogenesis 155 (2021) 104920

Contents lists available at ScienceDirect

Microbial Pathogenesis

journal homepage: www.elsevier.com/locate/micpath

Phylogenomic analysis of a highly virulent *Escherichia coli* ST83 lineage with potential animal-human transmission

Magdalena Ksiezarek [a], Ângela Novais [a], Helena Felga [b], Fátima Mendes [b], Marta Escobar [b], Luísa Peixe [a,*]

[a] UCIBIO-REQUIMTE. Laboratory of Microbiology, Faculty of Pharmacy, University of Porto, 4050, Porto, Portugal
[b] Clínica dos Gatos. Veterinary Clinic, 4100, Porto, Portugal

ARTICLE INFO

Keywords:
Pyelonephritis
Extraintestinal pathogenic *E. coli*
Virulence
Hemolysin A
Whole genome sequencing

ABSTRACT

The presence of specific virulence features conditions severe forms of urinary tract disease, but the frequency and distribution of these highly virulent extraintestinal pathogenic *Escherichia coli* strains in animals and humans is unclear. We used whole genome sequencing, comparative genomics, histological and clinical data to characterize the genetic basis for pathogenesis and origin of *E. coli* Ec_151217, a strain (B2, ST83, O83:H5:K5) that caused an extremely aggressive upper urinary tract infection (UTI) in a cat.

We show that Ec_151217 and 52% of other highly related ST83 genomes (O6 and O83) identified from different animal species and human infections carry two copies of the hemolysin A operon, though this duplication is infrequent (~1%) among closed ExPEC genomes from multiple sources.

Our data enlarges the list of *E. coli* genetic backgrounds carrying *hlyA* operon duplication which is potentially involved in severity of UTI, and demonstrates that it seems to occur infrequently amongst ExPEC. Its identification in *E. coli* lineages (diverse ST83 serotypes) of potential animal-human transmission is of concern and anticipates the need to screen larger collections.

1. Introduction

The characterization of the virulence genes content of *Escherichia coli* populations causing urinary tract infections (UTI) has been the subject of intensive research in order to find critical features for pathogenesis [1]. The presence of a wide array of virulence genes is correlated with increased pathogenesis, though most data are circumstantial [2,3]. Several *in vitro* and *in vivo* studies have supported the higher pathogenic potential observed by expression of specific toxins (e.g. *hlyA*, *cnf1*) or adhesins (e.g. P-fimbriae) though most available data result from characterization of a few strains [1,4]. More specifically, the duplication of particular virulence determinants such as hemolysin A (*hlyA*) has been associated with higher pathogenicity profile in pyelonephritic *E. coli* strain 536 (O6:K15), supported by *in vivo* models [5]. Furthermore, hemolysin A operon duplication has also been found in other human pyelonephritic *E. coli* strains such as J96 (O4:K6) or others (*E. coli* p19A, *E. coli* CP9), with several studies demonstrating differential HlyA expression and/or higher pathogenicity [5–8]. Nonetheless, the frequency with which this duplication occurs in extraintestinal pathogenic *Escherichia coli* (ExPEC) or its distribution in the ExPEC population is still unknown, hindering to establish the associated risk. In fact, frequency of *hlyA* duplication in isolates causing severe forms of upper UTI such as pyelonephritis or glomerulonephritis is not clear. Possibly because only a directed overlapping PCR approach would allow differentiating both operons [5, This study], and also because commonly used virulence screening databases on whole genome sequencing data do not allow identification of copies of the same target (e.g. hlyA) [9].

On the other hand, it has been widely discussed the extension and directionality of the contribution of ExPEC causing UTI in animals and humans. Studies addressing the ecology and potential transmission of ExPEC through the food chain, till recently, have been supported by methodologies with insufficient resolution for analysis of bacterial genetic similarity and biased towards a few clonal groups [10].

In this study, we combined detailed whole genome sequencing data, comparative genomics, histological and clinical data to characterize an *E. coli* strain that caused an acute pyelonephritis and glomerulonephritis in a cat, and to assess the distribution of similar highly aggressive strains in animals and humans by a genomics approach. Our data provides a

* Corresponding author.
E-mail address: lpeixe@ff.up.pt (L. Peixe).

https://doi.org/10.1016/j.micpath.2021.104920
Received 14 September 2020; Received in revised form 1 February 2021; Accepted 22 March 2021
Available online 27 April 2021

reliable genetic support on the role of *hlyA* duplication on the pathogenicity of a particularly virulent ExPEC lineage with potential animal-human transmission.

2. Material and methods

An *Escherichia coli* strain (Ec_151217) had caused a complicated UTI in a 10 years old Persian cat, that was finally diagnosed with a purulent interstitial glomerulonephritis, with pyelonephritis and urethritis. The histopathological analysis of the kidney revealed a severe inflammatory process, with a mixed, mononuclear infiltrate on the interstitial space and neutrophils (Supplementary Fig. S1).

The Ec_151217 strain identified in both purulent fluid and urine culture by phenotypic methods and PCR was characterized. *In vitro* antibiotic susceptibility testing was performed by disk diffusion for 16 antibiotics (beta-lactams, fluoroquinolones, aminoglycosides, tetracyclines, fosfomycin) and by broth microdilution method for colistin, according to EUCAST (www.eucast.org). Genomic DNA was extracted (Wizard Genomic DNA Purification Kit, PROMEGA) and sequenced by Illumina HiSeq, reads were assembled by SPAdes version 3.12.0 (http://cab.spbu.ru/software/spades/) and contigs were further annotated using RAST (http://rast.nmpdr.org/rast.cgi). The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession VORC00000000. The version described in this paper is version VORC01000000.

We used available in silico tools (http://www.genomicepidemiology.org/; http://clermontyping.iame-research.center/) to define serotype, multilocus sequence type (MLST, Achtman scheme), antimicrobial resistance and pathogenicity profiles, *E. coli* phylogenetic group. We searched 49 virulence genes (adhesins, toxins, siderophores, capsule type, protectins, invasins and miscellaneous) known to be involved in extraintestinal *E. coli* infections by PCR (supplementary Table S1), VirulenceFinder 2.0 (https://cge.cbs.dtu.dk/services/VirulenceFinder/) and using in-house databases. Twenty-seven ST83 *E. coli* genomes and associated metadata (when available) retrieved from NCBI and EnteroBase public databases were used for comparison (last accessed February 05, 2019; supplementary Table S2). SNP-based phylogeny was performed using CSI phylogeny 1.4 (https://cge.cbs.dtu.dk/services/CSIPhylogeny/) and represented with Interactive Tree of Life (iTOL, https://itol.embl.de). Duplication of *hlyA* operon was searched in 980 closed genomes from NCBI database from which ExPEC status was inferred by presence of >2 amongst *papAH* and/or *papC, sfa/focDE, afa/draBC, kpsM II* and *iutA* using ncbi-blast-2.8.1+ package (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/).

3. Results and discussion

3.1. hlyA duplication: genetics and distribution amongst ExPEC

Strain Ec_151217 has a total length of 4 967 437 bp and a G+C content of 50.5%. It belongs to phylogenetic group B2, O83:H5:K5 serotype and ST83, and was susceptible to all antibiotics tested. The isolate was predicted as a human pathogen with 466 matched potential pathogenic protein families and had a high virulence score (n = 28/49 virulence genes). Virulence gene profile included genes known to codify for different types of adhesins (*pap, fim, sfa/foc, matB, yfcV*), toxins (*hly,*

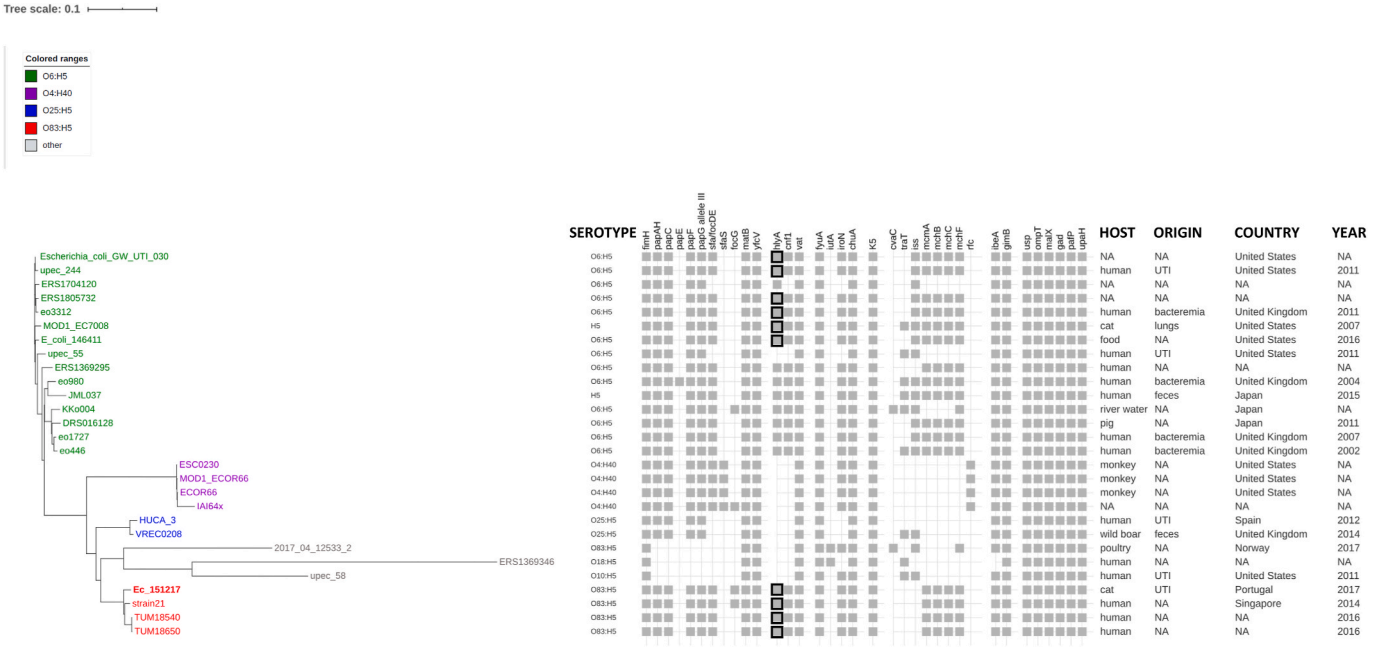


Fig. 1. Phylogeny and virulence gene profiles of ST83 *Escherichia coli* strains deposited in public databases.
Phylogenetic tree was obtained by SNP method using as reference *Escherichia coli* 536 (Accession number NC_008253.1) by CSI Phylogeny with default parameters. This figure shows virulence genes detected based on a cut-off of 90% identity and 60% query coverage to the reference nucleotide sequences. Framed squares represent putative double copies of hemolysin A operon.
Abbreviations: NA = not available, UTI = urinary tract infection.
Definitions are as follows: *fimH*, Type 1 fimbrin D-mannose specific adhesin; *papAH*, Pap fimbrial major pilin protein; *papC*, Outer membrane usher protein *PapC*; *papE*, Fimbrial protein *PapE*; *papF*, Fimbrial adapter *PapF*; *papG allele III*, Fimbrial adhesin *PapG* allele III; *sfa/focDE*, S fimbriae and F1C fimbriae; *sfaS*, S-fimbrial adhesin protein *SfaS*; *focG*, F1C minor fimbrial subunit protein G; *matB*, Common pilus major fimbrillin subunit *EcpA*; *yfcV*, Uncharacterized fimbrial-like protein *YfcV*; *hlyA*, Hemolysin A; *cnf1*, Cytotoxic necrotizing factor 1; *vat*, Vacuolating autotransporter toxin; *fyuA*, Pesticin/yersiniabactin receptor protein; *iutA*, Ferric aerobactin receptor; *iroN*, Catechole siderophore receptor; *chuA*, Outer membrane hemin receptor *ChuA*; *K5*, K5 variant of the group II capsule; *cvaC*, Colicin-V; *traT*, *TraT* complement resistance protein; *iss*, Increased serum survival protein; *mcmA*, Colicin-V; *mchB*, Microcin H47; *mchC*, *MchC* protein; *mchF*, ABC transporter protein *MchF*; *rfc*, O-antigen polymerase; *ibeA*, Invasion protein *IbeA*; *gimB*, Invasin (genetic island associated with newborn meningitis); *usp*, Uropathogenic specific protein; *ompT*, *OmpT* outer membrane protease; *malX*, Pathogenicity-associated island (PAI); *gad*, Glutamate decarboxylase alpha; *pafP*, Transcription regulator, associated with *metV* genomic island; *upaH*, Autotransporter *UpaH*.

cnf1, vat), siderophores (*fyuA, iroN*), invasins (*ibeA, gimB*) amongst others (e.g. *malX, usp, upaH*), many of which (alone or in combination) had been associated with a higher potential to adhere to the urinary epithelium and subsequently cause UTI (Fig. 1) [3].

Interestingly, the Ec_151217 strain contains two copies of the *hlyCABD* operon (7.2 Kb) encoding the toxin hemolysin (*hlyI* and *hlyII*, deposited at NCBI under accession numbers MN826072 and MN826073, respectively). The *hlyCABD* genes are 97,45% homologous and encode for two highly similar HlyA proteins (~97.6%). However, their 1.6 Kb upstream regions are highly dissimilar (~50.6% homology) while showing a conserved *ops* element and 39bp JUMPstart sequence for recognition by RfaH transcriptional regulator [11]. Differences at this region seem to condition differential expression levels of HlyA that might modulate variable virulence profiles at different stages of infection [5,11–13].

One of these HlyA operons and the corresponding upstream sequence (*hlyI*) is the one frequently found alone in a wide range of ExPEC clones including ST131 [11]. The second copy (*hlyII*) was detected in our highly pathogenic Ec_151217 strain, in other highly related ST83 genomes (Fig. 1, see below), and also in other uropathogenic strains (e.g. *E. coli* 536 ST127, *E. coli* J96 ST12) that caused acute pyelonephritis [5,11]. When closed genomes from putative ExPEC deposited at NCBI database were searched, this duplication was detected in only 1 out of 100 (1%) closed genomes, but this low occurrence needs to be confirmed in a larger and more representative population. These data suggest sporadic acquisition of a second *hlyA* operon by particular strains, some of which are known to have been involved in acute and severe infections, thus anticipating a higher risk of a poor UTI outcome. In fact, the presence of these two operons was shown to yield higher virulence than the presence of either *hlyI* or *hlyII* alone in well-known pyelonephritic 536 and J96 strains, compared to strains with only one hlyCABD operon (e.g. CFT073) [5,6,14]. Similar observations were also reported in other bacterial species [15].

Besides *hlyA* and other virulence genes, the presence of *pap*, *cnf1* might have also favoured the severe inflammatory response observed. In fact, *papG* allele III specificity for domestic animals, might be important for the establishment of pyelonephritis in the cat [16,17]. Genes encoding toxins like *hlyA* or *cnf1* (cytotoxic necrotizing factor), known to have necrotizing activity and high lethality in animal models, are frequently detected in *E. coli* strains causing life-threatening infections [3,13,18]. Contrariwise, their absence in closely related ST83 strains was associated with better clinical outcome [19].

Our study provides further genotypic and clinical evidence for the role of these two hemolysin A paralogues on higher virulence and severe clinical outcomes in a new *E. coli* genetic background. Besides, it also highlights the need to understand how often these gene duplication events occur in a larger population and how they condition pathogenesis of particular UTI strains. Finally, to know if there are one or two *hlyA* copies is essential for validation of functional or *in vivo* studies but this duplication might be missed by second-generation whole genome sequencing approaches [9].

3.2. Phylogeny and reservoir of this highly pathogenic E. coli strain

The SNP-based ST83 phylogenetic tree reveals that strains are grouped in four main clusters according to the serotype, most of them exhibiting a similar array of virulence genes (Fig. 1). Interestingly, many of these genomes were identified in different animal species or food suggesting possible transmission through the food chain.

Three other ST83 O83:H5 public genomes from humans were highly related to Ec_151217 (~200 SNP) (Fig. 1, Supplementary Fig. S2), suggesting a common phylogenetic origin, and all of them carried double *hlyA* copy. Highly similar ST83 O83:H5 isolates have also been previously identified causing severe or life-threatening infections in cats in the USA, but, unfortunately, the genomes are not available [18]. A high relatedness was also observed among ST83 O6:H5 genomes (n= 15, ~100–400 SNP; Supplementary Fig. S2), many of them from cat or human infections (including bacteremia), and 6 carried also the double *hlyA* operon. In fact, O6 and O83 are serogroups that had been frequently identified in strains responsible for UTI in domestic animals and humans, suggesting that these strains might be intermingled between these hosts [2,16]. Thus, the high frequency of detection of *hlyA* double operon in O83 and O6 lineages (n = 10/19; 52%) and previous reports [18] support a high pathogenic potential for these lineages and the risk for animal-human transmission.

Interestingly, other diverse and less related ST83 *E. coli* genomes (O4:H40, O25:H5 or others) lacked *cnf1*, *hlyA*, *mchABCF* operon or specific adhesins, and none of them had been recorded as being involved in a severe infection. All together, these data highlight the evolution of serogroup specific lineages within ST83, some of them with higher virulence potential for developing complicated infections in animals and humans, probably associated with the acquisition of a second *hlyA* operon. Particular strains contained genes conferring resistance to critical antibiotics (e.g. bla_{OXA-48}, bla_{KPC-2}) which represents an additional risk for human medicine [19] (data not shown).

4. Conclusions

We highlight, for the first time, the presence of a *hlyA* duplication on a highly pathogenic *E. coli* ST83 strain as well as in closely related genomes from different animal and human hosts from different continents suggesting a potential animal-human transmission of this *E. coli* lineage. We also show that this duplication seems to occur infrequently in the ExPEC population, but future studies are required to elucidate its occurrence in larger *E. coli* collections from different potential reservoirs (animals, food products, humans). This knowledge is of interest to improve strategies for detection and control of highly virulent uropathogenic *E. coli* strains for humans and animals.

Funding

This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is financed by national funds from FCT MCTES [UIDP/04378/2020 and UIDB/04378/2020]. Magdalena Ksiezarek is supported by FCT [SFRH/BD/132497/2017]. Ângela Novais is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0032].

Author statement

Magdalena Ksiezarek: Investigation, Methodology, Formal analysis, Data curation, Visualization, Writing – original draft Ângela Novais: Conceptualization, Validation, Supervision, Writing – review & editing Helena Felga: Methodology, Investigation Fátima Mendes: Methodology, Investigation Marta Escobar: Methodology, Investigation Luísa Peixe: Supervision, Funding acquisition, Resources, Writing - Reviewing and Editing.

Ethical statement

No ethical approval was required for this study.

Data availablility statement

The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession VORC00000000. The version described in this paper is version VORC01000000. *hlyI* and *hlyII* are deposited at NCBI under accession numbers MN826072 and MN826073, respectively.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Abbreviations

SC, subcutaneous; IV, intravenous; SID, once a day; BID, twice a day.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.micpath.2021.104920.

References

[1] J.R. Johnson, B.D. Johnston, S. Porter, P. Thuras, M. Aziz, L.B. Price, Accessory traits and phylogenetic background predict *Escherichia coli* extraintestinal virulence better than does ecological source, J. Infect. Dis. 219 (2019) 121–132.

[2] M. Blanco, J.E. Blanco, M.P. Alonso, J. Blanco, Virulence factors and O groups of *Escherichia coli* isolates from patients with acute pyelonephritis, cystitis and asymptomatic bacteriuria, Eur. J. Epidemiol. 12 (1996) 191–198.

[3] M. Tabasi, M.R.A. Karam, M. Habibi, E. Mostafavi, S. Bouzari, Genotypic characterization of virulence factors in *Escherichia coli* isolated from patients with acute cystitis, pyelonephritis and asymptomatic bacteriuria, J. Clin. Diagn. Res. 10 (2016) DC01–DC07.

[4] R.A. Welch, Uropathogenic *Escherichia coli*-associated exotoxins, Microbiol. Spectr. 4 (2016).

[5] G. Nagy, A. Altenhoefer, O. Knapp, E. Maier, U. Dobrindt, G. Blum-Oehler, R. Benz, L. Emody, J. Hacker, Both alpha-haemolysin determinants contribute to full virulence of uropathogenic *Escherichia coli* strain 536, Microb. Infect. 8 (2006) 2006–2012.

[6] E. Velasco, S. Wang, M. Sanet, J. Fernández-Vázquez, D. Jové, E. Glaría, A. F. Valledor, T.V. O'Halloran, C. Balsalobre, A new role for Zinc limitation in bacterial pathogenicity: modulation of α-hemolysin from uropathogenic *Escherichia coli*, Sci. Rep. 8 (2018) 6535.

[7] H.C. Mirsepasi-Lauridsen, Z. Du, C. Struve, G. Charbon, J. Karczewski, K. A. Krogfelt, A.M. Petersen, J.M. Wells, Secretion of alpha-hemolysin by *Escherichia coli* disrupts tight junctions in ulcerative colitis patients, Clin. Transl. Gastroenterol. 7 (2016) e149.

[8] Y.C. Smith, S.B. Rasmussen, K.K. Grande, R.M. Conran, A.D. O'Brien, Hemolysin of uropathogenic *Escherichia coli* evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice, Infect. Immun. 76 (2008) 2978–2990.

[9] J.F. Denton, J. Lugo-Martinez, A.E. Tucker, D.R. Schrider, W.C. Warren, M. W. Hahn, Extensive error in the number of genes inferred from draft genome assemblies, PLoS Comput. Biol. 10 (2014), e1003998.

[10] A.R. Manges, J.R. Johnson, Reservoirs of extraintestinal pathogenic *Escherichia coli*, Microbiol. Spectr. 3 (2015).

[11] N.T.K. Nhu, M.-D. Phan, B.M. Forde, A.M.V. Murthy, K.M. Peters, C.J. Day, J. Poole, T.J. Kidd, R.A. Welch, M.P. Jennings, G.C. Ulett, M.J. Sweet, S.A. Beatson, M.A. Schembri, Complex multilevel control of hemolysin production by uropathogenic *Escherichia coli*, mBio 10 (2019).

[12] R.A. Welch, S. Falkow, Characterization of *Escherichia coli* hemolysins conferring quantitative differences in virulence, Infect. Immun. 43 (1984) 156–160.

[13] L.C. Ristow, R.A. Welch, Hemolysin of uropathogenic *Escherichia coli*: a cloak or a dagger? Biochim. Biophys. Acta 1858 (2016) 538–545.

[14] E. Brzuszkiewicz, H. Brüggemann, H. Liesegang, M. Emmerth, T. Olschläger, G. Nagy, K. Albermann, C. Wagner, C. Buchrieser, L. Emody, G. Gottschalk, J. Hacker, U. Dobrindt, How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 12879–12884.

[15] X.H. Zhang, P.G. Meaden, B. Austin, Duplication of hemolysin genes in a virulent isolate of *Vibrio harveyi*, Appl. Environ. Microbiol. 67 (2001) 3161–3167.

[16] J.R. Johnson, P. Delavari, A.L. Stell, T.S. Whittam, U. Carlino, T.A. Russo, Molecular comparison of extraintestinal *Escherichia coli* isolates of the same electrophoretic lineages from humans and domestic animals, J. Infect. Dis. 183 (2001) 154–159.

[17] C. Féria, J. Machado, J. Duarte Correia, J. Gonçalves, W. Gaastra, Distribution of *papG* alleles among uropathogenic *Escherichia coli* isolated from different species, FEMS Microbiol. Lett. 202 (2001) 205–208.

[18] X. Liu, K. Thungrat, D.M. Boothe, Multilocus sequence typing and virulence profiles in uropathogenic *Escherichia coli* isolated from cats in the United States, PloS One 10 (2015), e0143335.

[19] M. de Toro, J. Fernández, V. García, A. Mora, J. Blanco, F. de la Cruz, M.R. Rodicio, Whole genome sequencing, molecular typing and *in vivo* virulence of OXA-48-producing *Escherichia coli* isolates including ST131 H30-Rx, H22 and H41 subclones, Sci. Rep. 7 (2017), 12103.

Additional material associated with the manuscript entitled “Phylogenomic analysis of a highly virulent *Escherichia coli* ST83 lineage with potential animal-human transmission” is available through the following links:

Table S1
https://drive.google.com/file/d/1-QKsiHPxH0Q8xt31VtkOeh65uFMIuJCL/view?usp=sharing

Table S2
https://drive.google.com/file/d/11O_1bNFE3Ex1k-4aRB6IE97FicE-O5XG/view?usp=sharing

Figure S1
https://drive.google.com/file/d/1OPx3RspY5HbsIfh44QmvnpSSeyf0NvyB/view?usp=sharing

Figure S2
https://drive.google.com/file/d/120ZoueL4jjw_sMuM9aJQvCQdPamIEGWM/view?usp=sharing

Limosilactobacillus spp. diversity and relevant features of urogenital species

Magdalena Ksiezarek[1,2], Teresa Gonçalves Ribeiro[1,2], Filipa Grosso[1,2], Luisa Peixe[1,2]

[1]UCIBIO – Applied Molecular Biosciences Unit, REQUIMTE, Faculty of Pharmacy, Department of Biological Sciences, Laboratory of Microbiology, University of Porto, 4050-313 Porto, Portugal

[2]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal

Manuscript submitted

ABSTRACT

Limosilactobacillus genus (*Lactobacillaceae* family) contains multiple species often isolated from human microbiome. However, since its recent reclassification from former *Lactobacillus*, several novel species had been described and its genomic particularities have not been explored. Here, we analyzed taxonomic and genomic diversity of *Limosilactobacillus* genus, pangenome of species/strains isolated from human origin and performed comparative genomics for urogenital species including novel *Limosilactobacillus portuensis* and *Limosilactobacillus urinaemulieris*, with the aim to better understand their nature and putative functionality. We present detailed taxonomic overview of *Limosilactobacillus* genus including genomes deposited in databases and demonstrated that public repositories contain genomes of still not described putatively novel species. We performed pangenome characterization of human *Limosilactobacillus* species and deeply investigated genomic content of novel urogenital species regarding their putative bacteriocins, probiotic activity, CRISPR-Cas system, antibiotic resistance and mobilome. We find that urogenital species share approximately 50% of predicted genes which could suggest that conserved functions are likely related to niche adaptation. Furthermore, some shared genes could be relevant for urogenital health maintenance e.g., genes conferring lactic acid production. This data highlights that those species might play a role in urogenital tract homeostasis, and they should be further investigated.

INTRODUCTION

Since several years, the genus *Lactobacillus* (family *Lactobacillaceae*) has been under investigation aiming proper classification of this bacterial group characterized by high phylogenetic and genotypic diversity [1]. In early 2020, taxonomy of the genus *Lactobacillus* was revisited based on polyphasic approach, resulting in reclassification of former *Lactobacillus* species into 25 genera, including *Limosilactobacillus* gen. nov. (formerly *Lactobacillus reuteri* group) [2].

To date, *Limosilactobacillus* comprises 23 validly published species (*Limosilactobacillus agrestis, Limosilactobacillus albertensis, Limosilactobacillus antri, Limosilactobacillus balticus, Limosilactobacillus caviae, Limosilactobacillus coleohominis, Limosilactobacillus equigenerosi, Limosilactobacillus fastidiosus, Limosilactobacillus fermentum, Limosilactobacillus frumenti, Limosilactobacillus gastricus, Limosilactobacillus gorillae, Limosilactobacillus ingluviei, Limosilactobacillus mucosae, Limosilactobacillus oris, Limosilactobacillus panis, Limosilactobacillus pontis, Limosilactobacillus portuensis,*

Limosilactobacillus reuteri, Limosilactobacillus rudii, Limosilactobacillus secaliphilus, Limosilactobacillus urinaemulieris, Limosilactobacillus vaginalis) of which 7 were published early this year [3, 4]. The genus' type strain - *L. fermentum* - and several other *Limosilactobacillus* species are well known from its implication in the food industry (fermentation) and probiotic market [5, 6]. These species are isolated from various sources including human, animals or food products [2].

In fact, some of these species are often found in healthy human microbiome (mainly gut and vagina), with *L. reuteri* being the most known and understood [6]. Other species reported from gastrointestinal tract include *L. fermentum* and *L. mucosae* [7–9], and occasionally *L. vaginalis* [10]. Regarding female urogenital microbiome (FUM), *Limosilactobacillus* have occasionally been identified i.e., *L. mucosae, L. coleohominis*, *L. reuteri*, *L. vaginalis*, *L. portuensis* and *L. urinaermulieris* [9, 11–15]. In fact, healthy human microbiome appears to be colonized in higher extent with *Lactobacillus* than *Limosilactobacillus* species [10, 14]. Nonetheless, it seems that in urogenital tract, species belonging to *Limosilactobacillus* co-exist with *Lactobacillus* species [2, 14]. Thus, *Limosilactobacillus* species seem to be also relevant for human urogenital health and deserves equal attention as their *Lactobacillus* relatives.

The aim of this study was to evaluate the taxonomic diversity of *Limosilactobacillus* genus and investigate genomic diversity of *Limosilactobacillus* species, originating from humans/human sources. We further performed comparative genomics for recently described *Limosilactobacillus portuensis*, *Limosilactobacillus urinaemulieris* and their closest relative *Limosilactobacillus vaginalis,* all isolated from human urogenital tract. To the best of our knowledge, this is the first study since recent *Lactobacillaceae* reclassification, focusing on pangenome and characterizing genomic diversity of *Limosilactobacillus* genus.

MATERIAL AND METHODS

Genomes database

A total of 336 genomes, including complete and draft genomes of 22 *Limosilactobacillus* species (**Supplementary Table S1**), were retrieved from the National Center for Biotechnology Information (NCBI) Assembly database on 6th September 2020. Additionally, 2 draft genomes representing recently described by our group *Limosilactobacillus* species were included in the analysis, namely c9Ua_26M^{T} (*L. urinaemulieris*) and c11Ua_112M^{T} (*L. portuensis*) [3].

Six genomes were excluded from further analysis based on worse assembly statistics since there was better representative assembly available for the same strains (**Supplementary Table S1**). Metadata related to downloaded assemblies was retrieved from NCBI BioSample database.

Average Nucleotide Identity and phylogenomic analysis

Average Nucleotide Identity based in BLAST+ (ANIb) analysis on 332 assemblies was performed with pyani (v0.2) [16]. Percentage identity matrix (**Supplementary Table S2)** was used to create heatmap representing ANIb clusters by pheatmap v1.0.12 package in R version 3.6.2 [17]. Phylogenomic analysis was performed using anvi'o v6.2. [18]. Single-copy core genes based on Bacteria_71 collection from hidden Markov model (HMM) profiles [19] were identified and 37 genes encoding ribosomal proteins were concatenated. FastTree version 2.1.11 [20] was used to create maximum likelihood phylogenomic tree with Jones-Taylor-Thorton substitution model, local support of SH-like 1000 and CAT approximation with 20 rate categories. The resulted phylogenomic tree was edited in iTOL [21].

Pangenome analysis

Nucleotide FASTA files were pretreated for compatibility with anvi'o v6.2. pipeline [18]. Anvi'o script was used to profile HMMs and genes prediction was performed by Prodigal [22]. Additionally, protein functions were annotated by Clusters of Orthologous Groups (COGs) database using anvi'o script. Pangenome analysis was performed by anvi'o, using MUSCLE for sequence alignment [23] and clustering with Markov Cluster algorithm [24]. Genes' collections were identified by applying filters into gene clusters search: core genome included gene clusters present in 100% of genomes; accessory genome was represented with three collections: softcore (gene clusters present in more than 95%), dispensable (gene clusters present in at least 2 genomes and in less than 95% of genome) and unique (singletons; genes present in just one unique genome). Pangenome was visualized by anvi'o and edited in Inkscape [25]. COGs categories distribution among strains was visualized in R v3.6.2 [17] ggplot2 package version 3.3.2 [26]. In case of multiple COG categories predicted per gene, the first one was used as the most significant hit.

Other WGS analysis

Additional analysis of the draft genomes of strains *L. urinaemulieris* c9Ua_26M^T and *L. portuensis* c11Ua_112M^T was performed using ResFinder v. 4.1, and PathogenFinder v. 1.1

to investigate the presence of acquired genes mediating antimicrobial resistance, and pathogenicity, respectively [27, 28]. BAGEL4 was used to predict in silico bacteriocin-encoding genes [29] and the presence of genes associated with lactic acid and hydrogen peroxide production i.e., *ldhA*, *ldh* and *poxL* were search by blastp using ncbi-blast-2.8.1+ package [30]. Identification of prophage sequences was determined using PHASTER (the Phage Search Tool Enhanced Release) [31]. Additionally, mobile genetic elements (MGEs) were investigated by the MobileElementFinder (v1.0.3) [32]. Coding sequences for Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated genes (Cas) were identified using CRISPRCasFinder v. 1.1.2—I2BC [33]. Only CRISPR showing an evidence level of 3 or 4 were considered.

RESULTS

Overview of *Limosilactobacillus* species diversity and taxonomy

A total of 332 genomes representing 22 *Limosilactobacillus* species were analyzed in this study (**Table 1, Supplementary Table S1**). At the date of genome retrieval, the NCBI Assembly database was lacking representative genome for *L. caviae*. Complete (55, 16.6%) and draft genomes (277, 83.4%) reveal that most of the strains were isolated from animals (n=167 genomes, 50.3%), followed by humans (n=74, 22.3%), and food products (n=69, 17.8%). *L. reuteri* and *L. fermentum* strains were the most sequenced, representing 60.2% and 23.8% of all available *Limosilactobacillus* genomes, respectively (**Table 1**).

Analysis of ANIb showed clear *Limosilactobacillus* species separation based on the widely accepted threshold of 95% for species discrimination [34] (**Fig. 1., Supplementary Table S2**). We observed near cut-off (94-95%) variations in ANIb values inside *L. reuteri* clade suggesting the existence of several subspecies as recently characterized by Li *et al.*, 2021 [4]. Moreover, publicly available genomes of strains VA24_5, Lr4000, W1P44.042, W1P28.032 deposited as *L. reuteri* and strain UMB0683 deposited as *L. pontis* should be reclassified since ANIb values were all below 95%. Strains VA24_5 and Lr4000 present ANIb values of ~82% with *L. reuteri* DSM 20016^T, and >95% with recently published *L. albertensis* Lr3000^T [4]. Strain W1P44.042 deposited as *L. reuteri* should be also reclassified as *L. mucosae* based on the ANIb value of 96% between strain W1P44.042 and the *L. mucosae* DSM 13345^T (**Fig. 1., Supplementary Table S2**). ANIb analysis also revealed that the most closely related species of strains W1P28.032 and UMB0683 was *L. pontis*. However, the ANIb value between strains W1P28.032 and UMB0683 was 73%, and between each strain and *L.*

pontis DSM 8475^T was ~82% and 84%, respectively, suggesting that these strains represent distinct and putative novel species.

To confirm the taxonomic position of strains that should be reclassified, we performed phylogenomic analysis based on 37 single-copy core ribosomal proteins including type strains of each species and putative novel *Limosilactobacillus* species (**Fig. 2**). This tree showed clear distinction between genomes of different species. Moreover, strains VA24_5 and Lr4000 cluster with the type strain of *L. albertensis*, strain W1P44.042 grouped with the type strain of *L. mucosae*, and strains W1P28.032 and UMB0683 form each an independent branch.

On the basis of this core genome analysis and ANIb values, we propose that strains VA24_5 and Lr4000 should be classified as members of the species *L. albertensis* and W1P44.042 strain classified as member of *L. mucosae*. Furthermore, 2 putative novel *Limosilactobacillus* species were identified.

Pangenome of *Limosilactobacillus* isolated from humans

We identified 74 *Limosilactobacillus* genomes isolated from humans, which included 11 *Limosilactobacillus* members (*L. albertensis*, *L. antri, L. coleohominis, L. fermentum, L. gastricus, L. mucosae, L. oris, L. portuensis, L. reuteri*, *L. urinaemulieris*, *L. vaginalis*) and the putative novel species represented by strain UMB0683, and here designated as *Limosilactobacillus* sp. nov. (**Table 2**). The average genome length was 2,039,809 bp, with a mean of 2056 genes predicted per genome. We found the highest number of assembled genomes belonged to *L. fermentum* (n= 36) and *L. reuteri* (n= 17), while remaining species had relatively poor representation (1-4 genome each). Most strains were isolated from gastrointestinal-related samples (n=31), followed by urogenital origin (n=9 vagina, n=7 urine), oral cavity (n=6) or breast milk (n=5) (**Supplementary Table S1**).

Overall, the *Limosilactobacillus* pangenome was represented by 10499 clusters of orthologous groups (COGs) of proteins (151749 genes), with 5052 clusters of known function (**Fig. 3**). A high genomic variability was observed, with the core genome represented only by 453 gene clusters (37033 genes). Within the accessory genome (10046 gene clusters, 114716 genes), we can distinguish three representative collections: softcore genome represented by 316 gene clusters (24553 genes), dispensable genome with 5967 gene clusters (86215 genes) and singletons with 3768 clusters (3948 genes). Most of the gene clusters in core genomes have known COG function (97.8%), contrary to accessory where most have unknown COG function (54%). Noteworthy, this pangenome analysis is sorted by frequencies of gene cluster

which incorporated 2 other species i.e., *L. albertensis* VA24_5 and *Limosilactobacillus* sp. nov. UMB0683 (**Fig. 1., Fig. 2.**) in the middle of *L. reuteri* clade (**Fig. 3.**).

The distribution of genes belonging to defined COG functional categories is presented in **Fig. 4**. **Fig. 4.A.** presents core and accessory genes grouped by COG functional categories, while **Fig. 4.B.** visualizes COG categories of accessory genome collections i.e., softcore, dispensable and singleton genes. Most of the genes in core genes collection (total 37033) were involved into translation, ribosomal structure and biogenesis (category J, 8236 genes), replication, recombination and repair (category L, 3183 genes), nucleotide transport and metabolism (category F, 3172 genes), amino acid transport and metabolism (category E, 2440 genes) and carbohydrate transport and metabolism (category G, 2311 genes). Interestingly, distribution of COG functional categories for softcore (**Fig. 4.B.**) demonstrated nearly the same functional pattern as core genes collection (**Fig. 4.A.**). Genes of known function representing singletons were associated with cell wall/membrane/envelope biogenesis (category M, 130 genes), carbohydrate transport and metabolism (category G, 113 genes), transcription (category K, 110 genes), replication, recombination and repair (category L, 93 genes) and mobilome: prophages, transposons (category X, 86 genes).

Pangenome of urogenital *Limosilactobacillus* species

We performed genomic comparisons of the two recently described urogenital species i.e., *L. urinaemulieris* and *L. portuensis*, together with their closely related *L. vaginalis*. Total of 5 genomes was available for comparison: 3 genomes representing *L. vaginalis*, 1 *L. urinaemulieris* and 1 *L. portuensis*. All but one strain (*L. vaginalis* LV515 from woman with trichomoniasis) were obtained from healthy women. Overall, 2716 clusters of orthologous groups of proteins were identified (9817 genes), with 1748 of known function**.** High strains relatedness was observed with core genes representing 49.6% of total gene clusters (1347 gene clusters, 7222 genes). Additionally, 88 gene clusters were identified as softcore genome, 483 gene clusters as dispensable genes and 798 gene clusters were identified as singletons (29.4%). Exact number of genes classified as core and accessory genome per each strain is presented in **Table 3**.

Distribution of COG categories for all detected genes among 5 strains is presented in **Fig. 5.** COGs functional groups were distributed relatively similar among 5 strains, with the highest number of genes with unclassified functionality (**Fig. 5.A.**). On average, 28% of genes found were unclassified or have unknown function. Most of the genes with known function (**Fig. 5.B.**) were involved in translation, ribosomal structure and biogenesis (mean 9.3%),

amino acid transport and metabolism (mean 7.1%), transcription (mean 5.5%), cell wall biogenesis (mean 5.3%), replication, recombination and repair (5.2%) or carbohydrate transport and metabolism (mean 4.8%).

Genomic characteristics specific to *L. urinaemulieris* and *L. portuensis*

To understand if the novel *Limosilactobacillus* urogenital species can provide benefits to niche they are inhabiting, we investigated genomic content that may be related to their putative probiotic activities such as lactic acid, hydrogen peroxide and bacteriocins production. Both type strains, *L. urinaemulieris* and *L. portuensis* possess genes required for D- and L-lactic acid synthesis i.e., *ldhA* and *ldh*, respectively. However, genes coding for pyruvate oxidase that is required to produce hydrogen peroxide were not found in the genomes of those strains. Additionally, *L. urinaemulieris* c9Ua_26M^T is not capable to produce any known bacteriocin, however in *L. portuensis* c11Ua_112M^T we found genes encoding for bacteriocin Enterolysin A. Nevertheless, appropriate phenotypic tests should be performed to confirm these in silico findings.

We further investigated the genetic content that distinguishes these urogenital strains/species. We found 164 unique genes for *L. urinaemulieris*, and 394 unique genes for *L. portuensis*, most of which were unclassified or have unknown function (119 and 296 genes, respectively) (**Supplementary Table S3**). Functional enrichment analysis resulted in 49 hits for *L. portuensis*, and 15 for *L. urinaemulieris*, however none of these COG functions were found to be enriched on significant level ($p < 0.05$) (**Supplementary Table S4**).

The largest number of known functions enriched in *L. portuensis* was related to defense mechanisms (COG category V; multidrug transporter EmrE, multidrug efflux pump subunit AcrB, toxin component of the Txe-Axe toxin-antitoxin module - Txe/YoeB family, antitoxin component of the RelBE or MazEF), and carbohydrate transport and metabolism mechanisms [COG category G; arabinose metabolism (L-arabinose isomerase Ara, L-ribulose-5-phosphate 4-epimerase AraD, phosphoribulokinase PrkB), alpha-L-fucosidase AfuC, teichoic acids export ATP-binding protein TagH].

The few hits found to be enriched in *L. urinaemulieris* were also mostly associated with defense mechanisms (COG category V; CRISPR-Cas system, and DNA methylase subunit YeeA). To investigate CRISPR loci and *cas* genes identification in more depth, the genome sequence of the *L. urinaemulieris* strain c9Ua_26M^T was uploaded to CRISPRcasFinder, and *cas* gene detection was performed via subtyping clustering model option. Strain c9Ua_26M^T harbored a CAS-Type IIC, and one confirmed CRISPR loci with a total of 7 spacers

(sequence length: 29 bp) (**Fig. 6**). Spacers in the CRISPR-Cas loci originate from foreign invaders and bacteriophages are the most common threats for bacteria, and in the case of a bacterial strain being invaded by a phage, the spacer sequences may contain a fragment corresponding to the phage genome [35]. In this sense, we further explored the origins of these foreign DNA fragments by performing a BLAST search of the extracted spacer sequences against the NCBI virus Refseq database. However, the nucleic acid source of the spacer sequences remains unknow as no positive hits were identified.

Pathogenicity, antibiotic resistance, and MGE in *L. urinaemulieris* and *L. portuensis*

Prediction of pathogenicity assessed with PathogenFinder tool revealed that calculated Matched Pathogenic Families for *L. urinaemulieris* c9Ua_26M^T and *L. portuensis* c11Ua_112M^T was 0, the Matched Not Pathogenic Families was 8 and 13, and the probability of the strains being a human pathogen was 0.175 and 0.212 (respectively). Thus, the strains were predicted as non-pathogenic to human.

Although acquired antibiotic resistance genes were not detected, different MGEs were identified within the genomes of *L. urinaemulieris* c9Ua_26M^T and *L. portuensis* c11Ua_112M^T (**Table 4**). Two predicted intact prophage regions were identified in the genome of *L. urinaemulieris* c9Ua_26M^T. The intact regions correspond to a *Lactobacillus* temperate phage phi g1e (GenBank Accession No. NC_004305), and a *Lactobacillus delbrueckii* subsp. *lactis* phage BK5-T (GenBank Accession No. NC_002796). MobileElementFinder identified two complete Insertion Sequences (ISs) belonging to IS*30* family.

Five prophage regions, two questionable and three incompletes, were detected in the genome of *L. portuensis* c11Ua_112M^T (**Table 4**). One complete IS belonging to IS*30* family, and a *repUS19* gene was identified by MobileElementFinder, the last sharing 100% similarity to that encoded in the cryptic plasmid pDLK3 (1.365 bp) from a *Staphylococcus aureus* isolate from Czech Republic (GenBank Accession No. GU562626) [36]. In addition, the screening of the NCBI databases (March 23, 2021) revealed 3 plasmids with ≥ 97.95% identical *repUS19* in other Gram-positive bacteria (*Staphylococcus* spp. and *Listeria monocytogenes*). The search for a complete pDLK3 in the draft genome of *L. portuensis* c11Ua_112M^T by MyDbFinder v 2.0 (https://cge.cbs.dtu.dk/services/MyDbFinder/) reveal that *L. portuensis* carried a 1.339 bp pDLK3 (GenBank Accession No. JABUXQ010000061.1), 26 bp less than the pDLK3 from the *S. aureus*.

DISCUSSION

In this study we report genomic features of 22 species classified to the recently proposed novel *Limosilactobacillus* genus [2].

NCBI public database contains over 300 genomes from *Limosilactobacillus* strains of various origin and mostly represented with draft genomes. The number of complete genomes is still low and reserved mostly to species with well explored beneficial activities, such as *L. reuteri* and *L. fermentum* [37]. Thus, for other species, comprehensive detection of certain genomic characteristics (e.g., MGEs) can be challenging.

Overall, our genome-based characterization, including ANIb and phylogenomic analysis, revealed that 2 strains (VA24_5 and Lr4000) should be reclassified as members of the species *L. albertensis*, 1 strain (W1P44.042) as member of *L. mucosae*, and 2 putative novel species (strains W1P28.032 and UMB0683) that should be further characterized to confirm it is potential as novel members of *Limosilactobacillus* genus (**Fig. 2**). In this sense, we highlight the need for a comprehensive revision of *Limosilactobacillus* content in public databases that still contain species that require reclassification and putative novel bacterial species which have not yet been described, as previously observed [38].

In our pangenome analysis, the core genome was represented by a higher number of genes than previously reported [39, 40]. However, previous studies analyzed former *Lactobacillus* members (before taxonomic reclassification), thus their genomes collection were more diverse than in our study, since we investigated only genomes of *Limosilactobacillus* strains isolated from humans. Nevertheless, the accessory genome for the 74 strains included 10046 gene clusters, which highlights a high species/strain genomic specificity. Although most of accessory genes have unknown function, a large part was involved in metabolism, cellular processes and signaling or MGEs. On the other hand, the highest number of core genes was related to housekeeping functions such as translation, ribosomal structure and biogenesis, and replication, recombination and repair, which was similar to previous observations for related species [8, 40–42], and played vital roles for bacterial survival. Moreover, genes related to information storage and processing had been previously suggested as more conserved than other functional groups [43].

While the majority of *Limosilactobacilus* genomic studies have focused on species most relevant to the industry e.g., probiotic market [44–47], we focused our analysis on *Limosilactobacillus* species found in the human urogenital microbiome. We observed that their core genes correspond to approximately 50% of all genes predicted, which is higher than in other inter-species studies [39, 43, 48] and more similar to intra-species genomic

comparisons [49–51], suggesting a high relatedness and common features of these urogenital species, likely promoting niche adaptation. Additionally, based on gene content, *L. portuensis* and *L. urinaemulieris* could produce D- and L-lactic acid, which similarly to related *Lactobacillaceae* species could contribute to maintenance of vaginal health [52]. *L. portuensis* also possess the gene encoding for enterolysin A which can be found in some *Lactobacillus* species [53] and can inhibit growth of certain Gram-positive bacteria, including *Lactobacillaceae* [54, 55]. Of note, we also found those species in vaginal microbiota of several asymptomatic women (unpublished data).

With further analysis, we observed some enriched functions in *L. portuensis* and/or L. *urinaemulieris* which support the phenotypic findings of Ksiezarek *et al*. for these species [3]. For instance, *L. portuensis* can use L-arabinose, while *L. urinaemulieris* cannot; only the *L. portuensis* genome encode a L-*arabinose* isomerase (*araA*), L-ribulose-5-phosphate 4-epimerase (*araD*), and L-ribulokinase. Moreover, different enriched functions associated with defense mechanisms were identified in these species. *L. portuensis* genome encode a multidrug transporter EmrE that confers resistance to a wide range of disinfectants and dyes known as quaternary cation compounds [56], and an efflux pump subunit AcrB that not only confers antibiotic resistance but was also demonstrated to protect *Escherichia coli* from the bile salts and fatty acids, thus influencing fitness responses [57]. Three type II toxin–antitoxin systems (*yefM-yoeB*, *mazEF*, and *relBE*) consisting of a toxic protein and its cognate antitoxin protein, the latter counteracting the toxicity of the former [58], known to interfere with essential cellular processes and being implicated in bacterial lifestyle adaptations such as persistence and the biofilm formation were identified in *L. portuensis*.

The Type II CRISPR-Cas system, a well-known molecular mechanism that provides adaptive immunity against exogenous genetic elements such as bacteriophages and plasmids in bacteria [59], was only found in the recently described *L. urinaemulieris* c9Ua_26M^T, and other few *Limosilactobacillus* species (*L. fermentum* and *L. mucosae*) [9, 47]. The presence of a CRISPR-Cas system indicates an advantage in promoting *L. urinaemulieris* genome stability by acting as a barrier to entry of foreign DNA elements. Moreover, *L. urinaemulieris* c9Ua_26M^T might be a main candidate for gene editing and cleavage of lytic bacteriophages in food industry, yet the functionality of the identified type II-A CRISPR-Cas system must be investigated.

Notably, a lack of antibiotic resistance genes was observed in *L. urinaemulieris* and *L. portuensis*, revealing that these strains might be susceptible to different antibiotic classes. Different MGEs seem to be contributing to *L. urinaemulieris* and *L. portuensis* genome

plasticity. IS elements from IS30 family identified in these novel species, are also commonly detected in prokaryotic genomes, yet and to the best of our knowledge, IS*Lp1* or IS*Lhe30* were only identified in two other *Limosilactobacillus* species, namely *L. fermentum* or *L. fermentum* and *L. reuteri*, respectively [47, 60]. Cryptic plasmids have been previously identified in *L. fermentum* and *L. reuteri* [61], and in this study in *L. portuensis*, but still there is a lack of information on biological advantages they provide to the bacteria. Thus, further studies are needed to better understand their function. Prophages are a common feature among prokaryotic genomes, including in *Lactobacillaceae* [62, 63]. In this study, different complete, questionable, and incomplete prophage region were reported *Limosilactobacillus* strains isolated from voided urine of asymptomatic women, which might suggest that human urogenital tract is also a reservoir of bacteriophages, as observed in the human gut [64].

CONCLUSIONS

Here, we investigated genomic diversity of recently reclassified *Limosilactobacillus* genus. We demonstrated existence of putative novel species in the public database, yet not validly described. Furthermore, we observed relatively small number of core genes for strains of human origin which suggests high species and strain-specificity within this genus. Accessory genome comprising various fitness and mobilome related genes likely represent adaptation and survival skills of these species.

We also demonstrated unique genomic features of recently published *L. urinaemulieris* and *L. portuensis* isolated from voided urine of healthy women. Their specific content, e.g., involved in defense mechanisms, suggests that they may play a role in urogenital tract homeostasis. Furthermore, they showed high level of genomic relatedness and conserved functions.

However, further sequencing to enlarge genomes collection and availability of complete genomes will contribute to better characterization of specific genomic structures and mobile genetic elements, in particular for urogenital *Limosilactobacillus* species to better evaluate their contribution to urogenital health.

AUTHOR STATEMENTS

Funding information

This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is financed by national funds from FCT (UIDP/04378/2020 and UIDB/04378/2020). MK is

supported by FCT [SFRH/BD/132497/2017]. FG is supported by national funds through FCT in the context of the transitional norm [DL57/2016/CP1346/CT0034].

Contributions

MK performed genomic analysis including ANIb, phylogenomic analysis, pangenome-related analyses; performed visualizations and wrote the manuscript. TGR characterized specific features of urogenital species, mobilome, genotypic antimicrobial resistance, revised and edited the manuscript. FG contributed to manuscript revision and editing. LP contributed to reviewing the manuscript, project design and administration, and funding.

Acknowledgements

We would like to thank Svetlana Ugarcina Perovic and Joana Rocha for their indispensable contribution to Female Urogenital Microbiome Project.

Conflicts of interest

The authors declare that there are no conflicts of interest.

ABBREVIATIONS

ANI, Average Nucleotide Identity; Cas, CRISPR-associated genes; COG, clusters of orthologous groups; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; HMM, hidden Markov model; IS, Insertion Sequence; LAB, lactic acid bacteria; MGE, mobile genetic element; NCBI, National Center for Biotechnology Information; WGS, whole genome sequencing.

REFERENCES

1. Salvetti E, Harris HMB, Felis GE, O'Toole PW. Comparative Genomics of the Genus *Lactobacillus* Reveals Robust Phylogroups That Provide the Basis for Reclassification. *Appl Environ Microbiol*;84. Epub ahead of print 01 2018. DOI: 10.1128/AEM.00993-18.
2. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, *et al.* A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 2020;2782–2858.
3. Ksiezarek M, Ribeiro TG, Rocha J, Grosso F, Perovic SU, *et al. Limosilactobacillus urinaemulieris* sp. nov. and *Limosilactobacillus portuensis* sp. nov. isolated from urine

of healthy women. *International Journal of Systematic and Evolutionary Microbiology* 2021;71:004726.

4. Li F, Cheng CC, Zheng J, Liu J, Quevedo RM, *et al. Limosilactobacillus balticus* sp. nov., *Limosilactobacillus agrestis* sp. nov., *Limosilactobacillus albertensis* sp. nov., *Limosilactobacillus rudii* sp. nov. and *Limosilactobacillus fastidiosus* sp. nov., five novel *Limosilactobacillus* species isolated from the vertebrate gastrointestinal tract, and proposal of six subspecies of *Limosilactobacillus reuteri* adapted to the gastrointestinal tract of specific vertebrate hosts. *Int J Syst Evol Microbiol*;71. Epub ahead of print February 2021. DOI: 10.1099/ijsem.0.004644.
5. Naghmouchi K, Belguesmia Y, Bendali F, Spano G, Seal BS, *et al. Lactobacillus fermentum*: a bacterial species with potential for food preservation and biomedical applications. *Crit Rev Food Sci Nutr* 2020;60:3387–3399.
6. Mu Q, Tavella VJ, Luo XM. Role of *Lactobacillus reuteri* in Human Health and Diseases. *Front Microbiol* 2018;9:757.
7. Park MR, Shin M, Mun D, Jeong S-Y, Jeong D-Y, *et al.* Probiotic *Lactobacillus fermentum* strain JDFM216 improves cognitive behavior and modulates immune response with gut microbiota. *Scientific Reports* 2020;10:21701.
8. Valeriano VDV, Oh JK, Bagon BB, Kim H, Kang D-K. Comparative genomic analysis of *Lactobacillus mucosae* LM1 identifies potential niche-specific genes and pathways for gastrointestinal adaptation. *Genomics* 2019;111:24–33.
9. Jia Y, Yang B, Ross P, Stanton C, Zhang H, *et al.* Comparative Genomics Analysis of *Lactobacillus mucosae* from Different Niches. *Genes (Basel)*;11. Epub ahead of print 14 2020. DOI: 10.3390/genes11010095.
10. Pasolli E, De Filippis F, Mauriello IE, Cumbo F, Walsh AM, *et al.* Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nature Communications* 2020;11:2610.
11. Coorevits L, Heytens S, Boelens J, Claeys G. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. *Eur J Clin Microbiol Infect Dis* 2017;36:635–639.
12. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, *et al.* Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. *Journal of Clinical Microbiology* 2014;52:871–876.
13. Puglisi E, Galletti S, Sagheddu V, Elli M. Genome Sequence Announcement of *Lactobacillus vaginalis* LMG S-26419, Isolated from a Healthy Woman. *Microbiology*

Resource Announcements;8. Epub ahead of print February 2019. DOI: 10.1128/MRA.01534-18.

14. Ksiezarek M, Ugarcina-Perovic S, Rocha J, Grosso F, Peixe L. Long-term stability of the urogenital microbiota of asymptomatic European women. *BMC Microbiol* 2021;21:64.
15. Nam H, Whang K, Lee Y. Analysis of vaginal lactic acid producing bacteria in healthy women. *J Microbiol* 2007;45:515–520.
16. 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* 2015;8:12–24.
17. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2018).
18. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology* 2021;6:3–6.
19. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 2019;35:4162–4164.
20. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* 2009;26:1641–1650.
21. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 2019;47:W256–W259.
22. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;11:119.
23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004;32:1792–1797.
24. van Dongen S, Abreu-Goodger C. Using MCL to extract clusters from networks. *Methods Mol Biol* 2012;804:281–295.
25. Inkscape Project. Inkscape [Internet]. 2020. Available from: https://inkscape.org
26. Wickham H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*. https://ggplot2.tidyverse.org (2016).
27. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;75:3491–3500.

28. Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O. PathogenFinder--distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One* 2013;8:e77302.
29. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, *et al.* BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research* 2018;46:W278–W281.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
31. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16-21.
32. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, *et al.* Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *Journal of Antimicrobial Chemotherapy* 2021;76:101–109.
33. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, *et al.* CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* 2018;46:W246–W251.
34. Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M, *et al.* A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* 2014;64:384–391.
35. Wang G, Liu Q, Pei Z, Wang L, Tian P, *et al.* The Diversity of the CRISPR-Cas System and Prophages Present in the Genome Reveals the Co-evolution of *Bifidobacterium pseudocatenulatum* and Phages. *Front Microbiol*;11. Epub ahead of print 26 May 2020. DOI: 10.3389/fmicb.2020.01088.
36. Kuntová L, Pantůček R, Rájová J, Růžičková V, Petráš P, *et al.* Characteristics and distribution of plasmids in a clonally diverse set of methicillin-resistant *Staphylococcus aureus* strains. *Arch Microbiol* 2012;194:607–614.
37. Cárdenas N, Laiño JE, Delgado S, Jiménez E, Juárez del Valle M, *et al.* Relationships between the genome and some phenotypical properties of *Lactobacillus fermentum* CECT 5716, a probiotic strain isolated from human milk. *Appl Microbiol Biotechnol* 2015;99:4343–4353.

38. Wittouck S, Wuyts S, Meehan CJ, van Noort V, Lebeer S. A Genome-Based Species Taxonomy of the *Lactobacillus* Genus Complex. *mSystems*;4. Epub ahead of print 3 September 2019. DOI: 10.1128/mSystems.00264-19.
39. Sun Z, Harris HMB, McCann A, Guo C, Argimón S, *et al.* Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nature Communications* 2015;6:8322.
40. Kant R, Blom J, Palva A, Siezen RJ, de Vos WM. Comparative genomics of *Lactobacillus*. *Microb Biotechnol* 2011;4:323–332.
41. Canchaya C, Claesson MJ, Fitzgerald GF, van Sinderen D, O'Toole PW. Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. *Microbiology (Reading)* 2006;152:3185–3196.
42. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, *et al.* Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* 2006;103:15611–15616.
43. Inglin RC, Meile L, Stevens MJA. Clustering of Pan- and Core-genome of *Lactobacillus* provides Novel Evolutionary Insights for Differentiation. *BMC Genomics* 2018;19:284.
44. Chen L, Gu Q, Li P, Chen S, Li Y. Genomic analysis of *Lactobacillus reuteri* WHH1689 reveals its probiotic properties and stress resistance. *Food Sci Nutr* 2019;7:844–857.
45. Xu S, Cheng J, Meng X, Xu Y, Mu Y. Complete Genome and Comparative Genome Analysis of *Lactobacillus reuteri* YSJL-12, a Potential Probiotics Strain Isolated From Healthy Sow Fresh Feces. *Evol Bioinform Online* 2020;16:1176934320942192.
46. Yoo D, Bagon BB, Valeriano VDV, Oh JK, Kim H, *et al.* Complete genome analysis of *Lactobacillus fermentum* SK152 from kimchi reveals genes associated with its antimicrobial activity. *FEMS Microbiology Letters*;364. Epub ahead of print 2 October 2017. DOI: 10.1093/femsle/fnx185.
47. Brandt K, Nethery MA, O'Flaherty S, Barrangou R. Genomic characterization of *Lactobacillus fermentum* DSM 20052. *BMC Genomics* 2020;21:328.
48. Mendes-Soares H, Suzuki H, Hickey RJ, Forney LJ. Comparative functional genomics of *Lactobacillus* spp. reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. *J Bacteriol* 2014;196:1458–1470.
49. Wegmann U, MacKenzie DA, Zheng J, Goesmann A, Roos S, *et al.* The pan-genome of *Lactobacillus reuteri* strains originating from the pig gastrointestinal tract. *BMC Genomics*;16. Epub ahead of print 1 December 2015. DOI: 10.1186/s12864-015-2216-7.

50. Zhang W, Wang J, Zhang D, Liu H, Wang S, *et al.* Complete Genome Sequencing and Comparative Genome Characterization of *Lactobacillus johnsonii* ZLJ010, a Potential Probiotic With Health-Promoting Properties. *Frontiers in Genetics* 2019;10:812.
51. Son S, Oh J-D, Lee SH, Shin D, Kim Y. Comparative genomics of canine *Lactobacillus reuteri* reveals adaptation to a shared environment with humans. *Genes Genomics* 2020;42:1107–1116.
52. Pino A, Bartolo E, Caggia C, Cianci A, Randazzo CL. Detection of vaginal lactobacilli as probiotic candidates. *Scientific Reports* 2019;9:3355.
53. de Jesus LCL, Drumond MM, Aburjaile FF, Sousa T de J, Coelho-Rocha ND, *et al.* Probiogenomics of *Lactobacillus delbrueckii subsp. lactis* CIDCA 133: In Silico, In Vitro, and In Vivo Approaches. *Microorganisms*;9. Epub ahead of print 14 April 2021. DOI: 10.3390/microorganisms9040829.
54. Khan H, Flint SH, Yu P-L. Determination of the mode of action of enterolysin A, produced by *Enterococcus faecalis* B9510. *J Appl Microbiol* 2013;115:484–494.
55. Nilsen T, Nes IF, Holo H. Enterolysin A, a Cell Wall-Degrading Bacteriocin from *Enterococcus faecalis* LMG 2333. *Appl Environ Microbiol* 2003;69:2975–2984.
56. Saleh M, Bay DC, Turner RJ. Few Conserved Amino Acids in the Small Multidrug Resistance Transporter EmrE Influence Drug Polyselectivity. *Antimicrob Agents Chemother*;62. Epub ahead of print August 2018. DOI: 10.1128/AAC.00461-18.
57. Li X-Z, Nikaido H. Efflux-mediated drug resistance in bacteria: an update. *Drugs* 2009;69:1555–1623.
58. Fraikin N, Goormaghtigh F, Van Melderen L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *J Bacteriol*;202. Epub ahead of print 11 March 2020. DOI: 10.1128/JB.00763-19.
59. Chylinski K, Makarova KS, Charpentier E, Koonin EV. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* 2014;42:6091–6105.
60. Alayande KA, Aiyegoro OA, Nengwekhulu TM, Katata-Seru L, Ateba CN. Integrated genome-based probiotic relevance and safety evaluation of *Lactobacillus reuteri* PNW1. *PLoS One*;15. Epub ahead of print 20 July 2020. DOI: 10.1371/journal.pone.0235873.
61. Fang F, Flynn S, Li Y, Claesson MJ, van Pijkeren J-P, *et al.* Characterization of endogenous plasmids from *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 2008;74:3216–3228.
62. Davidson BE, Kordias N, Dobos M, Hillier AJ. Genomic organization of lactic acid bacteria. *Antonie Van Leeuwenhoek* 1996;70:161–183.

63. Mercanti DJ, Carminati D, Reinheimer JA, Quiberoni A. Widely distributed lysogeny in probiotic lactobacilli represents a potentially high risk for the fermentative dairy industry. *Int J Food Microbiol* 2011;144:503–510.
64. Manrique P, Dills M, Young MJ. The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses*;9. Epub ahead of print 8 June 2017. DOI: 10.3390/v9060141.

Table 1. List of *Limosilactobacillus* species, number of genomes and origin available in NCBI database.

Species	**No of genomes**	**Host**
Limosilactobacillus agrestis	2	Rodents
Limosilactobacillus albertensis	2	Rodents, Lemur
Limosilactobacillus antri	2	Homo sapiens
Limosilactobacillus balticus	2	Rodents, Pheasant
Limosilactobacillus coleohominis	2	Homo sapiens
Limosilactobacillus equigenerosi	3	Horse
Limosilactobacillus fastidiosus	2	Rodents
Limosilactobacillus fermentum	79	Homo sapiens, cheese, fermented food products
Limosilactobacillus frumenti	3	Sourdough
Limosilactobacillus gastricus	3	Homo sapiens
Limosilactobacillus gorillae	1	Gorilla
Limosilactobacillus ingluviei	3	Pigeon
Limosilactobacillus mucosae	13	Homo sapiens, Pig, Boar, Cattle
Limosilactobacillus oris	3	Homo sapiens
Limosilactobacillus panis	1	Sourdough
Limosilactobacillus pontis	3	Homo sapiens, Sourdough
Limosilactobacillus portuensis	1	Homo sapiens
Limosilactobacillus reuteri	200	Homo sapiens, Rodents, Poultry, Pig, Cattle, Horse, Sheep, Goat, Badger, Sourdough, dairy products
Limosilactobacillus rudii	2	Rodents
Limosilactobacillus secaliphilus	1	Unknown
Limosilactobacillus urinaemulieris	1	Homo sapiens
Limosilactobacillus vaginalis	3	Homo sapiens

Table 2. Summary of 74 genomes belonging to 11 *Limosilactobacillus* species and 1 putative novel species isolated from human, including genome length, number of predicted genes and accession numbers.

	Species	**Strain**	**Nº of nucleotides**	**Nº of genes predicted**	**Assembly**	**BioSample**
1	*Limosilactobacillus albertensis*	VA24_5	2177265	2137	ASM1102800v1	SAMN14087229
2	*Limosilactobacillus antri*	DSM_16041^T	2302896	2150	ASM16083v1	SAMN00001477
3		MGYG_HGUT_01516	2241837	2151	MGYG-HGUT-01516	SAMEA5851019
4	*Limosilactobacillus coleohominis*	101_4_CHN	1725829	1758	ASM16193v1	SAMN02463725
5		DSM_14060^T	1716238	2738	ASM143505v1	SAMN02369394
6	*Limosilactobacillus fermentum*	ATCC_14931^T	1867005	1839	ASM15921v1	SAMN00001473
7		28_3_CHN	2026518	1988	ASM16239v1	SAMN02463745
8		FTDC8312	1966551	2007	ASM41700v1	SAMN02469912
9		3872	2330492	2280	ASM46678v3	SAMN02314197
10		Lf1	1815647	1865	LF1_1.0	SAMN02053534
11		NB_22	2011311	2040	LfermNB22_1.0	SAMN02470787
12		LfQi6	2098510	1836	ASM96683v2	SAMN03372370
13		779_LFER	1935807	1970	ASM107702v1	SAMN03197989
14		HFB3	2043356	2323	ASM129702v1	SAMN04038470
15		UCO_979C	2011828	2271	ASM129790v1	SAMN04100088
16		47_7	2098685	1866	ASM185410v1	SAMN05893390
17		SNUV175	2272328	2251	ASM194178v1	SAMN06174220
18		FTDC_8312	2239921	2163	ASM211964v1	SAMN06703219

19		LAC_FRN_92	2063606	1990	ASM219243v1	SAMN07192734
20		SK152	2092273	2033	ASM224261v1	SAMN04858206
21		SHI_2	1937976	1938	ASM259193v1	SAMN07267218
22		279	1983266	2002	ASM279427v1	SAMN08014151
23		311	2043115	2077	ASM279431v1	SAMN08014154
24		103	2048240	2096	ASM279437v1	SAMN07985466
25		UMB0187	1969291	1990	ASM286326v1	SAMN08193695
26		LfU21	1966592	1991	ASM286982v1	SAMN08290293
27		CBA7106	2042277	1958	ASM325587v1	SAMN07224244
28		CIM_MAG_1415	1758796	1853	ASM334358v1	SAMN08294985
29		AF16_22LB	1996988	1969	ASM346275v1	SAMN09734308
30		AF15_40LB	1968193	1954	ASM346428v1	SAMN09734289
31		AF11_4_H	1935804	2286	ASM346508v1	SAMN09734212
32		MTCC_5898	2098685	1834	ASM420881v1	SAMN10868412
33		BIO6529	2127042	2225	ASM886847v1	SAMN12856542
34		AGR1487	1939032	1901	ASM1103274v1	SAMN13639333
35		AGR1485	2226862	2126	ASM1103276v1	SAMN13241836
36		CVM_347	2091502	2326	ASM1129075v1	SAMN14329897
37		S30	2159265	2223	ASM1142035v1	SAMN10869160
38		L13	1947210	1984	ASM1207060v1	SAMN10906232
39		L18	2113188	2348	ASM1207062v1	SAMN10906314

40		HFD1	2101878	2009	ASM1227303v1	SAMN14470504
41		MGYG_HGUT_00166	1968193	1954	MGYG-HGUT-00166	SAMEA5849667
42	*Limosilactobacillus gastricus*	PS3	1904872	2002	ASM24777v2	SAMN02471359
43		DSM_16045^T	1848461	1896	ASM143436v1	SAMN02369419
44		LG045	1905155	2009	ASM964855v1	SAMN13111359
45	*Limosilactobacillus mucosae*	CRL573	2257701	2667	ASM76690v1	SAMN03081593
46		A1	2175821	2110	ASM1342386v1	SAMN15515560
47		INIA_P508	2172535	2114	INIA_P508	SAMEA5673459
48		MGYG_HGUT_02319	2369669	2207	UHGG_MGYG-HGUT-02319	SAMEA5851823
49	*Limosilactobacillus oris*	PB013_T2_3	2115990	2049	ASM18001v1	SAMN00116776
50		F0423	2174937	2081	ASM22150v2	SAMN00195308
51		DSM_4864^T	2031774	1956	ASM143446v1	SAMN02369408
52	*Limosilactobacillus portuensis*	c11Ua_112M^T	2115452	2208	ASM1483873v1	SAMN15159667
53	*Limosilactobacillus reuteri*	SD2112	2316838	2330	ASM15945v2	SAMN00001494
54		MM4_1A	2067914	2089	ASM15947v2	SAMN00001504
55		CF48_3A	2107903	2212	ASM15961v1	SAMN00001503
56		MM2_3	2015721	2067	ASM16071v1	SAMN00002242
57		IRT	1993967	1995	ASM104683v1	SAMN03382532
58		DSM_20016^T	1999618	2027	ASM1682v1	SAMN02598351
59		MD_IIE_43	2059118	2068	ASM200708v1	SAMN06315550
60		M27U15	2035662	2144	ASM211219v1	SAMN06640219

61		MM34_4A	2152944	2260	ASM211280v1	SAMN06640223
62		ATG_F4	2041516	2066	ASM420861v1	SAMN10888479
63		C93	1887942	1930	ASM434965v1	SAMN10689995
64		C88	2066995	2177	ASM434968v1	SAMN10689994
65		reuteri	2047619	2064	ASM918472v1	SAMN12990932
66		LMG_P_27481	2016419	2068	ASM1106543v1	SAMN13025177
67		Marseille_P5461	2039572	1801	PRJEB32461-P5461	SAMEA5606434
68		Marseille_P5460	2039540	1800	PRJEB32461-P5460	SAMEA5606433
69		Marseille_P4904	2039591	1698	PRJEB32461-P4904	SAMEA5606432
70	*Limosilactobacillus sp. nov.*	UMB0683	1873656	1749	ASM294094v1	SAMN07580856
71	*Limosilactobacillus urinaemulieris*	c9Ua_26M^T	1898259	1909	ASM1483874v1	SAMN15159665
72	*Limosilactobacillus vaginalis*	ATCC_49540^T	1877332	1896	ASM15943v1	SAMN00001485
73		CBA_L88_BV2	1879874	1907	ASM383315v1	SAMN10395645
74		LV515	1894710	1897	ASM936293v1	SAMN12784813

Table 3. Core and accessory genome characterization of 5 *Limosilactobacillus* strains (3 species) of urogenital origin.

			Core genome	**Accessory genome**		
Species	**Strain**	**Origin**	**No. of core genes**	**No. of softcore genes**	**No. of dispensable genes**	**No. of unique genes**
Limosilactobacillus urinaemulieris	c9Ua26^T	urine	1449	62	234	164
Limosilactobacillus portuensis	c11Ua_112^T	urine	1464	75	275	394
Limosilactobacillus vaginalis	ATCC_49540^T	vagina	1428	95	284	89
	CBA_L88_BV2	vagina	1441	77	199	190
	LV515	vagina	1440	96	355	6

Table 4. Mobilome of urogenital *L. urinaemulieris* and *L. portuensis*.

Strain	**Prophage**					**Plasmid *rep* gene**	**IS element/ IS family**
	Length (Kb)	**Completeness**	**GC%**	**Total proteins**	**Most common phage (GenBank Accession No.)**		
c9Ua_26M	55.6	intact	38.90	63	*Lactobacillus* tphage phi g1e (NC_004305)	-	IS*Lp1*/IS*30* IS*Lhe30*/IS*30*
	42.2	intact	38.75	75	*Lactobacillus* phage BK5-T (NC_002796)		
c11Ua_112M	26.0	questionable	39.21	38	*Lactobacillus* phage bIL286 (NC_002667)	*repUS19* (pDLK3)	IS*Lp1*/IS*30*
	11.4	questionable	40.29	15	*Levilactobacillus* phage Lb (NC_047983)		
	17.0	incomplete	34.46	17	*Lactiplantibacillus* phage Sha1 (NC_019489)		
	9.3	incomplete	34.90	15	*Lactocbacillus* phage phiadh (NC_000896)		
	14.3	incomplete	38.27	15	*Levilactobacillus* phage Lb (NC_047983)		

Figure 1. Heatmap representing percentage identity of ANIb for 332 *Limosilactobacillus* genomes. The largest clusters comprise strains identified as *L. reuteri* (n= 200 genomes), *L. fermentum* (n= 79 genomes) and *L. mucosae* (n= 13). The designation of the strains represents original identification under which strains were deposited in the public NCBI database.

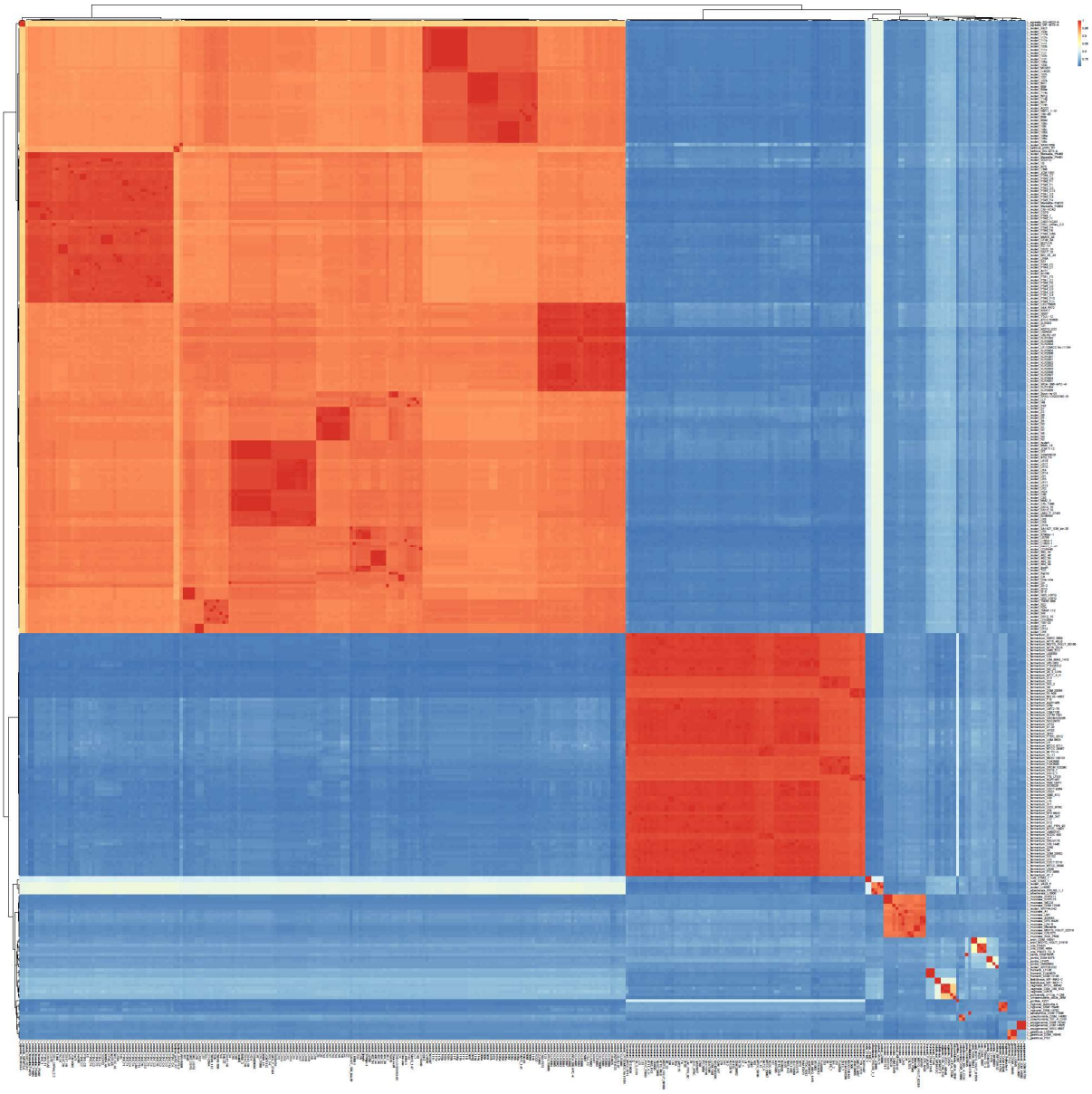


Figure 2. Phylogenomic tree based on single-copy core 37 genes encoding ribosomal proteins created with anvi'o v6 pipeline and edited in iTOL. Maximum likelihood tree with Jones-Taylor-Thorton substitution model was built with FastTree v 2.1.11 with SH-like 1000 support. Tree includes all validly described species type strains and not identified or not properly identified strains of *Limosilactobacillus* genus. Strains designated with one asterisk (*) at the end indicate strains originally deposited as *L. reuteri*, and with two asterisks (**) originally deposited as *L. pontis*.

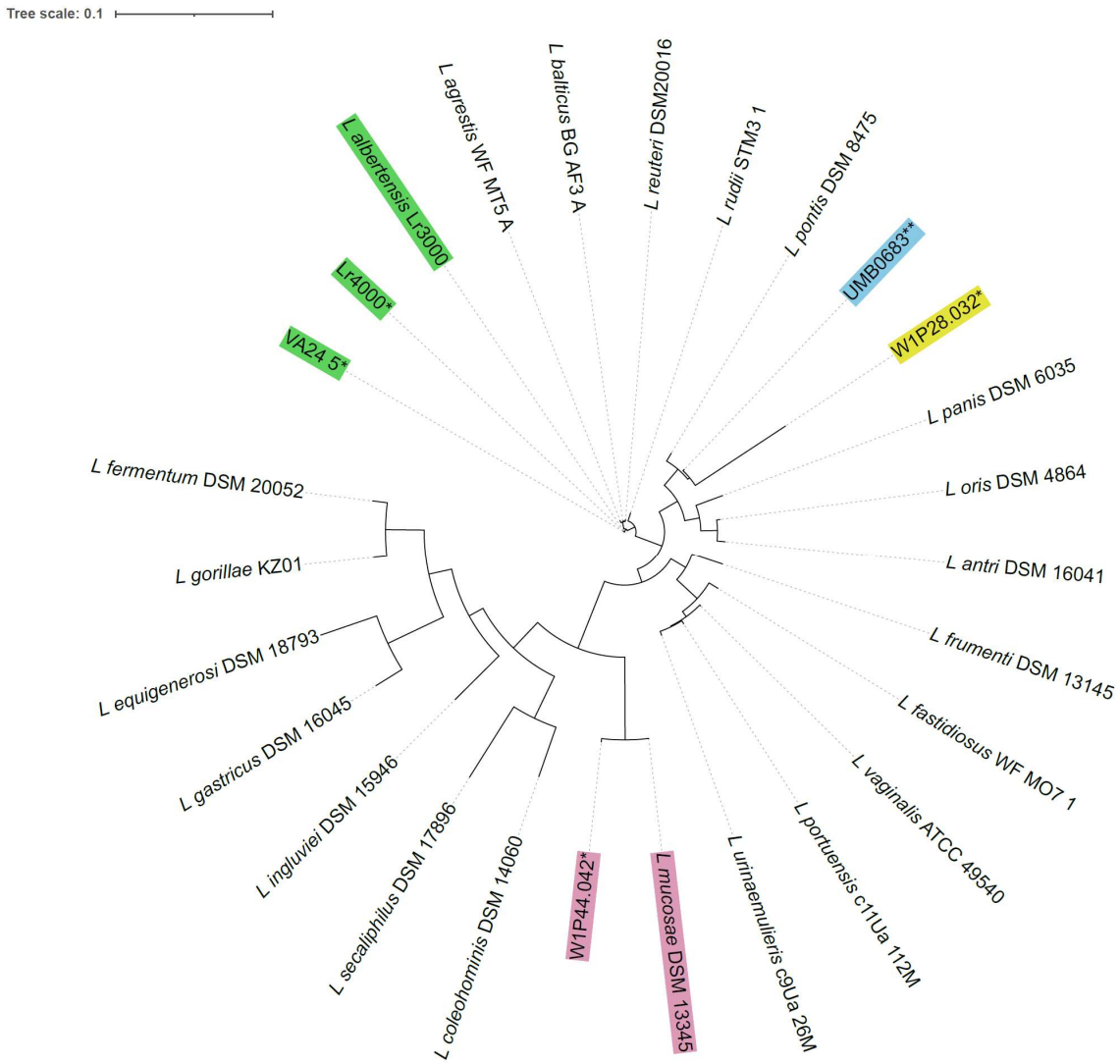


Figure 3. Pangenome of all *Limosilactobacillus* strains isolated from human host (11 species, 1 putative novel species, 74 genomes) generated by anvi'o. Genomes are organized based on tree of frequencies of gene clusters (right top). Each color represents different species. External white-green ring represents COGs functionality annotation, with green standing for known and white for unknown functions. Outside ring represents particular genes collections: red - core, black - accessory genome. Within accessory genome, softcore is marked in yellow, and singletons in blue. Additional information such as total length, GC content, completion, redundancy, number of genes, average gene length, number of genes per kbp, singleton gene clusters and number of gene clusters are represented by bars in top right.

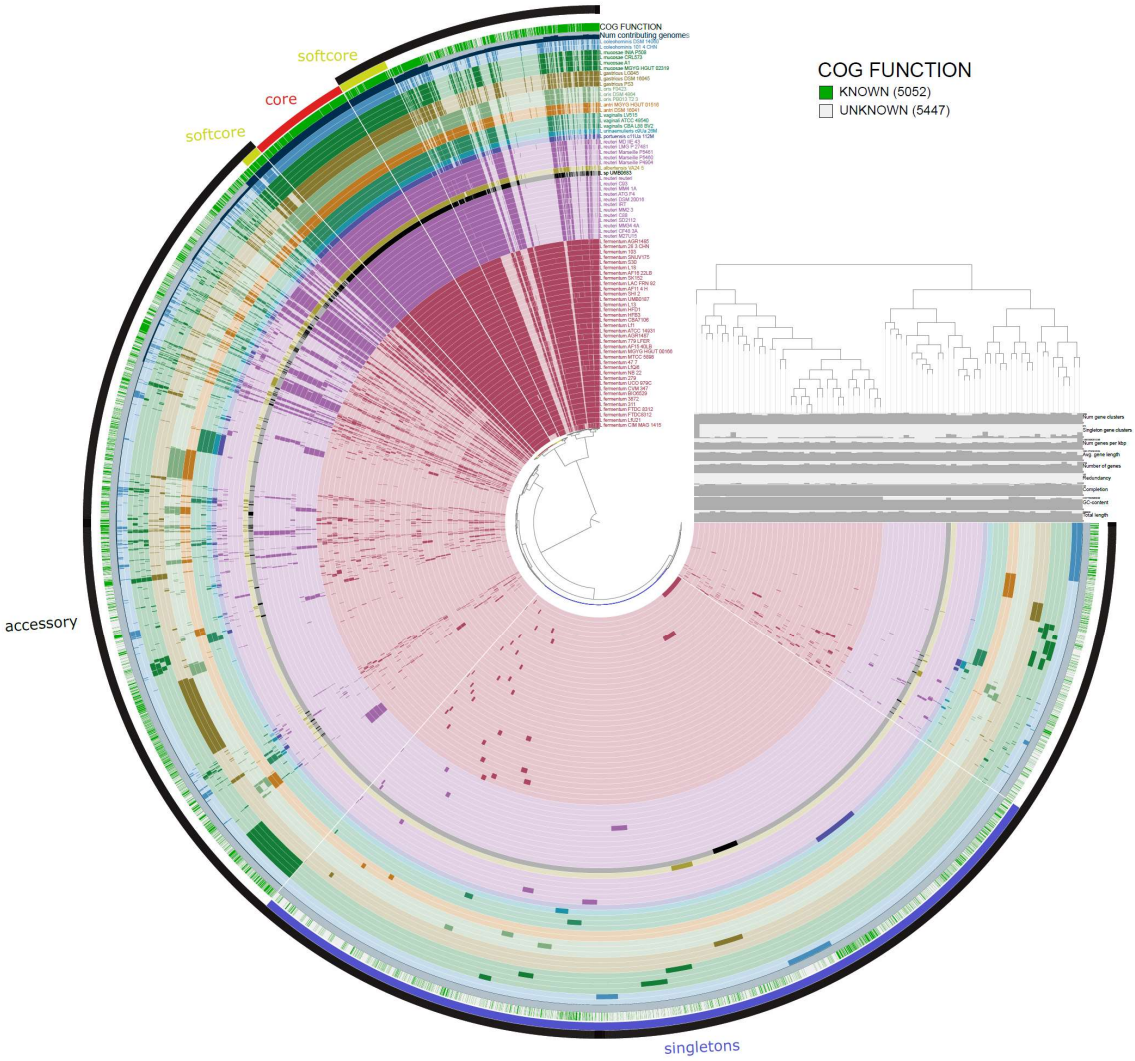


Figure 4. Pangenome specific genes collections grouped by COG functional categories.

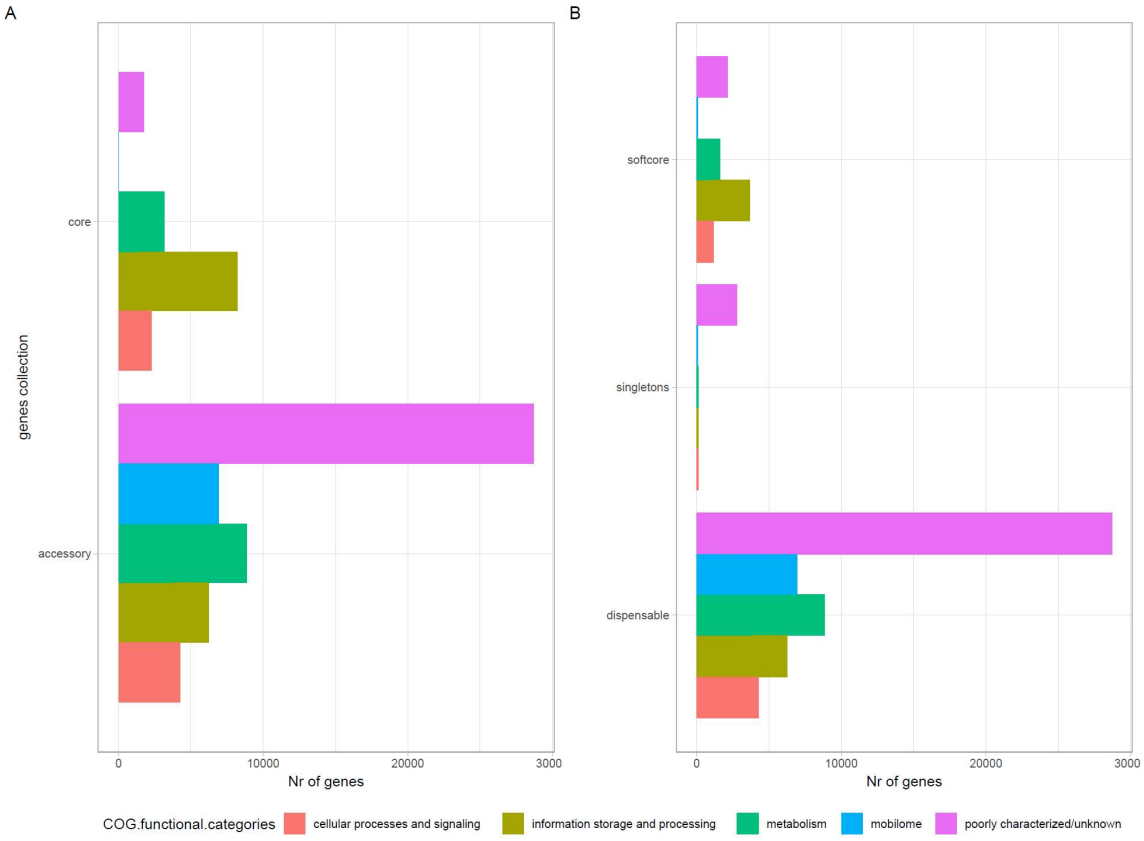


Figure 5. Distribution of COG categories for all detected genes among 5 *Limosilactobacillus* strains.

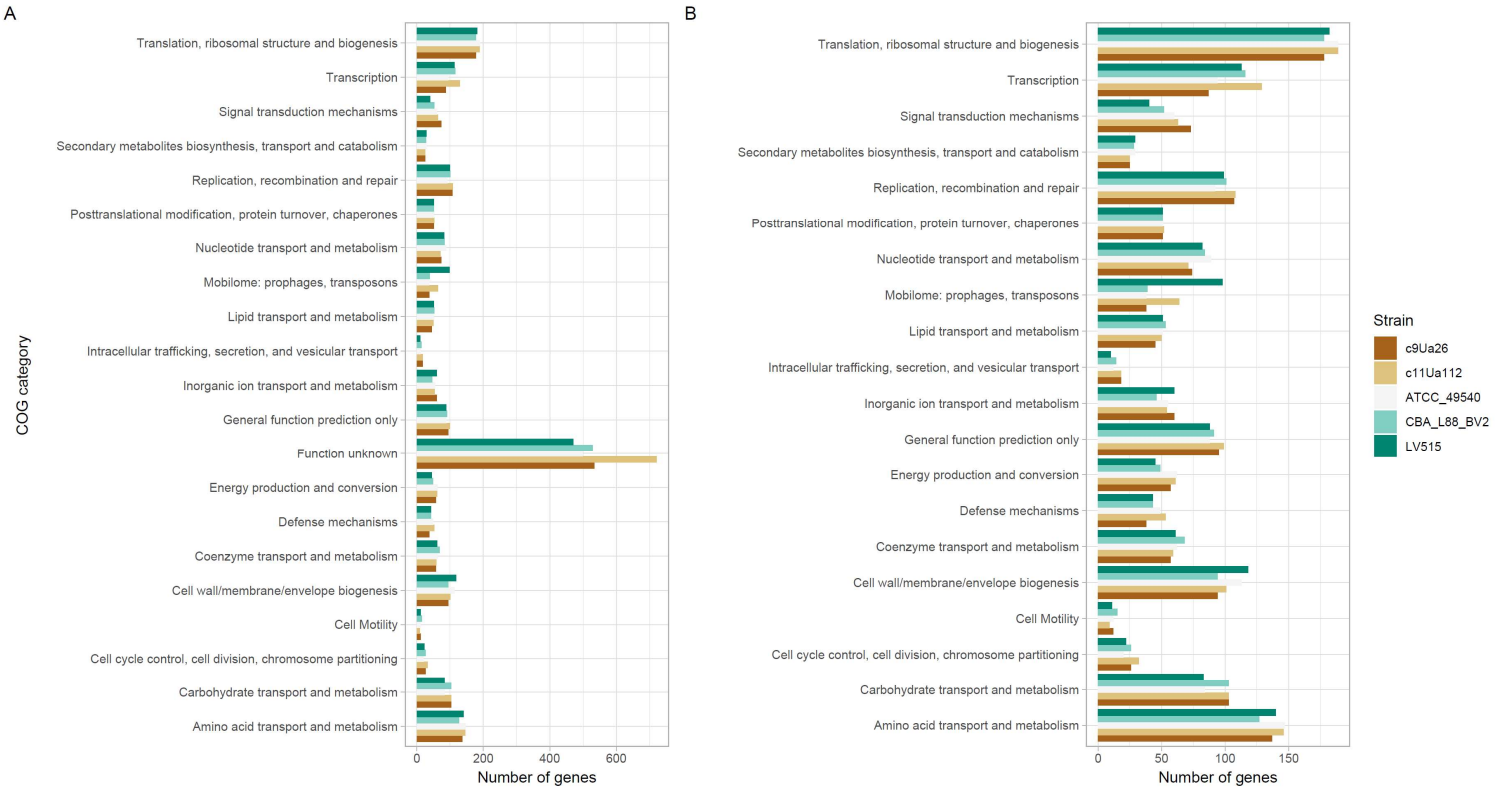
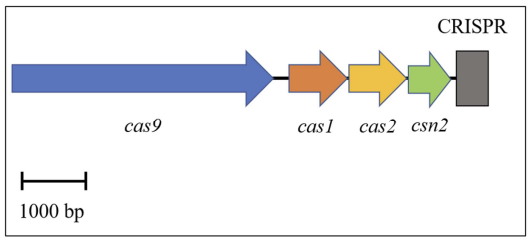


Figure 6. Schematic representation of CRISPR-Cas system in *L. urinaemulieris* c9Ua_26MT. The length of the arrows represents the length of the genes.



Additional material associated with the manuscript entitled "*Limosilactobacillus* spp. diversity and relevant features of urogenital species" is available through the following links:

Supplementary Material:

Table S1

https://docs.google.com/spreadsheets/d/138XOq7qjJxIs-NwfaCh7YHdKZPJBoQcQ/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Table S2

https://drive.google.com/file/d/179QE-NV88tUQTSUgb44Sl1wI8eDO1lln/view?usp=sharing

Table S3

https://docs.google.com/spreadsheets/d/17twPnAK0ocdwyUlGEE1KzSqoY96Ev2ne/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Table S4

https://docs.google.com/spreadsheets/d/19acGQ3qb7r1rU7O8k6wDASKJRvgjcHHm/edit?usp=sharing&ouid=115798665925427872829&rtpof=true&sd=true

Figures in high resolution:

https://drive.google.com/file/d/1h6MssB30epIht2VZDh3FiR4mB9-fEF07/view?usp=sharing

https://drive.google.com/file/d/1fMXMOmfZ4cFkSI-CyB6WBCmXuHYN5Lpr/view?usp=sharing

Discussion and conclusions

With this work, we enlarged the diversity of bacterial members and community structures of female urinary tract microbiome (FUM) and presented novel data on dynamics of FUM over a long-time period. New findings on the relationship between vaginal and urinary tract (UT) microbiome and a snapshot on the FUM structure in UT diseased hosts are also presented. Moreover, fine-grained bacterial analysis allowed the identification of 3 new species and 7 putative new ones as well as relevant virulence/properties of different bacterial groups (e.g., *Escherichia coli, Prevotella* sp. and *Limosilactobacillus* sp.). These achievements derived from a detailed data analyses and improved methodological approaches. Following, details on the main achievements and their relevance for improving FUM understanding are presented.

1. Comprehensive FUM structure composition.

Our data demonstrated that healthy FUM is characterized by a wide repertoire of bacterial species (8 phyla, 116 genera, 297 species; 53 species/sample). Overall, the FUM alpha diversity is intermediate comparing to other human body sites and species present relatively uneven distribution[5]. Moreover, healthy FUM presents a high degree of inter-individual species variability. Our data confirmed that previously defined urotypes such as *Lactobacillus*, *Streptococcus* or *Gardnerella* are indeed highly prevalent among healthy individuals, however for the first time we demonstrated high diversity within these urotypes. For instance, *Lactobacillus* urotype may comprise diverse bacterial communities, with six different species variable in abundance and/or presence (*L. crispatus, L. iners, L. gasserii, L. paragasseri, L. jensenii, L. mulieris*), which likely diverge in functional characteristics. Moreover, we uncovered diversity of *Gardnerella* genus, demonstrating that healthy FUM may comprise four different *Gardnerella* species (*G. vaginalis, G. swidsinskii, G. leopoldii, G.* genomospecies 3). Furthermore, certain bacterial species are highly prevalent (present in more than 50% of samples analyzed, e.g., *Corynebacterium tuberculostearicum, Finegoldia magna, Prevotella bivia*), however there is not a single species that would be common to all healthy FUM. Opportunistic pathogens associated with human UTI and/or genital tract diseases are often detected among healthy FUM (e.g., *Atopobium vaginae, Escherichia coli, Ureaplasma urealyticum*). These novel findings are of high importance since proper understanding of healthy FUM diversity is essential for further recognition of FUM dysbiosis and treatment development.

Our set of urine samples was subjected to analysis by two complementary methodologies i.e., improved protocol for 16S rRNA gene amplicon sequencing and optimized culturomic

approach. Considering that urine are low-biomass samples, we enlarged initial urine volume (20 ml) used for total DNA extraction, while previous studies used much smaller volume (usually 1 ml)[26,74,80]. Our study is also the first to apply long reads sequencing technology to characterize healthy FUM, which targets near full length of 16S rRNA gene. These improvements are highly important to increase accuracy of bacterial DNA detection from urine samples and especially species-level identification, since other studies usually apply sequencing only of selected 16S rRNA variable regions.

Furthermore, current culturomic-based urinary tract microbiome studies typically characterize a small number of bacterial isolates that represent most abundant morphotypes. We improved previously used culturomic protocol mainly by supplementing microbiological culture media to enhance recovery of fastidious bacteria (e.g., *Lactobacillus* sp., *Gardnerella* sp.) and by outstandingly broaden characterization of isolates recovered from each sample. Specifically, following our protocol, on average we characterized over 100 isolates/sample, while previous studies isolated and identified up to 5 isolates/sample[75,133].

For the advancements in uncovering diversity of healthy FUM going beyond commonly used MALDI-TOF MS, through housekeeping genes sequencing (e.g., 16S rRNA, *pheS, cpn60, rpoB*), was essential. This novel and exhaustive approach for microbiome members characterization allowed accurate identification of bacterial species in FUM (e.g., Lactobacillaceae, *Gardnerella* sp., *Corynebacterium* sp.) and detection of ten putative novel species (taxonomic description of *Lactobacillus mulieris*, *Limosilactobacillus urinaemulieris* and *Limosilactobacillus portuensis*). Genomic and biochemical characterization of these novel species will contribute to better understanding of their metabolism and fitness in healthy UT. Importantly, MALDI-TOF MS insufficiencies were also identified. Of note are those concerning the most abundant and/or prevalent bacterial species (e.g., Lactobacillaceae, *Gardnerella* sp.) that were often misidentified, mostly due to limitations of spectral reference databases. These culturomic improvements together with a more accurate identification allowed to enlarge the repertoire of bacterial species inhabiting healthy UT and a better understanding of healthy FUM structure.

The culturomic findings were complemented by data obtained from amplicon sequencing e.g., identification of slow-growing and/or fastidious species (e.g., *Campylobacter ureolyticus, Ureaplasma parvum*, *Sneathia sanguinegens*). Moreover, using two methodologies we

confirmed the presence of several low abundant bacteria detected by amplicon sequencing in healthy FUM (e.g., *Actinomyces urogenitalis*). This demonstrates, that currently used approach for amplicon sequencing data filtering (removal of taxa detected in < 0.1%) is inaccurate and will understate FUM diversity. Overall, applying both methodologies was shown to be beneficial for FUM compositional characterization (only 22% of species were detected by both methodologies). However, such extensive analysis may be associated with high costs and importantly, data integration may be challenging. Both approaches demonstrate specific advantages and limitations (e.g., quantification of bacteria in FUM). For instance, amplicon sequencing allows much faster FUM identification thus it could be more advantageous for diagnostic purposes, while culturomic does provide isolates for further manipulation and experimental testing and it seems more valuable in the research.

2.Compositional long-term stability of FUM in asymptomatic population.

Our data highlights that healthy microbiome composition may significantly change overtime (e.g., switch between different species such as *Gardnerella vaginalis* to *Gardnerella swidsinskii* or *Lactobacillus jensenii* to *Lactobacillus crispatus*), thus there might be different microbiome structures during the individual lifespan with similar functions. Importantly, these findings highlight the critical need for functional microbiome characterization, as different microbes may share the same functional traits. We demonstrated that certain bacterial communities appear to be more resilient in this niche (e.g., communities composed by *Lactobacillus crispatus*, *Bifidobacterium* sp., *Lactobacillus paragasseri*). Possibly, these communities have a greater fitness and are less susceptible to microbiome imbalances that could lead to disease. We also reported association of compositional FUM changes of *Finegoldia magna* and *Streptococcus anginosus* with host smoking status, which deserves further exploration, especially that these species have been previously reported from patients with urologic conditions.

3. Interconnection between urinary tract and vaginal microbiome.

Our findings demonstrated that urinary tract and vaginal microbiome are similar regarding overall structure (beta diversity, alpha diversity i.e., Pielou evenness and Shannon index).

However, several differences between microbiome in both niches can be observed at species level. For instance, differential abundance of certain species contributed to community disparities (e.g., urinary tract was enriched in *Gardnerella* genomospecies 3 while vaginal microbiome was enriched in *Limosilactobacillus urinaemulieris*). Moreover, certain differences were observed in co-occurrence correlation patterns, e.g., *Lactobacillus jensenii* and *Lactobacillus iners* are positively correlated in vaginal microbiome, while negatively correlated in urinary tract microbiome. Since many taxa are present in both urinary tract and vagina, there is a possibility that specific microbes may interchange their exact location and act complementary in both niches.

Importantly, most paired samples presented different microbiome structures, supporting usage of mid-stream voided urine as representative sample for lower urinary tract microbiome characterization.

4. FUM in urinary tract diseased hosts.

Our initial characterization suggests the microbiome of hosts with urologic conditions are enriched in putative uropathogenic species. We identified some UTI-associated species only in diseased cohorts e.g., *Proteus mirabilis, Pseudomonas luteola, Citrobacter portucalensis* in women diagnosed with overactive bladder syndrome, or *Klebsiella pneumoniae, Varibaculum* sp. in women with history of rUTI, however they were mostly in low relative abundance. Moreover, *Lactobacillus iners* was often present in high relative abundance, in both cohorts. Microbiome of OAB women was enriched in *Enterococcus faecalis, Campylobacter ureolyticus*, *Ureaplasma parvum* and occasionally with *Gardnerella* genomospecies 3 and *Gardnerella leopoldii*. Interestingly, OAB woman with the highest symptom severity score had bacterial community composed of potentially relevant e.g., *Gardnerella vaginalis* and *Ureaplasma urealyticum*. Nevertheless, this data should be further explored on larger well-characterized cohort since our sample size was relatively small and our asymptomatic controls were of different age.

5. Insights on relevant features of selected FUM members

Our analysis allowed to isolate multiple potentially relevant species to urinary tract health for which a little information is available. While characterizing novel species and/or poorly

explored FUM members we encountered members of the genus *Limosilactobacillus* and *Prevotella* that could have a role in urinary tract homeostasis. We also explored genomic diversity of *Escherichia coli* inhabiting asymptomatic population, since this species is the most common cause of UTI.

Since *Prevotella* genus is often involved in human infections/diseases, we explored genomic features of two closely related species i.e., *Prevotella corporis* and *Prevotella brunnea*, both isolated from our urine samples. We found that they are likely opportunistic pathogens, and they share conserved features that might be involved in their virulence (e.g., patatin-like phospholipases). Both species have also specific virulence-associated traits, especially strains isolated from diseased host (e.g., sialic acid synthesis, specific proteases, genes involved in arsenate resistance). Interestingly, both species can ferment mannose, which could have an impact on UT homeostasis. Particularly, they could contribute to reduction of quantities of mannose in bladder environment, which is proposed as supplementation to inhibit *E. coli* adhesion properties and thus, act preventively for UTI. We also provided data on resistance of these isolates, showing that they could contribute to spread of resistant genes (e.g., *tetQ*). Additionally, analyzing recently reclassified *Limosilactobacillus* genus, we highlighted several features that suggest that these bacteria are involved in urinary tract health. We demonstrated high variability within the genus, but high rate of conserved functions (e.g., lactic acid production) among strains of urogenital origin. Both works present novel descriptive genomic analysis of poorly explored species, which will provide basis for further studies investigating their role in urinary tract homeostasis.

Our compositional data revealed high prevalence of putative uropathogens in healthy population, which led to comprehensive characterization of *Escherichia coli* strains isolated from urogenital tract of asymptomatic women. Our data confirms that most strains inhabiting healthy women belong to extraintestinal pathogenic *E. coli* and have highly similar genomic background and virulence genes profile as those known to cause UTI. The factors that contribute to transition from commensal to pathogen remains unclear and should be further investigated. We also provided novel data on antimicrobial resistance of *E. coli* clones colonizing healthy women constituting a potential source of antimicrobial resistant genes to co-habiting microbiome members. Some strains were resistant to antibiotics commonly used in UTI treatment which should be of clinical concern, since it may increase risk of treatment failure if the host eventually develop UTI. We also contributed to phylogenomic

characterization of *E. coli* strain isolated from non-human severe UTI, demonstrating the importance and distribution of hemolysin A operon duplication in severity of UTI.

Overall, this work contributed to in-depth understanding of healthy FUM demonstrating that microbiome of urinary tract is composed of complex bacterial communities, and it is highly variable between individuals and within individuals over time. Lack of single bacterial species common to all asymptomatic individuals and especially, presence of putative uropathogenic species and without clinical signs of infection suggest that microbial interactions within microbiome play crucial role in urinary tract homeostasis. Additional future hypothesis would be that interlink between urinary tract and vaginal microbiome provide interchangeable transient species that may act prophylactically or deleteriously in each niche.

The future research effort should focus on identifying species/species combinations in FUM that could prevent disease/infection development. Further exploration of functional properties of other FUM members and experimental validation of genomic findings relevant for UT health is needed. Importantly, the research focus should be directed towards understanding microbial functionality in the community, since single bacterial isolate might behave differently. Moreover, advancing the field with transcriptomic and proteomic analysis may bring important data regarding strains colonizing urinary tract to distinguish putative pathogens from commensals.

Ultimately, the data presented in this thesis will largely contribute to future dysbiosis definition and lead to further development of therapeutic approaches involving microbiome interventions.

References

1. Leeuwenhoek, A. V. Observations, communicated to the publisher by Mr. Antony van Leewenhoeck, in a dutch letter of the 9th Octob. 1676. here English'd: concerning little animals by him observed in rain-well-sea- and snow water; as also in water wherein pepper had lain infused. *Philos Trans R Soc Lond* **12**, 821–831 (1677).
2. Mata, L. J., Carrillo, C. & Villatoro, E. Fecal microflora in health persons in a preindustrial region. *Appl Microbiol* **17**, 596–602 (1969).
3. Gorbach, S. L., Menda, K. B., Thadepalli, H. & Keith, L. Anaerobic microflora of the cervix in healthy women. *Am J Obstet Gynecol* **117**, 1053–1055 (1973).
4. Slots, J. Microflora in the healthy gingival sulcus in man. *Scand J Dent Res* **85**, 247–254 (1977).
5. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
6. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 103 (2020).
7. Yong, E. I contain multitudes: the microbes within us and a grander view of life. *Harper Collins Publishers* (2016).
8. Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat Med* **24**, 392–400 (2018).
9. Hooks, K. B. & O'Malley, M. A. Dysbiosis and Its Discontents. *mBio* **8**, e01492-17 (2017).
10. Lloyd-Price, J. *et al.* Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017).
11. The Integrative Human Microbiome Project: Dynamic Analysis of Microbiome-Host Omics Profiles during Periods of Human Health and Disease. *Cell Host & Microbe* **16**, 276–289 (2014).
12. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
13. Lane, E. R., Zisman, T. L. & Suskind, D. L. The microbiota in inflammatory bowel disease: current and therapeutic insights. *J Inflamm Res* **10**, 63–73 (2017).
14. Schirmer, M. *et al.* Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* **3**, 337–346 (2018).
15. Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. *Genome Medicine* **8**, 51 (2016).

16. Scotti, E. *et al.* Exploring the microbiome in health and disease: Implications for toxicology. *Toxicology Research and Application* **1**, 2397847317741884 (2017).
17. Dinan, T. G. & Cryan, J. F. The Microbiome-Gut-Brain Axis in Health and Disease. *Gastroenterol Clin North Am* **46**, 77–89 (2017).
18. Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. J. Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* **26**, 26191 (2015).
19. Strandwitz, P. *et al.* GABA-modulating bacteria of the human gut microbiota. *Nat Microbiol* **4**, 396–403 (2019).
20. Weingarden, A. R. & Vaughn, B. P. Intestinal microbiota, fecal microbiota transplantation, and inflammatory bowel disease. *Gut Microbes* **8**, 238–252 (2017).
21. Korpela, K. *et al.* Maternal Fecal Microbiota Transplantation in Cesarean-Born Infants Rapidly Restores Normal Gut Microbial Development: A Proof-of-Concept Study. *Cell* **183**, 324-334.e5 (2020).
22. Ni, J., Wu, G. D., Albenberg, L. & Tomov, V. T. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol* **14**, 573–584 (2017).
23. Amabebe, E. & Anumba, D. O. C. The Vaginal Microenvironment: The Physiologic Role of Lactobacilli. *Front Med* **5**, 181 (2018).
24. Wollina, U. Microbiome in atopic dermatitis. *Clin Cosmet Investig Dermatol* **10**, 51–56 (2017).
25. Nelson, D. E. *et al.* Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. *PLoS ONE* **5**, e14116 (2010).
26. Siddiqui, H., Nederbragt, A. J., Lagesen, K., Jeansson, S. L. & Jakobsen, K. S. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. *BMC Microbiol.* **11**, 244 (2011).
27. Wolfe, A. J. *et al.* Evidence of uncultivated bacteria in the adult female bladder. *J Clin Microbiol* **50**, 1376–1383 (2012).
28. Shannon, M. B. *et al.* Bladder urinary oxygen tension is correlated with urinary microbiota composition. *Int Urogynecol J* **30**, 1261–1267 (2019).
29. Sarigul, N., Korkmaz, F. & Kurultak, İ. A New Artificial Urine Protocol to Better Imitate Human Urine. *Sci Rep* **9**, 20159 (2019).
30. Mann, R., Mediati, D. G., Duggin, I. G., Harry, E. J. & Bottomley, A. L. Metabolic Adaptations of Uropathogenic *E. coli* in the Urinary Tract. *Front Cell Infect Microbiol* **7**, 241 (2017).

31. Alteri, C. J. & Mobley, H. L. T. Metabolism and Fitness of Urinary Tract Pathogens. *Microbiol Spectr* **3**, (2015) doi:10.1128/microbiolspec.MBP-0016–2015.
32. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat Rev Microbiol* **13**, 269–284 (2015).
33. Lacerda Mariano, L. & Ingersoll, M. A. The immune response to infection in the bladder. *Nat Rev Urol* **17**, 439–458 (2020).
34. Alfano, M. *et al.* The interplay of extracellular matrix and microbiome in urothelial bladder cancer. *Nat Rev Urol* **13**, 77–90 (2016).
35. Klingler, C. H. Glycosaminoglycans: How Much do We Know about their Role in the Bladder? *Urologia* **83**, 11–14 (2016).
36. Pastorello, I. *et al.* EsiB, a novel pathogenic *Escherichia coli* secretory immunoglobulin A-binding protein impairing neutrophil activation. *mBio* **4**, e00206-13 (2013).
37. Hiergeist, A. & Gessner, A. Clinical implications of the microbiome in urinary tract diseases. *Curr Opin Urol* **27**, 93–98 (2017).
38. Wilson, M. L. & Gaido, L. Laboratory diagnosis of urinary tract infections in adult patients. *Clin Infect Dis* **38**, 1150–1158 (2004).
39. Thomas-White, K. *et al.* Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat Commun* **9**, 1557 (2018).
40. Khasriya, R. *et al.* Spectrum of bacterial colonization associated with urothelial cells from patients with chronic lower urinary tract symptoms. *J. Clin. Microbiol.* **51**, 2054–2062 (2013).
41. Chen, Y. B. *et al.* The Urethral Microbiota - A Missing Link in the Female Urinary Microbiota. *J Urol* (2020) doi:10.1097/JU.0000000000000910.
42. Pohl, H. G. *et al.* The Urine Microbiome of Healthy Men and Women Differs by Urine Collection Method. *Int Neurourol J* **24**, 41–51 (2020).
43. Southworth, E. *et al.* A Cross-sectional Pilot Cohort Study Comparing Standard Urine Collection to the Peezy Midstream Device for Research Studies Involving Women. *Female Pelvic Med Reconstr Surg* **25**, e28 (2019).
44. Collins, L., Sathiananthamoorthy, S., Rohn, J. & Malone-Lee, J. A revalidation and critique of assumptions about urinary sample collection methods, specimen quality and contamination. *Int Urogynecol J* **31**, 1255–1262 (2020).

45. Collier, S., Matjiu, F., Jones, G., Harber, M. & Hopkins, S. A prospective study comparing contamination rates between a novel mid-stream urine collection device (Peezy) and a standard method in renal patients. *J Clin Pathol* **67**, 139–142 (2014).
46. Lagier, J.-C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* **1**, 1–8 (2016).
47. Lau, J. T. *et al.* Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med* **8**, 72 (2016).
48. Karstens, L. *et al.* Community profiling of the urinary microbiota: considerations for low-biomass samples. *Nat Rev Urol* **15**, 735–749 (2018).
49. Delanghe, J. & Speeckaert, M. Preanalytical requirements of urinalysis. *Biochem Med (Zagreb)* **24**, 89–104 (2014).
50. Roux, A., Thévenot, E. A., Seguin, F., Olivier, M.-F. & Junot, C. Impact of collection conditions on the metabolite content of human urine samples as analyzed by liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy. *Metabolomics* **11**, 1095–1105 (2015).
51. Bundgaard-Nielsen, C. *et al.* Voided Urinary Microbiota Is Stable Over Time but Impacted by Post Void Storage. *Front Cell Infect Microbiol* **10**, 435 (2020).
52. Jung, C. E. *et al.* Benchmarking urine storage and collection conditions for evaluating the female urinary microbiome. *Sci Rep* **9**, 13409 (2019).
53. Bilen, M. *et al.* The contribution of culturomics to the repertoire of isolated human bacterial and archaeal species. *Microbiome* **6**, 94 (2018).
54. Browne, H. P. *et al.* Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543–546 (2016).
55. Knight, R. *et al.* Best practices for analysing microbiomes. *Nat Rev Microbiol* **16**, 410–422 (2018).
56. Cui, H., Li, Y. & Zhang, X. An overview of major metagenomic studies on human microbiomes in health and disease. *Quant Biol* **4**, 192–206 (2016).
57. Hiergeist, A., Gläsner, J., Reischl, U. & Gessner, A. Analyses of Intestinal Microbiota: Culture versus Sequencing. *ILAR J* **56**, 228–240 (2015).
58. Weinstock, G. M. Genomic approaches to studying the human microbiota. *Nature* **489**, 250–256 (2012).
59. Fouts, D. E. *et al.* Integrated next-generation sequencing of 16S rDNA and metaproteomics differentiate the healthy urine microbiome from asymptomatic

bacteriuria in neuropathic bladder associated with spinal cord injury. *J Transl Med* **10**, 174 (2012).

60. Lewis, D. A. *et al.* The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. *Front Cell Infect Microbiol* **3**, 41 (2013).
61. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* **35**, 833–844 (2017).
62. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351 (2016).
63. Loman, N. J. *et al.* High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* **10**, 599–606 (2012).
64. Meisel, J. S. & Grice, E. A. The Human Microbiome. *Genomic and Precision Medicine* 63–77 (2017) doi:10.1016/B978-0-12-800681-8.00004-9.
65. Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of Next Generation Sequencing Technologies. *Curr Protoc Mol Biol* **122**, e59 (2018).
66. Bag, S. *et al.* An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. *Sci Rep* **6**, 26775 (2016).
67. Costea, P. I. *et al.* Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol* **35**, 1069–1076 (2017).
68. Teng, F. *et al.* Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci Rep* **8**, 16321 (2018).
69. Lim, M. Y., Song, E.-J., Kim, S. H., Lee, J. & Nam, Y.-D. Comparison of DNA extraction methods for human gut microbial community profiling. *Syst Appl Microbiol* **41**, 151–157 (2018).
70. Mai-Prochnow, A., Clauson, M., Hong, J. & Murphy, A. B. Gram positive and Gram negative bacteria differ in their sensitivity to cold plasma. *Sci Rep* **6**, 38610 (2016).
71. Qiagen. DNeasy Blood and Tissue Handbook. Online Handbook, accessed 20 September 2021, <https://www.qiagen.com> (2020).
72. Alakomi, H.-L. Weakening of the Gram-negative bacterial outer membrane: a tool for increasing microbiological safety. *VTT Publications* 638:5-95 (2007).
73. Natarajan, V. P., Zhang, X., Morono, Y., Inagaki, F. & Wang, F. A Modified SDS-Based DNA Extraction Method for High Quality Environmental DNA from Seafloor Environments. *Front Microbiol* **7**, 986 (2016).
74. Pearce, M. M. *et al.* The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *mBio* **5**, e01283-01214 (2014).

75. Hilt, E. E. *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* **52**, 871–876 (2014).
76. Moustafa, A. *et al.* Microbial metagenome of urinary tract infection. *Sci Rep* **8**, 4333 (2018).
77. Karstens, L. *et al.* Benchmarking DNA isolation kits used in analyses of the urinary microbiome. *Sci Rep* **11**, 6186 (2021).
78. Munch, M. M. *et al.* Optimizing bacterial DNA extraction in urine. *PLOS ONE* **14**, e0222962 (2019).
79. Karstens, L. *et al.* Does the Urinary Microbiome Play a Role in Urgency Urinary Incontinence and Its Severity? *Front Cell Infect Microbiol* **6**, 78 (2016).
80. Thomas-White, K. J. *et al.* Incontinence medication response relates to the female urinary microbiota. *Int Urogynecol J* **27**, 723–733 (2016).
81. Wu, P. *et al.* Urinary Microbiome and Psychological Factors in Women with Overactive Bladder. *Front Cell Infect Microbiol* **7**, 488 (2017).
82. Ammitzbøll, N. *et al.* Pre- and postmenopausal women have different core urinary microbiota. *Sci Rep* **11**, 2212 (2021).
83. Jacobs, K. M. *et al.* Cultivable Bacteria in Urine of Women With Interstitial Cystitis: (Not) What We Expected. *Female Pelvic Med Reconstr Surg* **27**, 322–327 (2021).
84. Cumpanas, A. A. *et al.* Urinary Microbiota-Are We Ready for Prime Time? A Literature Review of Study Methods' Critical Steps in Avoiding Contamination and Minimizing Biased Results. *Diagnostics (Basel)* **10**, 343 (2020).
85. Hoffman, C. *et al.* Species-Level Resolution of Female Bladder Microbiota from 16S rRNA Amplicon Sequencing. *mSystems* **0**, e00518-21 (2021).
86. Nelson, D. E. *et al.* Bacterial Communities of the Coronal Sulcus and Distal Urethra of Adolescent Males. *PLOS ONE* **7**, e36298 (2012).
87. Ackerman, A. L. & Underhill, D. M. The mycobiome of the human urinary tract: potential roles for fungi in urology. *Ann Transl Med* **5**, 31 (2017).
88. Wu, Y., Jiang, H., Tan, M. & Lu, X. Screening for chronic prostatitis pathogens using high-throughput next-generation sequencing. *The Prostate* **80**, 577–587 (2020).
89. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* **11**, 2639–2643 (2017).

90. Jovel, J. *et al.* Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Front Microbiol* **7**, (2016).
91. Cole, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**, D141-145 (2009).
92. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–D596 (2013).
93. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069–5072 (2006).
94. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336 (2010).
95. Schloss, P. D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**, 7537–7541 (2009).
96. Callahan, B. J. *et al.* DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).
97. Morgan, X. C. & Huttenhower, C. Chapter 12: Human Microbiome Analysis. *PLOS Comput Biol* **8**, e1002808 (2012).
98. Xia, Y. & Sun, J. Hypothesis Testing and Statistical Analysis of Microbiome. *Genes Dis* **4**, 138–148 (2017).
99. Starke, R., Pylro, V. S. & Morais, D. K. 16S rRNA Gene Copy Number Normalization Does Not Provide More Reliable Conclusions in Metataxonomic Surveys. *Microb Ecol* **81**, 535–539 (2021).
100. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* **69**, 330–339 (2007).
101. Martinez-Porchas, M., Villalpando-Canchola, E., Ortiz Suarez, L. E. & Vargas-Albores, F. How conserved are the conserved 16S-rRNA regions? *PeerJ* **5**, e3036 (2017).
102. PacBio. Pacific Biosciences, accessed 20 September 2021, < https://www.pacb.com>.
103. Větrovský, T. & Baldrian, P. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE* **8**, e57923 (2013).
104. Yang, B., Wang, Y. & Qian, P.-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* **17**, 135 (2016).

105. Durazzi, F. *et al.* Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep* **11**, 3030 (2021).
106. Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499–504 (2019).
107. Campanaro, S., Treu, L., Kougias, P. G., Zhu, X. & Angelidaki, I. Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. *Sci Rep* **8**, 1926 (2018).
108. Clooney, A. G. *et al.* Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PLoS One* **11**, e0148028 (2016).
109. Tessler, M. *et al.* Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci Rep* **7**, 6589 (2017).
110. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res* **17**, 377–386 (2007).
111. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**, 811–814 (2012).
112. Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533–1542 (2017).
113. Namiki, T., Hachiya, T., Tanaka, H. & Sakakibara, Y. MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Res* **40**, e155 (2012).
114. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824–834 (2017).
115. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
116. Imelfort, M. *et al.* GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* **2**, e603 (2014).
117. Eren, A. M. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
118. Zhu, A., Sunagawa, S., Mende, D. R. & Bork, P. Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol* **16**, 82 (2015).
119. Eren, A. M. *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* **6**, 3–6 (2021).

120. Papudeshi, B. *et al.* Optimizing and evaluating the reconstruction of Metagenome-assembled microbial genomes. *BMC Genomics* **18**, 915 (2017).
121. Adebayo, A. S. *et al.* The Urinary Tract Microbiome in Older Women Exhibits Host Genetic and Environmental Influences. *Cell Host Microbe* **28**, 298-305.e3 (2020).
122. Hasman, H. *et al.* Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. *J Clin Microbiol* **52**, 139–146 (2014).
123. Garretto, A., Thomas-White, K., Wolfe, A. J. & Putonti, C. Detecting viral genomes in the female urinary microbiome. *J Gen Virol* **99**, 1141–1146 (2018).
124. Santiago-Rodriguez, T. M., Ly, M., Bonilla, N. & Pride, D. T. The human urine virome in association with urinary tract infections. *Front Microbiol* **6**, 14 (2015).
125. Schmidt, K. *et al.* Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* **72**, 104–114 (2017).
126. Barraud, O. *et al.* Shotgun metagenomics for microbiome and resistome detection in septic patients with urinary tract infection. *Int J Antimicrob Agents* **54**, 803–808 (2019).
127. Greub, G. Culturomics: a new approach to study the human microbiome. *Clin Microbiol Infect* **18**, 1157–1159 (2012).
128. Lagier, J.-C. *et al.* Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* **18**, 1185–1193 (2012).
129. Kaur, J. & Kaur, H. Advantages and effectiveness of bacterial culture in medical laboratories. *IJAR* **3**, 8:1028-1039 (2015).
130. Overmann, J., Abt, B. & Sikorski, J. Present and Future of Culturing Bacteria. *Annu Rev Microbiol* **71**, 711–730 (2017).
131. Bonkat, G. *et al.* EAU Guidelines on urological infections. *European Association of Urology* 22-26 (2017).
132. Schmiemann, G., Kniehl, E., Gebhardt, K., Matejczyk, M. M. & Hummers-Pradier, E. The Diagnosis of Urinary Tract Infection. *Dtsch Arztebl Int* **107**, 361–367 (2010).
133. Coorevits, L., Heytens, S., Boelens, J. & Claeys, G. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. *Eur J Clin Microbiol Infect Dis* **36**, 635–639 (2017).

134. Clark, A. E., Kaleta, E. J., Arora, A. & Wolk, D. M. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin Microbiol Rev* **26**, 547–603 (2013).
135. Sauget, M., Valot, B., Bertrand, X. & Hocquet, D. Can MALDI-TOF Mass Spectrometry Reasonably Type Bacteria? *Trends Microbiol* **25**, 447–455 (2017).
136. Singhal, N., Kumar, M., Kanaujia, P. K. & Virdi, J. S. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Front Microbiol* **6**, 791 (2015).
137. Florio, W. *et al.* Detection of Antibiotic-Resistance by MALDI-TOF Mass Spectrometry: An Expanding Area. *Front Cell Infect Microbiol* **10**, 616 (2020).
138. Curtiss, N. *et al.* A case controlled study examining the bladder microbiome in women with Overactive Bladder (OAB) and healthy controls. *Eur J Obstet Gynecol Reprod Biol* **214**, 31–35 (2017).
139. Price, T. K. *et al.* The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. *J Clin Microbiol* **54**, 1216–1222 (2016).
140. Curtiss, N. *et al.* Age, menopausal status and the bladder microbiome. *Eur J Obstet Gynecol Reprod Biol* **228**, 126–129 (2018).
141. Pearce, M. M. *et al.* The female urinary microbiome in urgency urinary incontinence. *Am J Obstet Gynecol* **213**, 347.e1–11 (2015).
142. Jacobs, K. M., Thomas-White, K. J., Hilt, E. E., Wolfe, A. J. & Waters, T. P. Microorganisms Identified in the Maternal Bladder: Discovery of the Maternal Bladder Microbiota. *AJP Rep* **7**, e188–e196 (2017).
143. Gottschick, C. *et al.* The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. Microbiome **14**, 5(1):99 (2017).
144. Bresler, L. *et al.* Female lower urinary tract microbiota do not associate with IC/PBS symptoms: a case-controlled study. *Int Urogynecol J* **30**, 1835–1842 (2019).
145. Price, T. K. *et al.* The urobiome of continent adult women: a cross-sectional study. *BJOG* **127**, 193–201 (2020).
146. Penckofer, S. *et al.* Characteristics of the microbiota in the urine of women with type 2 diabetes. *J Diabetes Complications* **34**, 107561 (2020).
147. Romero, R. *et al.* The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* **2**, 4 (2014).
148. Price, T. K. *et al.* Temporal Dynamics of the Adult Female Lower Urinary Tract Microbiota. *mBio* **11**, e00475-20 (2020).

149. Thomas-White, K. J. *et al.* Evaluation of the urinary microbiota of women with uncomplicated stress urinary incontinence. *Am J Obstet Gynecol* **216**, 55.e1-55.e16 (2017).
150. Komesu, Y. M. *et al.* The urinary microbiome in women with mixed urinary incontinence compared to similarly aged controls. *Int Urogynecol J* **29**, 1785–1795 (2018).
151. Wolff, B. J., Price, T. K., Joyce, C. J., Wolfe, A. J. & Mueller, E. R. Oral probiotics and the female urinary microbiome: a double-blinded randomized placebo-controlled trial. *Int Urol Nephrol* **51**, 2149–2159 (2019).
152. Dong, Q. *et al.* The Microbial Communities in Male First Catch Urine Are Highly Similar to Those in Paired Urethral Swab Specimens. *PLoS One* **6**, 5:e19709 (2011).
153. Kinneman, L. *et al.* Assessment of the Urinary Microbiome in Children Younger Than 48 Months. *Pediatr Infect Dis J* **39**, 565–570 (2020).
154. Kassiri, B. *et al.* A Prospective Study of the Urinary and Gastrointestinal Microbiome in Prepubertal Males. *Urology* **131**, 204–210 (2019).
155. Gorbachinsky, I., Sherertz, R., Russell, G., Krane, L. S. & Hodges, S. J. Altered perineal microbiome is associated with vulvovaginitis and urinary tract infection in preadolescent girls. *Ther Adv Urol* **6**, 224–229 (2014).
156. Hickey, R. J. *et al.* Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women. *mBio* **6**, 2:e00097-15 (2015).
157. Thoma, M. E. *et al.* Longitudinal changes in vaginal microbiota composition assessed by gram stain among never sexually active pre- and postmenarcheal adolescents in Rakai, Uganda. *J Pediatr Adolesc Gynecol* **24**, 42–47 (2011).
158. Ollberding, N. J. *et al.* Urinary Microbiota Associated with Preterm Birth: Results from the Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) Study. *PLoS One* **11**, e0162302 (2016).
159. Yoo, J. Y. *et al.* 16S rRNA gene-based metagenomic analysis reveals differences in bacteria-derived extracellular vesicles in the urine of pregnant and non-pregnant women. *Exp Mol Med* **48**, e208 (2016).
160. Liu, F. *et al.* Alterations in the Urinary Microbiota Are Associated With Cesarean Delivery. *Front Microbiol* **9**, 2193 (2018).
161. Gulur, D. M. & Drake, M. J. Management of overactive bladder. *Nat Rev Urol* **7**, 572–582 (2010).
162. Kay, G. G. *et al.* Antimuscarinic drugs for overactive bladder and their potential effects on cognitive function in older patients. *J Am Geriatr Soc* **53**, 2195–2201 (2005).

163. Nightingale, G. Management of urinary incontinence. *Post Reprod Health* **26**, 63–70 (2020).
164. Nienhouse, V. *et al.* Interplay between Bladder Microbiota and Urinary Antimicrobial Peptides: Mechanisms for Human Urinary Tract Infection Risk and Symptom Severity. *PLoS One* **9**, 12:e114185 (2014).
165. Fok, C. S. *et al.* Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. *Int Urogynecol J* **29**, 1765–1771 (2018).
166. Thomas-White, K. J. *et al.* Urinary microbes and postoperative urinary tract infection risk in urogynecologic surgical patients. *Int Urogynecol J* **29**, 1797–1805 (2018).
167. Balachandran, A. A., Wildman, S. S., Strutt, M. & Duckett, J. Is chronic urinary infection a cause of overactive bladder? *Eur J Obstet Gynecol Reprod Biol* **201**, 108–112 (2016).
168. Gill, K., Kang, R., Sathiananthamoorthy, S., Khasriya, R. & Malone-Lee, J. A blinded observational cohort study of the microbiological ecology associated with pyuria and overactive bladder symptoms. *Int Urogynecol J* **29**, 1493–1500 (2018).
169. Thomas-White, K. *et al.* Vaginal estrogen therapy is associated with increased *Lactobacillus* in the urine of postmenopausal women with overactive bladder symptoms. *Am J Obstet Gynecol* **223**, 727.e1-727.e11 (2020).
170. Siddiqui, H., Lagesen, K., Nederbragt, A. J., Jeansson, S. L. & Jakobsen, K. S. Alterations of microbiota in urine from women with interstitial cystitis. *BMC Microbiology* **12**, 205 (2012).
171. Abernethy, M. G. *et al.* Urinary Microbiome and Cytokine Levels in Women With Interstitial Cystitis. *Obstet Gynecol* **129**, 500–506 (2017).
172. Nickel, J. C. *et al.* Assessment of the Lower Urinary Tract Microbiota during Symptom Flare in Women with Urologic Chronic Pelvic Pain Syndrome: A MAPP Network Study. *J Urol* **195**, 356–362 (2016).
173. Meriwether, K. V. *et al.* The Vaginal and Urinary Microbiomes in Premenopausal Women With Interstitial Cystitis/Bladder Pain Syndrome as Compared to Unaffected Controls: A Pilot Cross-Sectional Study. *Front Cell Infect Microbiol* **9**, 92 (2019).
174. Nickel, J. C. *et al.* Urinary fungi associated with urinary symptom severity among women with interstitial cystitis/bladder pain syndrome (IC/BPS). *World J Urol* **38**, 433–446 (2020).
175. Ananthakrishnan, A. N. *et al.* Environmental triggers in IBD: a review of progress and evidence. *Nat Rev Gastroenterol Hepatol* **15**, 39–49 (2018).

176. McWilliams, D. F. *et al.* Disease activity flares and pain flares in an early rheumatoid arthritis inception cohort; characteristics, antecedents and sequelae. *BMC Rheumatology* **3**, 49 (2019).
177. Hooton, T. M., Roberts, P. L., Cox, M. E. & Stapleton, A. E. Voided Midstream Urine Culture and Acute Cystitis in Premenopausal Women. *N Engl J Med* **369**, 1883–1891 (2013).
178. Yu, Y. *et al.* Diagnosing inflammation and infection in the urinary system via proteomics. *J Transl Med* **13**, 111 (2015).
179. Frølund, M. *et al.* The bacterial microbiota in first-void urine from men with and without idiopathic urethritis. *PLoS One* **13**, e0201380 (2018).
180. You, C. *et al.* The first report: An analysis of bacterial flora of the first voided urine specimens of patients with male urethritis using the 16S ribosomal RNA gene-based clone library method. *Microb Pathog* **95**, 95–100 (2016).
181. Yoo, J.-J. *et al.* Urinary Microbiome Characteristics in Female Patients with Acute Uncomplicated Cystitis and Recurrent Cystitis. *J Clin Med* **10**, 1097 (2021).
182. Hochstedler, B. R. *et al.* Urinary microbiota of women with recurrent urinary tract infection: collection and culture methods. *Int Urogynecol J* (2021) doi:10.1007/s00192-021-04780-4.
183. Liu, F. *et al.* Characterization of the urinary microbiota of elderly women and the effects of type 2 diabetes and urinary tract infections on the microbiota. *Oncotarget* **8**, 100678–100690 (2017).
184. Liu, F. *et al.* Alterations of Urinary Microbiota in Type 2 Diabetes Mellitus with Hypertension and/or Hyperlipidemia. *Front Physiol* **8**, 126 (2017).
185. Ling, Z., Liu, F., Shao, L., Cheng, Y. & Li, L. Dysbiosis of the Urinary Microbiota Associated With Urine Levels of Proinflammatory Chemokine Interleukin-8 in Female Type 2 Diabetic Patients. *Front Immunol* **8**, 1032 (2017).
186. Forster, C. S. *et al.* A cross-sectional analysis of the urine microbiome of children with neuropathic bladders. *J Pediatr Urol* **16**, 593.e1-593.e8 (2020).
187. Groah, S. L. *et al.* Redefining Healthy Urine: A Cross-Sectional Exploratory Metagenomic Study of People With and Without Bladder Dysfunction. *J Urol* **196**, 579–587 (2016).
188. Oresta, B. *et al.* The Microbiome of Catheter Collected Urine in Males with Bladder Cancer According to Disease Stage. *J Urol* **205**, 86–93 (2021).

189. Hussein, A. A. *et al.* Investigating the association between the urinary microbiome and bladder cancer: An exploratory study. *Urol Oncol* **39**, 370.e9-370.e19 (2021).
190. Chipollini, J. *et al.* Characterization of urinary microbiome in patients with bladder cancer: Results from a single-institution, feasibility study. *Urol Oncol* **38**, 615–621 (2020).
191. Mansour, B. *et al.* Bladder cancer-related microbiota: examining differences in urine and tissue samples. *Sci Rep* **10**, 11042 (2020).
192. Bi, H. *et al.* Urinary microbiota – a potential biomarker and therapeutic target for bladder cancer. *J Med Microbiol* **68**, 1471–1478 (2019).
193. Bučević Popović, V. *et al.* The urinary microbiome associated with bladder cancer. *Sci Rep* **8**, 12157 (2018).
194. Hourigan, S. K. *et al.* Studying the urine microbiome in superficial bladder cancer: samples obtained by midstream voiding versus cystoscopy. *BMC Urol* **20**, 5 (2020).
195. Wu, P. *et al.* Profiling the Urinary Microbiota in Male Patients With Bladder Cancer in China. *Front Cell Infect Microbiol* **8**, 167 (2018).
196. Mai, G. *et al.* Common Core Bacterial Biomarkers of Bladder Cancer Based on Multiple Datasets. *BioMed Res Int* **2019**, e4824909 (2019).
197. Liu, F. *et al.* Dysbiosis signatures of the microbial profile in tissue from bladder cancer. *Cancer Med* **8**, 6904–6914 (2019).
198. Pederzoli, F. *et al.* Sex-specific Alterations in the Urinary and Tissue Microbiome in Therapy-naïve Urothelial Bladder Cancer Patients. *Eur Urol Oncol* **3**, 784–788 (2020).
199. Zeng, J. *et al.* Alterations in Urobiome in Patients With Bladder Cancer and Implications for Clinical Outcome: A Single-Institution Study. *Front Cell Infect Microbiol* **10**, 555508 (2020).
200. Bajic, P., Wolfe, A. J. & Gupta, G. N. The Urinary Microbiome: Implications in Bladder Cancer Pathogenesis and Therapeutics. *Urology* **126**, 10–15 (2019).
201. Shrestha, E. *et al.* Profiling the Urinary Microbiome in Men with Positive versus Negative Biopsies for Prostate Cancer. *J Urol* **199**, 161–171 (2018).
202. Alanee, S. *et al.* A prospective study to examine the association of the urinary and fecal microbiota with prostate cancer diagnosis after transrectal biopsy of the prostate using 16sRNA gene analysis. *The Prostate* **79**, 81–87 (2019).
203. Wang, H. *et al.* Breast tissue, oral and urinary microbiomes in breast cancer. *Oncotarget* **8**, 88122–88138 (2017).

204. Liu, F. *et al.* Characteristics of the urinary microbiome in kidney stone patients with hypertension. *J Transl Med* **18**, 130 (2020).
205. Liu, F. *et al.* The pelvis urinary microbiome in patients with kidney stones and clinical associations. *BMC Microbiol* **20**, 336 (2020).
206. Xie, J. *et al.* Profiling the urinary microbiome in men with calcium-based kidney stones. *BMC Microbiol* **20**, 41 (2020).
207. Zampini, A., Nguyen, A. H., Rose, E., Monga, M. & Miller, A. W. Defining Dysbiosis in Patients with Urolithiasis. *Sci Rep* **9**, 5425 (2019).
208. Kachroo, N. *et al.* Standardization of microbiome studies for urolithiasis: an international consensus agreement. *Nat Rev Urol* **18**, 303–311 (2021).
209. Dornbier, R. A. *et al.* The microbiome of calcium-based urinary stones. *Urolithiasis* **48**, 191–199 (2020).
210. Kramer, H. *et al.* Diversity of the midstream urine microbiome in adults with chronic kidney disease. *Int Urol Nephrol* **50**, 1123–1130 (2018).
211. Rani, A. *et al.* A diverse virome in kidney transplant patients contains multiple viral subtypes with distinct polymorphisms. *Sci Rep* **6**, 33327 (2016).
212. Fricke, W. F., Maddox, C., Song, Y. & Bromberg, J. S. Human microbiota characterization in the course of renal transplantation. *Am J Transplant* **14**, 416–427 (2014).
213. Wu, J. F. *et al.* Urinary microbiome associated with chronic allograft dysfunction in kidney transplant recipients. *Clin Transplant* **32**, e13436 (2018).
214. Colas, L. *et al.* Unique and specific Proteobacteria diversity in urinary microbiota of tolerant kidney transplanted recipients. *Am J Transplant* **20**, 145–158 (2020).
215. Gerges-Knafl, D. *et al.* The urinary microbiome shows different bacterial genera in renal transplant recipients and non-transplant patients at time of acute kidney injury - a pilot study. *BMC Nephrol* **21**, 117 (2020).
216. Ning, Y. *et al.* Characteristics of the Urinary Microbiome From Patients With Gout: A Prospective Study. *Front. Endocrinol.* **11**, 272 (2020).
217. Mändar, R. *et al.* Seminal microbiome in men with and without prostatitis. *Int J Urol* **24**, 211–216 (2017).
218. Nickel, J. C. *et al.* Search for Microorganisms in Men with Urologic Chronic Pelvic Pain Syndrome: A Culture-Independent Analysis in the MAPP Research Network. *J Urol* **194**, 127–135 (2015).

219. Shoskes, D. A. *et al.* The Urinary Microbiome Differs Significantly Between Patients With Chronic Prostatitis/Chronic Pelvic Pain Syndrome and Controls as Well as Between Patients With Different Clinical Phenotypes. *Urology* **92**, 26–32 (2016).
220. Jiang, S. *et al.* Dysbiosis of urine microbiota in obstructive urinary retention patients revealed by next-generation sequencing. *Ann Clin Microbiol Antimicrob* **20**, 2 (2021).
221. Chiu, I. M. *et al.* Bacteria activate sensory neurons that modulate pain and inflammation. *Nature* **501**, 52–57 (2013).
222. Brubaker, L. *et al.* Urinary bacteria in adult women with urgency urinary incontinence. *Int Urogynecol J* **25**, 1179–1184 (2014).
223. Komesu, Y. M. *et al.* Defining the relationship between vaginal and urinary microbiomes. *American Journal of Obstetrics & Gynecology* **222**, 154.e1-154.e10 (2020).
224. Price, T. K. *et al.* Bladder bacterial diversity differs in continent and incontinent women: a cross-sectional study. *Am J Obstet Gynecol* **223**, 729.e1-729.e10 (2020).
225. Siddiqui, H. *et al.* Pathogens in Urine from a Female Patient with Overactive Bladder Syndrome Detected by Culture-independent High Throughput Sequencing: A Case Report. *Open Microbiol J* **8**, 148–153 (2014).
226. Nickel, J. C. *et al.* A Culture-Independent Analysis of the Microbiota of Female Interstitial Cystitis/Bladder Pain Syndrome Participants in the MAPP Research Network. *J Clin Med* **8**, 415 (2019).
227. Miller-Ensminger, T. *et al.* Bacteriophages of the Urinary Microbiome. *J Bacteriol* **200**, e00738-17 (2018).
228. Bajic, P. *et al.* Male Bladder Microbiome Relates to Lower Urinary Tract Symptoms. *Eur Urol Focus* **6**, 376–382 (2020).
229. Chen, C. *et al.* The female urinary microbiota in relation to the reproductive tract microbiota. *Gigabyte* **2020**, 1–9 (2020).
230. Felis, G. E. & Pot, B. The family Lactobacillaceae, Lactic Acid Bacteria: Biodiversity and Taxonomy. *John Wiley & Sons* 245–247 (2014).
231. Zheng, J. *et al.* A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *Int J Syst Evol Microbiol* **70**, 2782–2858 (2020).
232. Mokoena, M. P. Lactic Acid Bacteria and Their Bacteriocins: Classification, Biosynthesis and Applications against Uropathogens: A Mini-Review. *Molecules* **22**, 1255 (2017).

233. Castro-Bravo, N., Wells, J. M., Margolles, A. & Ruas-Madiedo, P. Interactions of Surface Exopolysaccharides From Bifidobacterium and *Lactobacillus* Within the Intestinal Environment. *Front Microbiol* **9**, 2426 (2018).
234. Chee, W. J. Y., Chew, S. Y. & Than, L. T. L. Vaginal microbiota and the potential of *Lactobacillus* derivatives in maintaining vaginal health. *Microbial Cell Factories* **19**, 203 (2020).
235. Chapot-Chartier, M.-P. Interactions of the cell-wall glycopolymers of lactic acid bacteria with their bacteriophages. *Front Microbiol* **5**, 236 (2014).
236. Laiño, J., Villena, J., Kanmani, P. & Kitazawa, H. Immunoregulatory Effects Triggered by Lactic Acid Bacteria Exopolysaccharides: New Insights into Molecular Interactions with Host Cells. *Microorganisms* **4**, 27 (2016).
237. Salvetti, E., Harris, H. M. B., Felis, G. E. & O'Toole, P. W. Comparative Genomics of the Genus *Lactobacillus* Reveals Robust Phylogroups That Provide the Basis for Reclassification. *Appl. Environ. Microbiol.* **84**, e00993-18 (2018).
238. Witkin, S. S. & Linhares, I. M. Why do lactobacilli dominate the human vaginal microbiota? *BJOG* **124**, 606–611 (2017).
239. Duar, R. M. *et al.* Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* **41**, S27–S48 (2017).
240. Dzidic, M. *et al.* Oral microbiome development during childhood: an ecological succession influenced by postnatal factors and associated with tooth decay. *The ISME Journal* **12**, 2292–2306 (2018).
241. Dal Bello, F. & Hertel, C. Oral cavity as natural reservoir for intestinal lactobacilli. *Syst Appl Microbiol* **29**, 69–76 (2006).
242. Mu, Q., Tavella, V. J. & Luo, X. M. Role of *Lactobacillus reuteri* in Human Health and Diseases. *Front Microbiol* **9**, 757 (2018).
243. Rodríguez-Sojo, M. J., Ruiz-Malagón, A. J., Rodríguez-Cabezas, M. E., Gálvez, J. & Rodríguez-Nogales, A. *Limosilactobacillus fermentum* CECT5716: Mechanisms and Therapeutic Insights. *Nutrients* **13**, 1016 (2021).
244. Terai, T. *et al.* Safety assessment of the candidate oral probiotic *Lactobacillus crispatus* YIT 12319: Analysis of antibiotic resistance and virulence-associated genes. *Food Chem Toxicol* **140**, 111278 (2020).
245. Salvetti, E. & O'Toole, P. W. The Genomic Basis of Lactobacilli as Health-Promoting Organisms. *Microbiol Spectr* **5**, (2017) doi: 10.1128/microbiolspec.BAD-0011-2016.

246. Eberly, A. R. *et al.* Biofilm Formation by Uropathogenic *Escherichia coli* Is Favored under Oxygen Conditions That Mimic the Bladder Environment. *Int J Mol Sci* **18**, 2077 (2017).
247. Ceccarani, C. *et al.* Diversity of vaginal microbiome and metabolome during genital infections. *Sci Rep* **9**, 14095 (2019).
248. O'Hanlon, D. E., Moench, T. R. & Cone, R. A. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PLoS One* **8**, e80074 (2013).
249. Pino, A., Bartolo, E., Caggia, C., Cianci, A. & Randazzo, C. L. Detection of vaginal lactobacilli as probiotic candidates. *Sci Rep* **9**, 3355 (2019).
250. Ragaliauskas, T. *et al.* Inerolysin and vaginolysin, the cytolysins implicated in vaginal dysbiosis, differently impair molecular integrity of phospholipid membranes. *Sci Rep* **9**, 10606 (2019).
251. France, M. T., Mendes-Soares, H. & Forney, L. J. Genomic Comparisons of *Lactobacillus crispatus* and *Lactobacillus iners* Reveal Potential Ecological Drivers of Community Composition in the Vagina. *Appl. Environ. Microbiol.* **82**, 7063–7073 (2016).
252. Petrova, M. I., Reid, G., Vaneechoutte, M. & Lebeer, S. *Lactobacillus iners*: Friend or Foe? *Trends Microbiol* **25**, 182–191 (2017).
253. Vaneechoutte, M. *Lactobacillus iners*, the unusual suspect. *Res Microbiol* **168**, 826–836 (2017).
254. van Houdt, R. *et al. Lactobacillus iners*-dominated vaginal microbiota is associated with increased susceptibility to *Chlamydia trachomatis* infection in Dutch women, a case control study. *Sex Transm Infect* **94**, 117–123 (2018).
255. Putonti, C. *et al.* Draft Genome Sequences of Six *Lactobacillus gasseri* and Three *Lactobacillus paragasseri* Strains Isolated from the Female Bladder. *Microbiol Resour Announc* **8**, e00973-19 (2019).
256. Nasioudis, D. *et al.* The composition of the vaginal microbiome in first trimester pregnant women influences the level of autophagy and stress in vaginal epithelial cells. *J Reprod Immunol* **123**, 35–39 (2017).
257. Edelman, S. M. *et al.* Identification of a high-molecular-mass *Lactobacillus* epithelium adhesin (LEA) of *Lactobacillus crispatus* ST1 that binds to stratified squamous epithelium. *Microbiology* **158**, 1713–1722 (2012).
258. Bönisch, E. *et al.* Lipoteichoic acid mediates binding of a *Lactobacillus* S-layer protein. *Glycobiology* **28**, 148–158 (2018).

259. Sun, Z. *et al.* Characterization of a S-layer protein from *Lactobacillus crispatus* K313 and the domains responsible for binding to cell wall and adherence to collagen. *Appl Microbiol Biotechnol* **97**, 1941–1952 (2013).
260. Abramov, V. *et al.* Probiotic Properties of *Lactobacillus crispatus* 2,029: Homeostatic Interaction with Cervicovaginal Epithelial Cells and Antagonistic Activity to Genitourinary Pathogens. *Probiotics Antimicrob Proteins* **6**, 165–176 (2014).
261. Muscariello, L., De Siena, B. & Marasco, R. *Lactobacillus* Cell Surface Proteins Involved in Interaction with Mucus and Extracellular Matrix Components. *Curr Microbiol* **77**, 3831–3841 (2020).
262. Vélez, M. P., De Keersmaecker, S. C. J. & Vanderleyden, J. Adherence factors of *Lactobacillus* in the human gastrointestinal tract. *FEMS Microbiol Lett* **276**, 140–148 (2007).
263. Ojala, T. *et al.* Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*. *BMC Genomics* **15**, 1070 (2014).
264. Ghartey, J. P. *et al. Lactobacillus crispatus* Dominant Vaginal Microbiome Is Associated with Inhibitory Activity of Female Genital Tract Secretions against *Escherichia coli*. *PLoS One* **9**, e96659 (2014).
265. Atassi, F., Pho Viet Ahn, D. L. & Lievin-Le Moal, V. Diverse Expression of Antimicrobial Activities Against Bacterial Vaginosis and Urinary Tract Infection Pathogens by Cervicovaginal Microbiota Strains of *Lactobacillus gasseri* and *Lactobacillus crispatus*. *Front. Microbiol.* **10**, 2900 (2019).
266. Marziali, G., Foschi, C., Parolin, C., Vitali, B. & Marangoni, A. In-vitro effect of vaginal lactobacilli against group B *Streptococcus*. *Microb Pathog* **136**, 103692 (2019).
267. Kalyoussef, S. *et al. Lactobacillus* proteins are associated with the bactericidal activity against *E. coli* of female genital tract secretions. *PLoS One* **7**, e49506 (2012).
268. Barrett, E., Ross, R. P., O'Toole, P. W., Fitzgerald, G. F. & Stanton, C. γ-Aminobutyric acid production by culturable bacteria from the human intestine. *J Appl Microbiol* **113**, 411–417 (2012).
269. Boonstra, E. *et al.* Neurotransmitters as food supplements: the effects of GABA on brain and behavior. *Front Psychol* **6**, 1520 (2015).
270. Yunes, R. A. *et al.* GABA production and structure of *gadB*/*gadC* genes in *Lactobacillus* and *Bifidobacterium* strains from human microbiota. *Anaerobe* **42**, 197–204 (2016).

271. Dinan, T. G., Stanton, C. & Cryan, J. F. Psychobiotics: a novel class of psychotropic. *Biol Psychiatry* **74**, 720–726 (2013).
272. Puglisi, E., Galletti, S., Sagheddu, V. & Elli, M. Genome Sequence Announcement of *Lactobacillus vaginalis* LMG S-26419, Isolated from a Healthy Woman. *Microbiol Resour Announc* **8**, e01534-18 (2019).
273. Jia, Y. *et al.* Comparative Genomics Analysis of *Lactobacillus mucosae* from Different Niches. *Genes (Basel)* **11**, 95 (2020).
274. Pasolli, E. *et al.* Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nat Commun* **11**, 2610 (2020).
275. Al-Hadidi, A., Navarro, J., Goodman, S. D., Bailey, M. T. & Besner, G. E. *Lactobacillus reuteri* in Its Biofilm State Improves Protection from Experimental Necrotizing Enterocolitis. *Nutrients* **13**, 918 (2021).
276. Liu, Y., Fatheree, N. Y., Mangalat, N. & Rhoads, J. M. Human-derived probiotic *Lactobacillus reuteri* strains differentially reduce intestinal inflammation. *Am J Physiol Gastrointest Liver Physiol* **299**, G1087-1096 (2010).
277. Vicariotto, F., Mogna, L. & Del Piano, M. Effectiveness of the Two Microorganisms *Lactobacillus fermentum* LF15 and *Lactobacillus plantarum* LP01, Formulated in Slow-release Vaginal Tablets, in Women Affected by Bacterial Vaginosis: A Pilot Study. *J Clin Gastroenterol* **48**, S106 (2014).
278. Miljkovic, M., Thomas, M., Serror, P., Rigottier-Gois, L. & Kojic, M. Binding activity to intestinal cells and transient colonization in mice of two *Lactobacillus paracasei* subsp. *paracasei* strains with high aggregation potential. *World J Microbiol Biotechnol* **35**, 85 (2019).
279. Nishiyama, K. *et al.* Adhesion properties of *Lactobacillus rhamnosus* mucus-binding factor to mucin and extracellular matrix proteins. *Biosci Biotechnol Biochem* **79**, 271–279 (2015).
280. Turovskiy, Y., Ludescher, R. D., Aroutcheva, A. A., Faro, S. & Chikindas, M. L. Lactocin 160, a Bacteriocin Produced by Vaginal *Lactobacillus rhamnosus*, Targets Cytoplasmic Membranes of the Vaginal Pathogen, *Gardnerella vaginalis*. *Probiotics Antimicrob Proteins* **1**, 67–74 (2009).
281. Bertuccini, L., Russo, R., Iosi, F. & Superti, F. Effects of *Lactobacillus rhamnosus* and *Lactobacillus acidophilus* on bacterial vaginal pathogens. *Int J Immunopathol Pharmacol* **30**, 163–167 (2017).

282. Allonsius, C. N. *et al.* Inhibition of *Candida albicans* morphogenesis by chitinase from *Lactobacillus rhamnosus* GG. *Sci Rep* **9**, 2900 (2019).
283. Chen, C.-C. *et al.* Antimicrobial Activity of *Lactobacillus* Species Against Carbapenem-Resistant Enterobacteriaceae. *Front. Microbiol.* **10**, 789 (2019).
284. Homayouni, A. *et al.* Effects of probiotics on the recurrence of bacterial vaginosis: a review. *J Low Genit Tract Dis* **18**, 79–86 (2014).
285. Marcotte, H. *et al.* Characterization and complete genome sequences of *L. rhamnosus* DSM 14870 and *L. gasseri* DSM 14869 contained in the EcoVag® probiotic vaginal capsules. *Microbiol Res* **205**, 88–98 (2017).
286. Macklaim, J. M., Clemente, J. C., Knight, R., Gloor, G. B. & Reid, G. Changes in vaginal microbiota following antimicrobial and probiotic therapy. *Microb Ecol Health Dis* **26**, 27799 (2015).
287. Vaneechoutte, M. *et al.* Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. *Int. J. Syst. Evol. Microbiol.* **69**, 679–687 (2019).
288. Schellenberg, J. J., Patterson, M. H. & Hill, J. E. *Gardnerella vaginalis* diversity and ecology in relation to vaginal symptoms. *Res Microbiol* **168**, 837–844 (2017).
289. Hill, J. E., Albert, A. Y. K. & VOGUE Research Group. Resolution and Cooccurrence Patterns of *Gardnerella leopoldii*, *G. swidsinskii*, *G. piotii*, and *G. vaginalis* within the Vaginal Microbiome. *Infect Immun* **87**, e00532-19 (2019).
290. Castro, J., Jefferson, K. K. & Cerca, N. Genetic Heterogeneity and Taxonomic Diversity among *Gardnerella* Species. *Trends Microbiol* **28**, 202–211 (2020).
291. Pleckaityte, M. Cholesterol-Dependent Cytolysins Produced by Vaginal Bacteria: Certainties and Controversies. *Front. Cell. Infect. Microbiol.* **9**, 452 (2020).
292. Gelber, S. E., Aguilar, J. L., Lewis, K. L. T. & Ratner, A. J. Functional and Phylogenetic Characterization of Vaginolysin, the Human-Specific Cytolysin from *Gardnerella vaginalis*. *J Bacteriol* **190**, 3896–3903 (2008).
293. Hardy, L., Cerca, N., Jespers, V., Vaneechoutte, M. & Crucitti, T. Bacterial biofilms in the vagina. *Res Microbiol* **168**, 865–874 (2017).
294. Cerca, N. Could targeting neighboring bacterial populations help combat bacterial vaginosis? *Future Microbiol* **14**, 365–368 (2019).

295. Breshears, L. M., Edwards, V. L., Ravel, J. & Peterson, M. L. *Lactobacillus crispatus* inhibits growth of *Gardnerella vaginalis* and *Neisseria gonorrhoeae* on a porcine vaginal mucosa model. *BMC Microbiol* **15**, 276 (2015).
296. Castro, J., Rosca, A. S., Cools, P., Vaneechoutte, M. & Cerca, N. *Gardnerella vaginalis* Enhances *Atopobium vaginae* Viability in an in vitro Model. *Front Cell Infect Microbiol* **10**, 83 (2020).
297. Castro, J., Machado, D. & Cerca, N. Unveiling the role of *Gardnerella vaginalis* in polymicrobial Bacterial Vaginosis biofilms: the impact of other vaginal pathogens living as neighbors. *ISME J* **13**, 1306–1317 (2019).
298. Lewis, A. L. & Gilbert, N. M. Roles of the vagina and the vaginal microbiota in urinary tract infection: evidence from clinical correlations and experimental models. *GMS Infect Dis* **8**, Doc02 (2020).
299. O'Brien, V. P., Joens, M. S., Lewis, A. L. & Gilbert, N. M. Recurrent *Escherichia coli* Urinary Tract Infection Triggered by *Gardnerella vaginalis* Bladder Exposure in Mice. *J Vis Exp* **166**, 10.3791/61967 (2020).
300. Gilbert, N. M., O'Brien, V. P. & Lewis, A. L. Transient microbiota exposures activate dormant *Escherichia coli* infection in the bladder and drive severe outcomes of recurrent disease. *PLOS Pathog* **13**, e1006238 (2017).
301. Gilbert, N. M. & Lewis, A. L. Covert pathogenesis: Transient exposures to microbes as triggers of disease. *PLoS Pathog* **15**, e1007586 (2019).
302. Putonti, C. *et al.* Genome Investigation of Urinary *Gardnerella* Strains and Their Relationship to Isolates of the Vaginal Microbiota. *mSphere* **6**, e00154-21 (2021).
303. Malki, K. *et al.* Genomes of *Gardnerella* Strains Reveal an Abundance of Prophages within the Bladder Microbiome. *PLoS One* **11**, e0166757 (2016).
304. Adeolu, M., Alnajar, S., Naushad, S. & S Gupta, R. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol* **66**, 5575–5599 (2016).
305. Brannon, J. R. *et al.* Invasion of vaginal epithelial cells by uropathogenic *Escherichia coli*. *Nat Commun* **11**, 2803 (2020).
306. Terlizzi, M. E., Gribaudo, G. & Maffei, M. E. UroPathogenic *Escherichia coli* (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. *Front Microbiol* **8**, 1566 (2017).

307. Murray, B. O. *et al.* Recurrent Urinary Tract Infection: A Mystery in Search of Better Model Systems. *Front Cell Infect Microbiol* **11**, 691210 (2021).
308. Haugan, M. S. *et al.* Growth Rate of *Escherichia coli* During Human Urinary Tract Infection: Implications for Antibiotic Effect. *Antibiotics* **8**, 92 (2019).
309. Reitzer, L. & Zimmern, P. Rapid Growth and Metabolism of Uropathogenic *Escherichia coli* in Relation to Urine Composition. *Clin Microbiol Rev* **33**, e00101-19 (2019).
310. Subashchandrabose, S. & Mobley, H. L. T. Virulence and Fitness Determinants of Uropathogenic *Escherichia coli*. *Microbiol Spectr* **3**, (2015) doi: 10.1128/microbiolspec.UTI-0015-2012.
311. Manges, A. R. *et al.* Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. *Clin. Microbiol. Rev.* **32**, 3 (2019).
312. Flament-Simon, S.-C. *et al.* Clonal Structure, Virulence Factor-encoding Genes and Antibiotic Resistance of *Escherichia coli*, Causing Urinary Tract Infections and Other Extraintestinal Infections in Humans in Spain and France during 2016. *Antibiotics (Basel)* **9**, 161 (2020).
313. Fibke, C. D. *et al.* Genomic Epidemiology of Major Extraintestinal Pathogenic *Escherichia coli* Lineages Causing Urinary Tract Infections in Young Women Across Canada. *Open Forum Infect Dis* **6**, ofz431 (2019).
314. Tabasi, M., Karam, M. R. A., Habibi, M., Mostafavi, E. & Bouzari, S. Genotypic Characterization of Virulence Factors in *Escherichia coli* Isolated from Patients with Acute Cystitis, Pyelonephritis and Asymptomatic Bacteriuria. *J Clin Diagn Res* **10**, DC01–DC07 (2016).
315. Luna-Pineda, V. M. *et al.* Curli of Uropathogenic *Escherichia coli* Enhance Urinary Tract Colonization as a Fitness Factor. *Front Microbiol* **10**, 2063 (2019).
316. Bien, J., Sokolova, O. & Bozko, P. Role of Uropathogenic *Escherichia coli* Virulence Factors in Development of Urinary Tract Infection and Kidney Damage. *Int J Nephrol* **2012**, 681473 (2012).
317. Tamadonfar, K. O., Omattage, N. S., Spaulding, C. N. & Hultgren, S. J. Reaching the End of the Line: Urinary Tract Infections. *Microbiol Spectr* **7**, 3 (2019).
318. Zhao, F., Yang, H., Bi, D., Khaledi, A. & Qiao, M. A systematic review and meta-analysis of antibiotic resistance patterns, and the correlation between biofilm formation with virulence factors in uropathogenic *E. coli* isolated from urinary tract infections. *Microb Pathog* **144**, 104196 (2020).

319. O'Brien, V. P. *et al.* A mucosal imprint left by prior *Escherichia coli* bladder infection sensitizes to recurrent disease. *Nat Microbiol* **2**, 16196 (2016).
320. Yang, C.-Y. *et al.* Oral Microbiota Community Dynamics Associated With Oral Squamous Cell Carcinoma Staging. *Front Microbiol* **9**, 862 (2018).
321. Moschioni, M., Pansegrau, W. & Barocchi, M. A. Adhesion determinants of the *Streptococcus* species. *Microb Biotechnol* **3**, 370–388 (2010).
322. Asam, D. & Spellerberg, B. Molecular pathogenicity of *Streptococcus anginosus*. *Mol Oral Microbiol* **29**, 145–155 (2014).
323. Nobbs, A. H., Lamont, R. J. & Jenkinson, H. F. *Streptococcus* adherence and colonization. *Microbiol Mol Biol Rev* **73**, 407–450 (2009).
324. Götz, F., Perconti, S., Popella, P., Werner, R. & Schlag, M. Epidermin and gallidermin: Staphylococcal lantibiotics. *Int J Med Microbiol* **304**, 63–71 (2014).
325. Kim, P. I. *et al.* Characterization and structure identification of an antimicrobial peptide, hominicin, produced by *Staphylococcus hominis* MBBL 2-9. *Biochem Biophys Res Commun* **399**, 133–138 (2010).
326. Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511–516 (2016).
327. Iwase, T. *et al. Staphylococcus epidermidis* Esp inhibits *Staphylococcus aureus* biofilm formation and nasal colonization. *Nature* **465**, 346–349 (2010).
328. Sugimoto, S. *et al. Staphylococcus epidermidis* Esp degrades specific proteins associated with *Staphylococcus aureus* biofilm formation and host-pathogen interaction. *J Bacteriol* **195**, 1645–1655 (2013).
329. Arciola, C. R., Campoccia, D., Ravaioli, S. & Montanaro, L. Polysaccharide intercellular adhesin in biofilm: structural and regulatory aspects. *Front Cell Infect Microbiol* **5**, 7 (2015).
330. Giormezis, N. *et al.* Coagulase-negative staphylococcal bloodstream and prosthetic-device-associated infections: the role of biofilm formation and distribution of adhesin and toxin genes. *J Med Microbiol* **63**, 1500–1508 (2014).
331. Josse, J., Laurent, F. & Diot, A. Staphylococcal Adhesion and Host Cell Invasion: Fibronectin-Binding and Other Mechanisms. *Front Microbiol* **8**, 2433 (2017).
332. Namvar, A. E. *et al.* Clinical characteristics of *Staphylococcus epidermidis*: a systematic review. *GMS Hyg Infect Control* **9**, Doc23 (2014).

333. Limoli, D. H., Jones, C. J. & Wozniak, D. J. Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. *Microbiol Spectr* **3**, (2015) doi: 10.1128/microbiolspec.MB-0011-2014.
334. Flemming, H.-C., Neu, T. R. & Wozniak, D. J. The EPS matrix: the 'house of biofilm cells'. *J Bacteriol* **189**, 7945–7947 (2007).
335. Burkovski, A. Cell envelope of corynebacteria: structure and influence on pathogenicity. *ISRN Microbiol* **2013**, 935736 (2013).
336. Burkovski, A. The role of corynomycolic acids in *Corynebacterium*-host interaction. *Antonie Van Leeuwenhoek* **111**, 717–725 (2018).
337. Tauch, A. & Burkovski, A. Molecular armory or niche factors: virulence determinants of *Corynebacterium* species. *FEMS Microbiol Lett* **362**, fnv185 (2015).
338. Chiller, K., Selkin, B. A. & Murakawa, G. J. Skin microflora and bacterial infections of the skin. *J Investig Dermatol Symp Proc* **6**, 170–174 (2001).
339. Kaźmierczak, A. K., Szarapińska-Kwaszewska, J. K. & Szewczyk, E. M. Opportunistic coryneform organisms--residents of human skin. *Pol J Microbiol* **54**, 27–35 (2005).
340. Ramana, K. V. *et al.* Non Diphtheritic Corynebacteria (NDC) and Their Clinical Significance: Clinical Microbiologist's Perspective. *AJEID* **2**, 83–87 (2014).
341. Altonsy, M. O. *et al. Corynebacterium tuberculostearicum*, a human skin colonizer, induces the canonical nuclear factor-κB inflammatory signaling pathway in human skin cells. *Immun Inflamm Dis* **8**, 62–79 (2020).
342. Hinić, V. *et al. Corynebacterium tuberculostearicum*: a Potentially Misidentified and Multiresistant *Corynebacterium* Species Isolated from Clinical Specimens. *J Clin Microbiol* **50**, 2561–2567 (2012).
343. Omland, S. H., Ekenberg, C., Henrik-Nielsen, R. & Friis-Møller, A. Pancreatic panniculitis complicated by infection with *Corynebacterium tuberculostearicum*: A case report. *IDCases* **1**, 45–46 (2014).
344. Kline, K. A., Fälker, S., Dahlberg, S., Normark, S. & Henriques-Normark, B. Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* **5**, 580–592 (2009).
345. Oliveira, A. *et al.* Insight of Genus *Corynebacterium*: Ascertaining the Role of Pathogenic and Non-pathogenic Species. *Front Microbiol* **8**, 1937 (2017).
346. Ramsey, M. M., Freire, M. O., Gabrilska, R. A., Rumbaugh, K. P. & Lemon, K. P. *Staphylococcus aureus* Shifts toward Commensalism in Response to *Corynebacterium* Species. *Front Microbiol* **7**, 1230 (2016).

347. Yan, M. *et al.* Nasal microenvironments and interspecific interactions influence nasal microbiota complexity and *S. aureus* carriage. *Cell Host Microbe* **14**, 631–640 (2013).
348. Gupta, V. K., Chaudhari, N. M., Iskepalli, S. & Dutta, C. Divergences in gene repertoire among the reference *Prevotella* genomes derived from distinct body sites of human. *BMC Genomics* **16**, 153 (2015).
349. Tett, A., Pasolli, E., Masetti, G., Ercolini, D. & Segata, N. *Prevotella* diversity, niches and interactions with the human host. *Nat Rev Microbiol* **19**, 585-599 (2021).
350. Zhang, Y. *et al.* Population-Genomic Insights into Variation in *Prevotella intermedia* and *Prevotella nigrescens* Isolates and Its Association with Periodontal Disease. *Front. Cell. Infect. Microbiol.* **7**, 409 (2017).
351. Eiring, P., Waller, K., Widmann, A. & Werner, H. Fibronectin and laminin binding of urogenital and oral *Prevotella* species. *Zentralbl Bakteriol* **288**, 361–372 (1998).
352. Randis, T. M. & Ratner, A. J. *Gardnerella* and *Prevotella*: Co-conspirators in the Pathogenesis of Bacterial Vaginosis. *J Infect Dis* **220**, 1085-1088 (2019).
353. Gilbert, N. M. *et al. Gardnerella vaginalis* and *Prevotella bivia* Trigger Distinct and Overlapping Phenotypes in a Mouse Model of Bacterial Vaginosis. *J Infect Dis* **220**, 1099–1108 (2019).
354. Filardo, S. *et al.* Selected Immunological Mediators and Cervical Microbial Signatures in Women with *Chlamydia trachomatis* Infection. *mSystems* **4**, e00094-19 (2019).
355. Alauzet, C., Marchandin, H. & Lozniewski, A. New insights into *Prevotella* diversity and medical microbiology. *Future Microbiol* **5**, 1695–1718 (2010).
356. Castro, J., Rosca, A. S., Muzny, C. A. & Cerca, N. *Atopobium vaginae* and *Prevotella bivia* Are Able to Incorporate and Influence Gene Expression in a Pre-Formed *Gardnerella vaginalis* Biofilm. *Pathogens* **10**, 247 (2021).
357. McEwen, S. A. & Collignon, P. J. Antimicrobial Resistance: a One Health Perspective. *Microbiol Spectr* **6**, 2 (2018).
358. Ferri, M., Ranucci, E., Romagnoli, P. & Giaccone, V. Antimicrobial resistance: A global emerging threat to public health systems. *Crit Rev Food Sci Nutr* **57**, 2857–2876 (2017).
359. Christaki, E., Marcou, M. & Tofarides, A. Antimicrobial Resistance in Bacteria: Mechanisms, Evolution, and Persistence. *J Mol Evol* **88**, 26–40 (2020).
360. Brinkac, L., Voorhies, A., Gomez, A. & Nelson, K. E. The Threat of Antimicrobial Resistance on the Human Microbiome. *Microb Ecol* **74**, 1001–1008 (2017).
361. Baron, S. A., Diene, S. M. & Rolain, J.-M. Human microbiomes and antibiotic resistance. *Hum Microbiome J* **10**, 43–52 (2018).

Appendix

APPENDIX 1

Optimization of culturomic protocol[1] (**Table 5**) was based on three main steps:

1) modification of culture media and size of Petri dishes to facilitate capturing morphologically different bacteria from urine samples

- protocol included chromogenic agar supplemented with 2% (w/v) gelatin, 0.5% (w/v) yeast extract, 0.1% (w/v) starch, 0.1% (w/v) glucose which was previously described to improve growth of *Gardnerella* sp.[2] and with 0.1% (v/v) Tween 80 which is beneficial for growth of most *Lactobacillacea*[3]
- size of Petri dishes was increased from commonly used 90 mm diameter plates to 140 mm diameter plates.

2) characterization of enhanced number of isolates to improve bacterial diversity captured from the samples

- up to 5 isolates of the same morphotype were isolated, including all morphotypes possible to be visually identified.

3) improvement of identification methods to provide reliable identification at species level

- Sanger sequencing of specific marker genes including 16S rRNA gene, *pheS* for *Lactobacillaceae*, *cpn60* for *Gardnerella*, *rpoB* for *Acinetobacter* and *Staphylococcus*, *recN* for *Citrobacter*
- additional identification by species specific PCRs namely *dltS* for *Streptococcus agalactiae*, *sodA* for *Enterococcus faecalis*, *malB* for *Escherichia coli*

Table 5. The modified culturomic protocol used in all tested samples.

Collection method	Urine volume	Culture media	Atmospheric condition	Temperature	Incubation time	Isolate identification
MSU	100 µl	BAP, sCAP, FTM	aerobic	37°C	48 h	MALDI TOF-MS, sequencing of marker genes, species specific PCR, complementary laboratory identification based on Gram-staining or biochemical tests
	100 µl	BAP, sCAP	microaerophilic (5% CO_2)	37°C	48 h	
	100 µl	BAP	anaerobic	37°C	48 h	

MSU - midstream urine, BAP - Blood agar plate, sCAP - supplemented Chromogenic agar plate, FTM - Fluid Thioglycollate Medium, MALDI-TOF MS - matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, PCR - polymerase chain reaction.

Appendix 1 - References

1. Hilt EE. *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* **52,** 871-876 (2014).
2. Alves P, Castro J, Sousa C, Cereija TB, Cerca N. *Gardnerella vaginalis* outcompetes 29 other bacterial species isolated from patients with bacterial vaginosis, using in an in vitro biofilm formation model. *J Infect Dis* **210**, 593-596 (2014).
3. Man JCD, Rogosa M, Sharpe ME. A Medium for the Cultivation of Lactobacilli. *J Appl Microbiol* **23**, 130-135 (1960).