U.PORTO

PHD IN PHARMACEUTICAL SCIENCES
MICROBIOLOGY SPECIALITY

Exploring the Diversity and Features of Gardnerella spp.: Implications for Interactions in the Female Urogenital Microbiome and the Host

Márcia Daniela Rocha de Sousa

D

2025

U.PORTO

Márcia Sousa, Exploring the Diversity and Features of Gardnerella spp.: Implications for Interactions in the Female Urogenital Microbiome and the Host

D.FFUP 2025

Exploring the Diversity and Features of Gardnerella spp.: Implications for Interactions in the Female Urogenital Microbiome and the Host

Márcia Daniela Rocha de Sousa

Faculdade de Farmácia

D
2025

U.PORTO

PHD IN PHARMACEUTICAL SCIENCES

MICROBIOLOGY SPECIALITY

Exploring the Diversity and Features of *Gardnerella* spp.: Implications for Interactions in the Female Urogenital Microbiome and the Host

Márcia Daniela Rocha de Sousa

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

February, 2025

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

Márcia Daniela Rocha de Sousa

This work was performed at the BacTDrugs Research Lab, at the Microbiology Laboratory, Biological Sciences Department of the Faculty of Pharmacy, University of Porto (FFUP, Porto, Portugal). 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. The author received financial support through an individual doctoral fellowship funded by the Foundation for Science and Technology (FCT) of Portugal with the reference 2020.05038.BD.

Acknowledgements

Quero expressar a minha profunda gratidão à minha orientadora de doutoramento, Dra. Filipa Grosso, cujo apoio inabalável e orientação foram fundamentais para a conclusão bem-sucedida deste doutoramento. Estou igualmente muito grata às minhas coorientadoras, Professora Dra. Luísa Peixe e Dra. Teresa Gonçalves Ribeiro, pela dedicação incansável, paciência e entusiasmo ao longo da nossa colaboração diária, que foram essenciais para a qualidade da minha investigação.

Além disso, quero expressar o meu apreço à Dra. Ângela Novais, Dra. Ana Freitas, Professora Carla Novais, Professora Patrícia Antunes e Bárbara Duarte pelo aconselhamento profissional, apoio e encorajamento. Estou também grata pelo apoio dos técnicos Cristina Pinto da Costa, Daniel Nunes e Joana Fernandes, do Laboratório de Microbiologia da Faculdade de Farmácia.

À Fundação para a Ciência e a Tecnologia (FCT) de Portugal pelo financiamento (referência SFRH/BD/05038/2020) e pela oportunidade de dar continuidade a este projeto de doutoramento.

Estou profundamente grata pelo aos meus colegas de bancada, especialmente à Ana Beatriz, pelas discussões científicas, compreensão e amizade, e ao Mike, pela amizade, apoio e companheirismo. Agradeço à Lili e à Rita, que, mesmo chegando mais tarde ao laboratório, se tornaram amigas tão especiais. O vosso apoio foi inestimável e sou eternamente grata pela amizade, carinho e companheirismo que me proporcionaram ao longo desta jornada académica.

Estou imensamente grata a todos os meus amigos “fora do laboratório”, que sempre me apoiaram ao longo deste percurso. Agradeço especialmente às minhas amigas e companheiras de aventuras, Mariana, Dani, Cé, Catarina e Cátia! Sem vocês a minha vida não seria tão divertida!

Por fim, quero agradecer à minha família, especialmente aos meus pais e ao meu irmão, pelo apoio e incentivo, dia após dia. Palavras não chegam para agradecer tudo o que fazem e fizeram para que eu pudesse seguir os meus sonhos.

A todos que, de alguma forma, contribuíram para a realização desta tese, deixo o meu mais sincero agradecimento.

Abstract

For decades, *Gardnerella vaginalis* was the sole recognized species of its genus, mainly associated with bacterial vaginosis (BV), a condition with significant health implications for women of reproductive age. Recent taxonomic revisions have expanded it to four named species (*Gardnerella vaginalis*, *Gardnerella piotii*, *Gardnerella swidsinskii* and *Gardnerella leopoldii*) and ten genomic species, with *Gardnerella* now detected in BV, lower urinary tract disorders (e.g., urgency urinary incontinence, overactive bladder), and asymptomatic women.

This thesis investigates *Gardnerella* species- and strain-specific traits shaping their roles in the urogenital microbiome, focusing on virulence, metabolism, and antimicrobial resistance (AMR).

The bacterial collection analyzed comprised 33 *Gardnerella* isolates obtained from urine and vaginal swabs during a previous Female Urogenital Microbiome (FUM) study conducted by our research team. These isolates were used for various phenotypic assays, including antimicrobial susceptibility, enzymatic activity, and inhibition tests. Additionally, 120 publicly available genomes were incorporated to support comprehensive genomic analyses.

This thesis presents the most comprehensive *Gardnerella* pangenome analysis to date, revealing a highly dynamic accessory genome that contributes to differences in virulence, metabolism, and antimicrobial resistance (AMR). Of note, 57% of public genomes remain misclassified, underscoring the need for whole-genome sequencing (WGS) over 16S rRNA or *cpn60* sequencing methods. Two novel species, *Gardnerella pickettii* and *Gardnerella greenwoodii* were characterized, refining *Gardnerella* taxonomy.

Virulence profiling identified NanH1 and NanH3 as prevalent sialidases, NanH2 as dysbiosis-linked, and a novel NanH4 enzyme, with NanH3's inhibition by DANA/Zanamivir suggesting therapeutic potential. Vaginolysin (VLY) occurred in 89.1% of strains, with a new Type 3B variant and five genetic platforms indicating diverse evolutionary origins.

We report the first neopullulanase and the presence of various carbohydrate-degrading enzymes and transporters, such as α-amylase and MalXFGK respectively, which emphasizes the genus's role in glycogen degradation and uptake, critical for adapting to the urogenital microbiome and influencing resource availability for coexisting species.

AMR analysis detected *lsaC* as the dominant ARG, in addition to *mefA, msrD, aph(3')-Ia, tetM* and *ermX,* yet MIC values for antibiotics tested were low for most strains. Elevated MICs to metronidazole and vancomycin in few isolates could not be explained by known resistance mechanisms, suggesting the involvement of novel genetic or regulatory factors.

Integrating WGS and phenotypic data from symptomatic and asymptomatic isolates, this study provides new insights into *Gardnerella* taxonomy, virulence, metabolism, and antimicrobial resistance, laying the foundation for improved diagnostic, treatment strategies, and microbiome research. Urging future focus on sialidase inhibition, metabolic competition, and resistance profiling will provide further advances in optimizing infection prevention and therapeutic approaches.

Keywords: *Gardnerella*; Microbiota; Pangenome; Virulence factors; Metabolic adaptation

Resumo

Durante décadas, *Gardnerella vaginalis* foi a única espécie reconhecida do seu género, sendo sobretudo associado a vaginose bacteriana (VB), uma condição com implicações significativas para a saúde de mulheres em idade reprodutiva. Revisões taxonómicas recentes expandiram este género para quatro espécies (*Gardnerella vaginalis, Gardnerella piotii, Gardnerella swidsinskii* e *Gardnerella leopoldii*) e dez genomospécies. Atualmente, *Gardnerella* é detetada não só na VB, mas também em perturbações do trato urinário inferior (por ex., incontinência urinária de urgência, bexiga hiperativa) e em mulheres assintomáticas.

Esta tese explora características específicas de espécies e estirpes de *Gardnerella* que moldam o seu papel no microbioma urogenital, com foco na virulência, metabolismo e resistência a antimicrobianos (AMR).

A coleção bacteriana analisada incluiu 33 isolados de *Gardnerella* obtidos a partir de amostras de urina e exsudados vaginais durante um estudo anterior do Microbioma Urogenital Feminino (*Female Urogenital Microbiome* – FUM), conduzido pela nossa equipa de investigação. Estes isolados foram submetidos a vários ensaios fenotípicos, incluindo suscetibilidade antimicrobiana, atividade enzimática e testes de inibição. Adicionalmente, foram incorporados 120 genomas disponíveis publicamente para suportar as análises genómicas efetuadas.

Esta tese apresenta a análise mais abrangente do pangenoma de *Gardnerella* até à data, revelando um genoma acessório altamente dinâmico que contribui para diferenças na virulência, metabolismo e resistência a antimicrobianos. É de salientar que 57% dos genomas públicos continuam a ser incorretamente classificados, destacando a necessidade de sequenciação genómica total (*whole-genome sequencing* – WGS) em detrimento dos métodos baseados na sequenciação dos genes 16S rRNA ou *cpn60*. Duas novas espécies, *Gardnerella pickettii* e *Gardnerella greenwoodii*, foram caracterizadas, consolidando a taxonomia do género *Gardnerella.*

O perfil de virulência identificou NanH1 e NanH3 como as sialidases mais prevalentes, NanH2 sobretudo associada a disbiose, e uma nova enzima NanH4. Adicionalmente, a inibição de NanH3 por DANA/Zanamivir sugere um potencial alvo terapêutico. A vaginolisina (VLY) foi detetada em 89,1% das estirpes, com a identificação de uma nova variante do Tipo 3B, e cinco plataformas genéticas distintas, indicando origens evolutivas diversas.

Identificámos a primeira ocorrência de neo-pululanase em *Gardnerella*, bem como a presença de várias enzimas e transportadores envolvidos na degradação de hidratos de carbono, como α-amilase e MalXFGK, respetivamente. Estes resultados destacam o papel de *Gardnerella* na degradação e captação de glicogénio, um processo essencial para a adaptação ao microbioma urogenital, que também contribui para a disponibilidade de recursos para as espécies coexistentes.

A análise de genes de resistência a antimicrobianos identificou *lsaC* como o mais frequente, juntamente com *mefA, msrD, aph(3')-Ia, tetM e ermX*. No entanto, os valores de MIC para os antibióticos testados foram baixos para a maioria dos isolados. Foram observadas concentrações mínimas inibitórias elevadas para metronidazol e vancomicina em alguns isolados, sem explicação através dos mecanismos de resistência conhecidos, sugerindo o envolvimento de novos fatores genéticos ou reguladores.

Ao integrar dados de WGS e fenotípicos de isolados de dadoras sintomáticas e assintomáticas, este estudo fornece novos conhecimentos sobre a taxonomia, virulência, metabolismo e resistência a antimicrobianos em *Gardnerella*, estabelecendo bases para possíveis avanços no diagnóstico, tratamento e investigação do microbioma. O foco futuro na inibição das sialidases, na competição metabólica e no perfil de resistência poderá impulsionar avanços na prevenção e no tratamento de patologias associadas a *Gardnerella*.

Palavras-chave: *Gardnerella,* Microbiota, Pangenoma, Fatores de virulência, Metabolismo

Table of Contents

List of Boxes

List of Tables

List of Abbreviations

ARG: Antimicrobial resistance gene

AMR: Antimicrobial Resistance

ANI: Average Nucleotide Identity

ANIb: Average Nucleotide Identity based on blast

AAI: Average amino acid Identity

ARG: Antimicrobial Resistance Gene

BV: Bacterial Vaginosis

CFU: Colony Forming Unit

CLSI: Clinical and Laboratory Standards Institute

COG: Cluster of Orthologous Group

CST: Community State Types

dDDH: digital DNA-DNA hybridization

DNA: DeoxyriboNucleic Acid

ECOFF: Epidemiological *cut-off*

EUCAST: European Committee on Antimicrobial Susceptibility Testing

EQUC: Enhanced Quantitative Urine Culture

FUM: Female Urogenital Microbiome

GC content: Guanidine + Cytosine content

HGT: Horizontal Gene Transfer

HIV: Human Immunodeficiency Virus

HMP: Human Microbiome Project

HPV: Human Papillomavirus

LUTD/LUTS: Lower Urinary Tract Disorders/ Lower Urinary Tract Symptoms

MALDI-TOF MS: Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry

Mbp: Mega base pairs

MIC: Minimum inhibitory concentration

MSU: Midstream Urine

NCBI: National Center for Biotechnology Information

OAB: overactive bladder

OABSS: Overactive Bladder Symptom Score

POP: Pelvic Organ Prolapse

RA: Relative Abundance

rRNA: ribosomal RNA

SPA: SupraPubic bladder Aspiration

STI: Sexually Transmitted Infection

SUC: standard urine culture

SUI: Stress Urinary Incontinence

tRNA: transfer RNA

PUS: Peri-Urethral Swab

RA: Relative Abundance

rRNA: Ribosomal Ribonucleic Acid

rUTI: recurrent Urinary Tract Infection

TNA: Tetra-Nucleotide Analysis

TUC: TransUrethral Catheter

UI: Urinary Incontinence

UT: Urotype

UTI: Urinary Tract Infection

UUI: Urgency Urinary Incontinence

VS: Vaginal Swab

WGS: Whole-Genome Shotgun sequence

WHO: World Health Organization

1. Introduction

Gardnerella has long been associated with bacterial vaginosis (BV), a condition with significant health implications for women of reproductive age (1,2). However, recent studies on the female urogenital microbiome (FUM) have identified various *Gardnerella* species and genomic species not only in BV-positive women and those with urinary tract infections (UTIs) or lower urinary tract disorders/syndromes (LUTD/LUTS), such as overactive bladder (OAB) and urgency urinary incontinence (UUI), but also in healthy cohorts (3–5).

For seven decades, the genus *Gardnerella* was thought to contain only a single species, *Gardnerella vaginalis* (6). However, recent research identified novel species, highlighting the genus genetic and functional diversity (7–9). In recent years, the taxonomic classification of the *Gardnerella* genus has undergone significant expansion and refinement. In 2019, comparative genomics and protein signatures analyses allowed an amended description of *G. vaginalis*, recognition of three new species (*Gardnerella piotii, Gardnerella swidsinskii* and *Gardnerella leopoldii*), and 9 genomic species (10). Later, a 10th *Gardnerella* genomic species was proposed (11,12).

These findings reinforce the possibility that *Gardnerella*'s role in pathology may be species or even strain-dependent and arises from a complex interplay of genetic diversity, host responses, and interactions within the FUM, underscoring the need to investigate specific virulence features, antibiotic resistance genes, and other genetic factors that may drive the potential pathogenicity of *Gardnerella* (13).

Nevertheless, the distribution and abundance of different *Gardnerella* species in FUM of symptomatic and asymptomatic women is barely explored. Only a few recent studies have examined bacterial compositions that include the newly described *Gardnerella* species (14,15).

The following sections provide a comprehensive review of current knowledge regarding *Gardnerella´*s complex role in the urogenital tract, integrating recent findings at species-level diversity, metabolic and virulence potential, interactions within the FUM, and their implications for female urogenital health and disease.

It is important to mention that most papers published before 2019, and referring only to "*Gardnerella*" or "*G. vaginalis*", could contain inaccurate species information taking into consideration the recent taxonomic restructuration of the genus. To address this limitation

and ensure clarity, the term "*Gardnerella*" will be used throughout this work to collectively refer to the *Gardnerella* species reported in studies that did not use the terminology proposed by Vaneechoutte in 2019 (10). Whenever possible, we will convert any classifications by ecotype, group, or clade, into the corresponding updated taxonomic structure.

1.1 Female Urogenital Microbiome

The urogenital tract encompasses the organs and structures involved in both urinary and reproductive functions, reflecting their proximity, common embryological origin, and shared pathways. In this study, the term "urogenital" specifically refers to the female structures of the bladder, urethra, and vagina, highlighting the interconnection between them. Consequently, the interpretation of results obtained from urine samples collected by the midstream voided urine (MSU) method must consider that the detected microbes may originate from both the urinary and genital tracts (16,17). For female donors, self-collection methods like MSU and vaginal swabs are optimal. MSU collection minimizes contamination from urethral bacteria and debris by discarding the initial portion of the urine stream before collecting a midstream sample in a sterile container. This non-invasive and generally painless technique provides a sample more representative of bladder urine, making it suitable for both clinical and research settings. Vaginal swabs, similarly, minimally invasive and causing little to no discomfort, offer a direct means of sampling the diverse microbiota within the vaginal canal.

Besides the importance of understanding how samples are collected, it is also essential to define the key terms used in microbiome research to accurately interpret findings from urine and vaginal samples.

According to the currently accepted definitions, the term “microbiota” refers to all living microorganisms (bacteria, archaea, fungi, small protists and algae) present in a defined environment (18). As phages, viruses, plasmids, prions, viroids, and free DeoxyriboNucleic Acid (DNA) are not considered as living microorganisms, they are not considered in the microbiota (18). The definition of “microbiome” includes the community of the microorganisms (microbiota), their structural elements (nucleic acids, proteins, lipids, polysaccharides, metabolites as signaling molecules, toxins, organic, and inorganic molecules) and environmental conditions such as molecules produced by coexisting microorganisms and host and controlled by the surrounding environmental conditions (18).

Despite representing distinct concepts, the terms “microbiota” and “microbiome” are frequently used interchangeably in practice.

Launched in 2008 by the National Institutes of Health (NIH), the Human Microbiome Project (HMP) aimed to characterize the microbiome across various body sites, including the gastrointestinal tract, oral cavity, vagina, skin, and nasal cavity, using culture-independent methods (19). Initially, the urinary tract was excluded, as urine from healthy individuals was traditionally considered sterile (20). This misconception persisted largely due to the limitations of Standard Urine Culture (SUC), which failed to detect low-abundance or fastidious uropathogens (21). The detection of microbial DNA in asymptomatic female urine samples through 16S rRNA amplicon sequencing challenged the sterility paradigm (22,23). However, this method cannot differentiate between live, dead, or lysed bacteria (24). The Expanded Quantitative Urine Culture (EQUC) protocol addressed this limitation by incorporating larger urine volumes, diverse atmospheric conditions, and extended incubation times, allowing the cultivation of previously undetected bacteria (25). EQUC findings confirmed that sequencing-identified DNA corresponded to viable bacterial populations.

Recent studies have identified *Lactobacillus*, *Gardnerella*, and *Streptococcus* as key genera within the FUM, often in combination with other genera such as *Corynebacterium*, *Escherichia*, and *Staphylococcus* (15,21,25). Moreover, alterations in microbial composition have been linked to urinary tract disorders. For instance, *Gardnerella* abundance is higher, while *Lactobacillus* levels are lower in women with urgency urinary incontinence (UUI) compared to those without (26). Using 16S rRNA sequencing, nine genera (*Actinobaculum*, *Actinomyces*, *Aerococcus*, *Arthrobacter*, *Corynebacterium*, *Gardnerella*, *Oligella*, *Staphylococcus*, and *Streptococcus*) were more frequently detected in UUI urine samples using the EQUC method (26). Similarly, *Bifidobacterium*, *Enterococcus*, *Lactobacillus*, and *Streptococcus* were less prevalent in overactive bladder (OAB) patients, whereas *Actinomyces* and *Corynebacterium* were more common (21).

Evidence suggests that FUM can be dynamic and fluctuate in each person, over time and with age (15,25). *Gardnerella* and *Lactobacillus* are more common in reproductive aged women, whereas *Escherichia* and *Mobiluncus* are more common in older, more likely to be postmenopausal females (27–29). Longitudinal studies have confirmed microbial shifts within individuals over time. For example, urine microbiota sampled 2.5 years apart revealed conversions between bacterial taxa, such as *Lactobacillus jensenii* shifting to *Lactobacillus crispatus*, or *Gardnerella swidsinskii* and *Fannyhessea vaginae* (formerly

Atopobium vaginae) transitioning to *Gardnerella vaginalis*, *Bifidobacterium* spp., *Cutibacterium avidum* and *Citrobacter koseri* converting to *Escherichia coli* (15).

Regarding the vaginal microbiota, cultivation-independent methods have shown that vaginal bacterial communities can be clustered into anywhere from three to nine discrete groups, most of which dominated by lactobacilli (30). Another widely used classification method to characterize the vaginal microbiota was proposed by Ravel *et al*. as community state types (CST), according to the dominance of a bacterial species (31). Therefore, the vaginal microbiota can be divided into 5 CSTs: CST I associated with a dominance of *L. crispatus*; CST II with *Lactobacillus gasseri*, CST III with *Lactobacillus iners;* CST V with *L. jensenii*; and CST IV was heterogeneous and typified by a higher proportion of obligate anaerobic bacteria, including *F. vaginae*, *Gardnerella*, and *Prevotella* spp. (31). CST terminology is helpful to compare different groups and to analyze fluctuations among cohorts. In ethnic studies applying CST terminology it was observed that African and Hispanic women were more likely to harbor CST IV (*Gardnerella* most abundant), in Asian women CST III appeared as the most common group, while Caucasian women had a higher prevalence of CST I and CST V (31,32). Further investigation of the vaginal microbiota using longitudinal study design has shown that vaginal communities are dynamic and capable of rapid shifts, although in many women, the microbiota is fairly stable (30). Emerging evidence appears to show that CST I tends to be most stable and to promote vaginal community stability (33), whereas CST IV appears to frequently transition to multiple other states (34). The CST framework undergoes continuous refinement, as evidenced by advancements such as the VAginaL community state typE Nearest Centro/d clAssifier (VALENCIA) study (35). This innovation provides a crucial solution for enhancing the reliability and consistency of assigned vaginal community state types. It also enables robust comparisons across datasets and supports meta-analyses that integrate data from diverse sources. Significant evidence now indicates that the vaginal microbiota dominated by *Lactobacillus* species is optimal for vaginal health, particularly *L. crispatus*, which was strongly correlated with the absence of BV (30), and women with a dominant population of vaginal lactobacilli are at a lower risk of UTI compared to women with more diverse microbiota, consisting of Gram-negative anaerobes, *Actinobacteria*, and *Firmicutes* (3).

In addition, further evidence of an interconnected urogenital microbiota has emerged from whole-genome phylogenetic analyses and 16S rRNA amplicon sequencing, comparing bacterial strains from urine and vaginal samples within the same individual (16,36). These studies identified highly similar *E. coli*, *L. iners*, *L. crispatus*, and *Streptococcus anginosus* strains in paired vaginal and urine samples (16). Moreover, *Lactobacillus* was the most

abundant genus in both urine and vaginal samples from healthy women (36), reinforcing the concept of a shared microbial ecosystem between the vagina and urinary tract.

1.2 *Gardnerella*: a Relevant Member of the Urogenital Microbiome

The role of the *Gardnerella* genus in the FUM remains a subject of debate, as it has been identified in both asymptomatic and symptomatic women (15,26,27). Defining a "core" healthy or unhealthy urogenital microbiome is also particularly challenging due to the dynamic nature of this niche, which is influenced by numerous factors, including age, physiological conditions, menstrual cycle, medical treatments, and lifestyle (15,25,29,37). In this section, we will explore *Gardnerella*'s role in both health and disease conditions, exploring the key features that enable its adaptation within the urogenital microbiome.

1.2.1 History of the Genus *Gardnerella*

Gardnerella genus belongs to the order *Bifidobacteriales*, and family *Bifidobacteriaceae* (38). During almost 70 years, *G. vaginalis* was the only species reported within the genus. *G. vaginalis* was first isolated in 1953 from cervix swabs of women and urine of men and characterized by Leopold who described a small, non-motile, non-capsulated, pleomorphic Gram-negative rod, with slow growth and capacity to form tiny colonies surrounded by zones of haemolysis (6).

In 1955, Gardner and Dukes proposed the name *Haemophilus vaginalis* for the bacterium they believed to be the primary cause of what was then called “nonspecific vaginitis” (now known as BV) (39). *H. vaginalis* was widely accepted and remained unquestioned until 1963, when Zinnemann and Turner concluded that *H. vaginalis* was Gram-positive, probably a *Corynebacterium*, and proposed the specific name *Corynebacterium vaginale* based on metabolic features and Gram staining (40). The misclassification possibly relied on the unusual thinness of the bacterium cell wall (8-12 nm), which may lead to inconsistent Gram staining results, as the peptidoglycan layer is too thin to retain the crystal violet stain effectively, but still differs from typical Gram-negative bacteria in its composition and structure: which was confirmed by electron microscopy, absence of diaminopimelic in the peptidoglycan and the lack of lipopolysaccharide in the cell wall (10,41-43). Still, nowadays this bacterium is often classified as Gram-variable, due to its biochemistry and cell wall ultrastructure, resulting in variations between negative and positive Gram staining reaction (44). Meanwhile, many authors used the new nomenclature, *C. vaginale,* despite the accumulating evidence that the organism was not a *Corynebacterium* (45).

In 1980, Greenwood and Pickett proposed the new genus *Gardnerella* (Gard,ner.el'la. M.L. dim. ending -ella; M. L. fem. n. *Gardnerella*; named after H. L. Gardner) for inclusion of organisms previously designated as either *H. vaginalis* or *C. vaginale* based on Adansonian classification and biochemical analysis, DNA-DNA hybridizations (DDH), biochemical analysis of cell envelopes, and electronic microscopy (46) (**BOX 1**). At that time, the genus was not assignable to a family because of the unusual cell wall (46). The genus is restricted to facultatively anaerobic bacteria that use carbohydrates as their major energy source and includes catalase and oxidase negative bacteria with laminated cell walls that produce acetic acid as the major final product of fermentation (46,47). The type species of the genus *Gardnerella* is *G. vaginalis* (vag.in.al'is. L. adj. *vaginalis,* of the vagina), which is a fastidious organism, requiring biotin, folic acid, niacin, thiamine, riboflavin, and two or more purines-pyrimidines for growth. It grows on sheep blood agar, Thayer-Martin agar, or 0.01% tellurite agar. It is not able to reduce nitrates and is indole- and urease-negative. Colonies are round, opaque, and smooth with β-haemolysis on human and rabbit blood agars, but not in sheep blood (46). In 1997, the genus *Gardnerella* was finally assigned to the family *Bifidobacteriaceae* (38).

Although through the years several classification schemes (e.g. biochemical tests, molecular methods based on 16S rRNA gene or *cpn60,* concatenated sequences of 473 genes common to a set of 17 *Gardnerella* isolates) have been proposed in an attempt to delineate subgroups within *Gardnerella* (7,8), *G. vaginalis* was the only species recognized until 2019, when Vaneechoutte and colleagues proposed an emended description of *G. vaginalis*, introduced three new species; *Gardnerella piotii, Gardnerella leopoldii,* and *Gardnerella swidsinskii*, and nine genomic species (10). This proposal was based on Whole-Genome Shotgun (WGS) sequence comparisons (digital DNA-DNA hybridization, dDDH; and average nucleotide identity, ANI), matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis, and biochemical properties (10). The species could not be delineated using full-length 16S rRNA gene sequences (10). *In silico* analysis based on core genome analysis and ANI comparisons pointed out for a putative 10th genomic species (12). Nevertheless, Potter and colleagues suggested that depending on the bioinformatic tools used, the number of *Gardnerella* genomic species differ (12 with ANI, 14 with average nucleotide identity based on blast - ANIb, 8 with average amino acid identity - AAI and 9 with correlation indexes of tetra-nucleotide signatures - Tetra) (11) (**BOX 1**). Analyses using various bioinformatic tools have grouped *G. leopoldii* and *G. swidsinskii* together, suggesting they may represent the same species (11). In support of this, Bulavaitė *et al.,* (48) demonstrated that MALDI-TOF MS could only

differentiate between the phylogenetically diverse groups composed of *G. vaginalis/ G. piotii* and *G. leopoldii/ G. swidsinskii*, failing to resolve *G. leopoldii* and *G. swidsinskii* into distinct species. This contrasts with the study by Vaneechoutte and colleagues, who reported a set of peaks (single and pairs) obtained by MALDI-TOF MS analysis that could be used to distinguish the four *Gardnerella* species (10).

In **Table 1** are the main characteristics of *Gardnerella* species, according to Vaneechoutte *et al.,* (10).

Table 1: Main characteristics of *Gardnerella* species according to Vaneechoutte *et al.,* (10).

Methodology	**Species**			
	G. vaginalis	***G. leopoldii***	***G. piotii***	***G. swidsinskii***
Gram	Gram-positive to Gram-variable coccobacilli			
Genome size (Mbp)	1.66 - 2.01	1.47 - 1.56	1.51 - 1.52	1.59 - 1.62
Size (µm)	0.5 X 1.5			
Growth	Chocolate agar at CO_2-enriched conditions (5 %) for 24-120 hours			
Morphology	Pinpoint white to greyish with a smooth surface			
β-galactosidase	-	-	+	-
Sialidases	+/-	-	+	-
DNA GC content (mol%)	41.0 - 42.8	41.9 - 43.2	41.1 - 42.3	41.4 - 42.3

Abbreviations - DNA: DeoxyriboNucleic Acid, GC: Guanine and Cytosine content, +: positive, - : negative.

BOX 1. Genomic methodologies implemented to classify *Gardnerella* strains

Adansonian: Also known as numerical taxonomy or phenetics, is a method of taxonomic classification in which organisms are grouped together based on their overall similarity across a large number of equally weighted traits or characteristics (49). The more features two organisms have in common, the more likely they are to be classified together, regardless of the evolutionary significance of those traits**.**

DDH: A molecular biology technique that measures the genetic similarity between DNA sequences. The melting temperature of a double-strand depends on the degree of matching base pairings between both strands; genomic (dis)similarity can be inferred from the melting temperature. DDH values ≤70% are considered as an indication that the tested organism belongs to a different species than the type strain(s) used as reference(s) (50).

dDDH: An *in silico* method used to estimate genome-to-genome similarity by comparing sequenced genomes, serving as a more accurate and reproducible alternative to traditional DDH for bacterial species delineation. It calculates pairwise genome distances and provides a similarity percentage, facilitating species classification and taxonomic studies with greater precision (50,51).

AAI: A pairwise genome comparisons and averaging the sequence identities of shared orthologous genes (amino acid). An AAI value of >90% was determined to include *Gardnerella* strains in the same species (11). AAI may be suitable for comparisons of distant genomes rather than ANI, because the latter may contain more evolutionary noise (52). A score of >90% indicates that *Gardnerella* strains belong to the same species (11).

Accessory genome: Genes not present in all strains of a given taxa. Large differences in the accessory gene contents were observed for six *Gardnerella* genomic species (11).

ANI: A pairwise genome comparisons and averaging the sequence identities of shared orthologous genes (nucleotides). Typically, a *cut-off* score of >95% indicates that they belong to the same species (53). An ANI value of >96% was determined to include *Gardnerella* strains in the same species (10).

ANIb: ANIb refers to the calculation of ANI between genomic sequences using the BLAST algorithm. This method compares homologous genomic regions to assess genetic relatedness, providing a percentage identity score for species delineation.

Core genome: Genes present in 100% strains at 70% nucleotide identity (11). Potter *et al.,* observed that the genomic species can be distinguished by the similarity of their 200 core genes (11).

MALDI-TOF MS: The sample is co-crystallized with the matrix on the sample target, to be desorbed and ionized by the MALDI ion source (e.g. ultraviolet laser). The ion molecules, including the microbial peptides/proteins, are accelerated by the electric field into the TOF analyser. All the ions are separated by TOF in accordance with the *m/z* ratio and a mass spectrum. Individual mass peaks are used for microorganism identification and provide valuable information for the fingerprinting of bacteria (54). In the study conducted by Vaneechoutte *et al.,* peaks at *m/z* 4422/4429, 5162, 6855/6885, and 8842/8857 differentiate *G. vaginalis* from *G. piotii*. A single peak at *m/z* 2704 was proposed to discriminate *G. leopoldii* from *G. swidsinskii*. The absence or presence of a single peak at *m/z* 5349 and the peak pairs at *m/z* 4849/4928 and 9795/9853 differentiate *G. vaginalis/G. piotii* and *G. leopoldii/G. swidsinskii* (10).

Tetra-Nucleotide Analysis (TNA): A tetra-nucleotide is a sequence of four DNA bases (e.g., AGTC, CTTA). Tetra-nucleotide Analysis (TNA) calculates a correlation coefficient between the frequency patterns of these four-base sequences in different DNA samples. This coefficient, ranging from 0 to 1, serves as an indicator of genetic relatedness between sequences. A value of 1 indicates identical tetra-nucleotide usage patterns, typically seen when comparing a genome to itself, while lower values suggest less relatedness. TNA is a valuable tool in comparative genomics for assessing similarities between DNA sequences without requiring full sequence alignment (53).

1.2.2 *Gardnerella* in the Healthy Urogenital Microbiota

New studies suggest that *Gardnerella*'s presence may reflect varying bacterial interactions and host state. In healthy women of reproductive age, *Gardnerella* often coexists with other commensal bacteria, particularly *Lactobacillus*, which usually dominates the female urogenital microbiome (15,55,56). Several species of the *Lactobacillaceae* family*,* particularly *L. crispatus,* help prevent pathogen growth by maintaining a low vaginal pH through lactic acid production and producing antimicrobial compounds like hydrogen peroxide and bacteriocin-like substances (57,58). This coexistence suggests that *Gardnerella* may also play a role in maintaining the ecosystem's balance. However, the relationship between *Gardnerella* and urogenital health is complex, with different species potentially contributing to either healthy or unhealthy conditions.

The urogenital microbiome is dynamic and can change due to various factors throughout life in response to age, health conditions, sexual activity, and during menstruation and pregnancy, thus some differences have been reported regarding the presence of *Gardnerella* in urine and vagina (15,59–61). For instance, during menstruation, the vaginal pH rises, leading to a decrease in *Lactobacillus* populations and an increase of other facultative and anaerobic species within the microbiome. Post-menstruation, the decline in vaginal pH promotes *Lactobacillus* recovery while suppressing the growth of facultative and anaerobic bacteria (62). Moreover, after childbirth, the vaginal microbiome in women experiences a decrease in *Lactobacillus* and a rise in diverse anaerobes, often lasting up to a year (34,44). Interestingly, women vaccinated against human papillomavirus (HPV) have been observed to have reduced *Gardnerella* populations, although species-specific associations remain to be elucidated (61).

Several *Gardnerella* species have been detected in healthy urogenital microbiomes, including *G. vaginalis, G. leopoldii, G. swidsinskii,* and *G. pickettii* (15,63). Particularly, in the healthy vagina have been identified *G. piotii*, with occasional presence of *G. vaginalis, G. leopoldii* and *G. swidsinskii.* Other microorganisms commonly found alongside *Gardnerella* in healthy FUM include *Lactobacillus* spp. (namely *L. crispatus, L. jensenii* or *L. gasseri*), *Prevotella, Gemella, F. vaginae, Bifidobacterium* spp., *Corynebacterium*, and *Streptococcus anginosus* (21,26,31,64,65). **Table 2** presents a summary of studies reporting the detection of *Gardnerella* in healthy FUM.

These findings underscore the complex nature of *Gardnerella* in the urogenital microbiome. The presence of this genus is not a simple indicator of pathology but rather a nuanced

component of a sophisticated microbial ecosystem. Researchers emphasize the critical need to continue exploring the potential roles of specific *Gardnerella* species and to understand the complex microbial interactions that contribute to urogenital health.

Table 2: Summary of studies reporting the identification of *Gardnerella* and their main findings on healthy female urogenital microbiome.

Population	Sample collection method	Methodology used for *Gardnerella* identification	Main findings	Reference
Non-pregnant reproductive aged women	VS	16S rRNA amplicon sequencing	• *Gardnerella* is present in co-occurrence with *Cryptobacterium, Gemella*, *Aerococcus, Prevotellaceae, Ruminococcaceae* and *Anaeroglobus* • Vaginal microbial communities in a state of dynamic equilibrium and resilience • Difference between vaginal microbiota of Asian, Caucasian, African and Hispanic women	Ravel *et al.,* 2011 (31)
Non-pregnant reproductive aged women	MSU	16S rRNA amplicon sequencing	• *Lactobacillus*, *Prevotella* and *Gardnerella* are the most dominant • Asymptomatic urine microbiome is characterized by complex microbial composition with intra-individual variations	Siddiqui *et al.,* 2011 (65)
Non-pregnant reproductive aged women, Menopausal** women	VS	16S rRNA amplicon sequencing, *cpn60* sequencing	• CSTs dominated by *Lactobacillus* spp. • Difference between vaginal microbiota of Asian and Caucasian women • *Gardnerella* in co-occurrence with *Lactobacillus iners, Lactobacillus jensenii, Lactobacillus gasseri, Bifidobacterium, Fannyhessea vaginae,* and *Escherichia coli*	Albert *et al.,* 2015 (64)

			● Co-occurrence and co-dominance of *Gardnerella vaginalis*, Gardnerella leopoldii*, Gardnerella swidsinskii* and Gardnerella greenwoodii**	
Non-pregnant reproductive aged women, Menopausal** women	MSU	Culturomics of resuspended sediment of urine, 16S rRNA amplicon sequencing	● Sixty 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 ● *Gardnerella* genus was identified in both pre and post-menopausal women ● *Lactobacillus* sp. was found to be more common in pre-menopausal women, while *Mobiluncus* sp. was more common genus in post-menopausal women ● No correlation observed between age and number of genera identified	Curtiss *et al.,* 2018 (28)

Non-pregnant reproductive aged women	VS	Sequencing of *cpn60* gene	• *Cpn60* sequence was used to *Gardnerella* spp. identification and differentiate • *Gardnerella vaginalis, Gardnerella swidsinskii, Gardnerella leopoldii, Gardnerella piotii and Gardnerella pickettii** frequently identified • Frequent co-occurrence of *G. vaginalis, G. swidsinskii, G. leopoldii, G. piotii* and *G. pickettii** • Low co-occurrence of *G. leopoldii* and *G. swidsinskii*	Hill and Albert, 2019 (14)
Non-pregnant reproductive aged women	MSU, PUS	EQUC, MALDI-TOF MS	• *Gardnerella, Lactobacillus, Streptococcus agalactiae, Staphylococcus epidermidis* and *Corynebacterium tuberculostearicum* are the most dominant • Oral probiotic usage (*Lactobacillus rhamnosus* GR-1 and *Lactobacillus reuteri* RC-14) did not change the ratio of urogenital microbiota • Probiotic species were not recovered in final samples	Wolff *et al.,* 2019 (66)
Menopausal** women	TUC	SUC, EQUC, 16S rRNA amplicon sequencing	• *Lactobacillus*, *Streptococcus*, *Gardnerella* and *Escherichia* are the most dominant • No correlation between *Lactobacillus* and age, menopausal status, parity or vaginal intercourse	Price *et al.,* 2020 (29)

Non-pregnant reproductive aged women	MSU, PUS	EQUC, 16S rRNA amplicon sequencing	● Differences in microbiota were associated with personal factors e.g., menstruation, vaginal intercourse, sexual partners ● Increase in *Streptococcus* and *Staphylococcus* species after vaginal intercourse ● *Lactobacillus* and *Gardnerella* predominance	Price *et al.,* 2020 (25)
Non-pregnant reproductive aged women, Men	MSU, TUC	SUC, 16S rRNA amplicon sequencing	● *Lactobacillus, Streptococcus, Prevotella, Veillonella,* and *Gardnerella* in high prevalence in women ● Women had higher prevalence of *Lactobacillus* and *Prevotella* than men, but less *Streptococcus, Veillonella* and *Gardnerella* ● MSU and TUC share bacterial genera but in different RA	Pohl *et al.,* 2020 (67)
Non-pregnant reproductive aged women	MSU	EQUC, MALDI-TOF MS, 16S rRNA amplicon sequencing, *cpn60*-based amplicon sequencing	● Changes were detected at two distant timepoints ● FUM can be dynamic over time ● *Gardnerella vaginalis, Gardnerella swidsinskii, and Gardnerella pickettii** identified for the first time in urogenital microbiota of asymptomatic women ● *Lactobacillus crispatus,* often in combination with *G. vaginalis* or *G. pickettii**	Ksiezarek *et al.,* 2021 (15)

Non-pregnant reproductive aged women	Human vaginal metagenome datasets in the HMP, VS	Whole-genome metagenomic sequencing	• *Gardnerella vaginalis, Lactobacillus jensenii, Lactobacillus crispatus, Lactobacillus gasseri, Lactobacillus iners, Fannyhessea vaginae* and *Prevotella amnii* are the most predominant • Difference between vaginal microbiota of Asian and Caucasian women	Liu *et al.,* 2021 (32)
Non-pregnant reproductive aged women	MSU	EQUC, MALDI-TOF MS, 16S rRNA amplicon sequencing, biomarkers (*pheS, cpn60, rpoB, recN, dltS, sodA, malB*)	● 297 bacterial species (median of 53 species/sample) were identified ● Only 22% of the species were detected by culture and sequencing methods ● *Gardnerella vaginalis, Gardnerella leopoldii, Gardnerella swidsinskii* and *Gardnerella pickettii** were identified ● *G. swidsinskii* (RA, 49.26%) identified in combination with *Fannyhessea vaginae* ● *G. vaginalis* (RA, 92.18%) identified in combination with *Lactobacillus gasseri* (RA, 6.39%) ● *G. leopoldii* (RA, 24.55%) identified in combination with *Alloscardovia omnicolens* (RA, 24.55%), *Bifidobacterium spp.* (RA, 24.55%), *Winkia neuii* (RA, 20.62%), *Streptococcus anginosus* (RA, 1.23%)	Perovic *et al.,* 2022 (63)

Non-pregnant reproductive aged women	VS	16S rRNA amplicon sequencing	• *Lactobacillus crispatus* dominated the largest number of samples, followed by the *Lactobacillus iners* and *Gardnerella* • *Lactobacillus jensenii*, *Prevotella*, *Lactobacillus gasseri*, *Bifidobacterium* and *Streptococcus* were dominant in much smaller subsets • *L. crispatus*, *L. jensenii* and *Limosilactobacillus* taxa was positively linked to oestrogen levels and contraceptive use and negatively linked to childbirth and breastfeeding • *Gardnerella, Prevotella* and *Bacteroides*, correlated with multiple partners, menopause, menstrual hygienic products (such as menstrual cup or pads), hygienic habits and contraceptive use	Lebeer *et al.,* 2023 (61)
Pregnant women, complete and draft *Gardnerella* assemblies (GenBank in October 2020)	VS	Whole-genome metagenomic sequencing, *Gardnerella* phylogeny based on 16S rRNA blast database	• *Gardnerella vaginalis*, Gardnerella pickettii*, Gardnerella piotti**, *Gardnerella leopoldii** and *Gardnerella swidsinskii** were the most frequently detected • *G. vaginalis*, G. pickettii*, G. piotti**, *G. leopoldii** and *G. swidsinskii** typically co-occurred with each other and with *Fannyhessea vaginae* and *Lactobacillus iners*	Berman *et al.,* 2024 (60)

			• Authors failed to find a clear association of species with preterm birth	

Abbreviations - CST: Community State Types, EQUC: Enhanced Quantitative Urine Culture, FUM: Female Urogenital Microbiome, HMP: Human Microbiome Project, MALDI-TOF MS (MALDI-TOF VITEK MS): Matrix-Assisted Laser Desorption Ionization Time-Of-flight Mass Spectrometry, Metagenomic analysis: Genomic analysis of microorganisms in a community, by independent-culture analysis techniques, MSU: Midstream Urine, PUS: Peri-Urethral Swab, RA: Relative Abundance, rRNA: Ribosomal Ribonucleic Acid, SUC: Standard Urine Culture, TUC: TransUrethral Catheter, VS: Vaginal Swab, *Previously classified as groups or clades, these are now redefined and assigned to their respective species, **Menopausal: women's menopausal status, encompassing premenopausal, menopausal (transition), or postmenopausal stages.

1.2.3 *Gardnerella* Clinical Manifestations

Gardnerella spp. have been associated with a spectrum of clinical manifestations. It has been considered that *Gardnerella* can become opportunistically pathogenic when its abundance increases relative to other members of the microbial community. This shift in microbial composition, known as dysbiosis, can lead to a breakdown in homeostasis within the urogenital tract (68). *Gardnerella* overgrowth has been linked to several disorders in the FUM, including lower urinary tract dysfunction (LUTD) and lower urinary tract symptoms (LUTS) such as overactive bladder (OAB) and urge urinary incontinence (UUI). Additionally, *Gardnerella* is sometimes detected in polymicrobial urinary tract infections (UTIs) and has been associated with an increased risk of sexually transmitted infections (STIs), possibly due to its ability to compromise mucosal barriers. Most importantly, *Gardnerella* dominance in the vaginal microbiota is a hallmark of bacterial vaginosis (BV), a common condition characterized by vaginal discomfort and discharge (69).

Beyond the urogenital tract, *Gardnerella* infections have been sporadically reported in extra-urogenital sites. *Gardnerella* sepsis (septic gardnerellosis), a life-threatening condition caused by a dysregulated host immune response, has been diagnosed in both healthy and immunocompromised women, as well as in cases of postpartum endometritis and C-section incision infections (70–72). Neonatal infections have also been documented (73,74). *Gardnerella* bacteremia in men is rare, with almost all infections originating in the urinary tract (75).

1.2.3.1 *Gardnerella* and Lower Urinary Tract Syndromes/Disorders

OAB is a subset of LUTS/LUTD, without recognized etiology but thought to be multifactorial (5). It is characterized by a sudden, urgent need to urinate (“urgency”), with or without urinary incontinence, in the absence of a UTI, and is often attributed to disorders of neuro-muscular regulation (76). OAB significantly impacts physical and social functioning, disrupting sleep, work, and social relationships, and urgency incontinence further diminishes health-related quality of life (77). OAB is prevalent in both sexes, and although affects primarily women, new research is now focusing on its impact on men (78). In 2007, the Overactive Bladder Symptom Score (OABSS) questionnaire, self-administered, was developed to provide a validated assessment of OAB symptoms and their impact on

patients' quality of life (79). The OABSS consists of four questions assessing OAB symptoms: daytime frequency, nighttime frequency, urgency, and urge incontinence (79). Patients rate symptom severity, with the total score ranges from 0 to 15, (mild: 3-5, moderate: 6-11, severe: 12-15), with higher scores indicating more severe OAB (80,81). *Gardnerella* has the ability to release large amounts of ATP, inducing Ca^{2+} influx and contraction into urothelial bladder cell lines, suggesting that it might cause the uncontrolled bladder contraction and subsequent voiding in OAB, thus, it is certainly plausible that microbial products could affect neuro-muscular regulation (76). Besides, significant differences were found in the prevalence of certain organisms in the bladder of women with and without OAB. The genera commonly isolated from female patients diagnosed with OAB were *Lactobacillus* (namely *Lactobacillus gasseri*) *Corynebacterium, Streptococcus, Actinomyces, Staphylococcus, Aerococcus* and *Gardnerella* (21). Viral genomes as complete JC Virus (subtype 7A and 7B) and BK Virus (double-stranded DNA viruses and belong to the *Polyomaviridae* family), were also detected in urinary samples from OAB patients (82). Interestingly, a correlation was observed between the urinary microbiome composition and the efficacy of Mirabegron in treating OAB. The findings suggest that a higher prevalence of *Lactobacillus* or a reduction in *Prevotella* and *Gardnerella* may enhance the treatment's effectiveness (83).

Affected patients with UUI, experience urgency, with involuntary urine loss, sometimes followed by immediate leakage of a large volume of urine (84). These conditions have a high impact on patient's quality of life and frequently affect postmenopausal women (26,84). In UUI, as occurs in OAB, *Gardnerella´*s release of large amounts of ATP, enough to stimulate cellular Ca^{2+} influx, may justify the uncontrolled voiding feeling in UUI patients (76). Nevertheless, the bladder microbiome of women with UUI is less diverse than bladder microbiome of asymptomatic women (26,81), being mostly dominated by members of the genera *Gardnerella, Lactobacillus, Actinomyces* and *Prevotella* and members of family *Enterobacteriaceae* (4,26). Additionally, Putonti *et al.,* identified *G. vaginalis, G. swidsinskii* and *G. leopoldii* in samples from patients with UUI (12).

Recently, *Lactobacillus, Gardnerella, Prevotella, Escherichia*, and other bacterial taxa were identified in women with mixed UI (symptoms of both stress and urge urinary incontinence). A lower abundance of *Lactobacillus* and increased bacterial diversity in the urinary microbiome was associated with greater UI severity, suggesting a potential link between microbial composition and disease progression (85).

UTIs are most common in women of reproductive age and can be caused by “classic” uropathogens Gram-negative, as *Escherichia coli* and *Klebsiella pneumoniae* or by Gram-positive as *Staphylococcus saprophyticus, Enterococcus faecalis* and *S. agalactiae* (3). However, fastidious anaerobic bacteria may be easily overlooked due to the use of SUC in laboratories, as the conditions employed do not support their detection (3). Using EQUC, Price *et al.,* identified *Gardnerella* and *Streptococcus* spp. in UTI urines, however, in lower prevalence when compared with urine from women without UTI symptoms (87). The presence of *Gardnerella* in urine has been linked to recurrent urinary tract infections (rUTI), particularly when risk factors such as high *Gardnerella* proliferation or other causative agents of BV are present (88). In fact, in a mouse model, *G. piotii* strain JCP8151B has been shown to induce epithelial exfoliation and promote uropathogenic *E. coli* emergence, which can contribute to recurrent UTIs (89). This action is probably due to the presence of NanH3 alone or in co-occurrence with NanH2, enzymes responsible for extracellular sialidase activity, which may support the observed findings (90).

Table 3 presents a summary of studies reporting the detection of *Gardnerella* in different urogenital disorders.

Table 3: Summary of studies reporting the identification of *Gardnerella* and their main findings on female urinary microbiome in unhealthy conditions.

Population	Sample collection method	Methodology used for *Gardnerella* identification	Main findings	Reference
Overactive bladder (OAB)				
OAB (n=41) Asymptomatic (n=24)	TUC	EQUC, MALDI-TOF MS, 16S rRNA amplicon sequencing	● Evidence of live bacteria in the adult female bladder ● Most bacterial species detected by the EQUC protocol were also identified through 16S rRNA gene sequencing in urine samples (n=4) from the same OAB-women. ● In both cohorts, the most prevalent genera isolated were *Lactobacillus*, *Corynebacterium*, *Streptococcus*, *Actinomyces*, and *Staphylococcus*. Other genera commonly isolated include *Aerococcus, Gardnerella, Bifidobacterium*, and *Actinobaculum* ● *Gardnerella vaginalis* co-occurred with *Lactobacillus jensenii*, *Lactobacillus iners, Rothia dentocariosa, Streptococcus anginosus, Aerococcus urinae, Enterococcus faecalis* in urine of OAB-women	Hilt *et al.,* 2014 (21)

OAB, UTI, SUI and diabetes (n=39) Asymptomatic (n=38)	TUC, VS	SUC, EQUC, modified EQUC protocol, Whole-genome metagenomic sequencing	• *Bifidobacterium bifidum*, was detected in bladder, gut, and vagina • Asymptomatic women contain bacterial genera typically not detected by standard urine culture, including *Lactobacillus, Gardnerella, Streptococcus, Staphylococcus,* and *Corynebacterium*, in low RA • Uropathogenic species, such as *E. coli, Klebsiella pneumoniae, Proteus mirabilis, Enterobacter cloacae, Morganella morganii*, and *Pseudomonas aeruginosa*, represented only 7.7% of microorganisms cultured from bladder • *P. aeruginosa, K. pneumoniae*, and *E. cloacae* were isolated only from the bladder of symptomatic women • *Actinomyces neuii, Lactobacillus crispatus, Lactobacillus gasseri,* and *Lactobacillus jensenii* were highly similar between the bladder and vagina • Bacteria that can reside in both the bladder and vagina could provide protection against urinary infection	Thomas-White *et al.,* 2018 (16)
OAB (n=7), SUI (n = 1) Diabetes (n=1)	TUC	EQUC, MALDI-TOF MS, Whole-genome sequencing	• None of *Gardnerella* species are associated with either continence or OAB with statistical significance	Putonti *et al.,* 2021 (12)

kidney stone (n=1) 103 publicly available *Gardnerella* genomes			• *Gardnerella vaginalis, Gardnerella swidsinskii, Gardnerella leopoldii,* and *Gardnerella pickettii** strains were isolated from asymptomatic patients, as well as individuals with urinary symptoms, without significant differences between *Gardnerella* species and urinary symptoms • Genes that encode sialidase A and O-sialoglycoprotein endopeptidase were identified in *G. vaginalis, Gardnerella piotii*, *G. pickettii*, G.* genomic species 2 and *G.* genomic species 11, while in *Gardnerella* genomic species 7, 12 and 13 genes were not detected	
OAB (n=64)	TUC	SUC, 16S rRNA amplicon sequencing	• Patients with severe OAB had higher bacterial diversity • The urinary microbiome of women who responded effectively to Mirabegron treatment was predominantly dominated by *Lactobacillus* • Higher abundances of *Gardnerella* and *Prevotella* were observed in cases were the treatment was ineffective	Zhou *et al.,* 2022 (83)
Urgency Urinary Incontinence (UUI)/ Urinary Incontinence (UI)/ Urinary Tract Infections (UTI)				

UI, POP, Benign gynecologic conditions (healthy bladder)	MSU, VS, TUC, SPA	SUC, 16S rRNA amplicon sequencing	• Existence of microbial community in the healthy bladder, supported by additional skin and needle samples analysis • *Gardnerella*, *Burkholderia, Ralstonia,* and *Streptococcus* among the low-abundance genera	Wolfe *et al.,* 2012 (23)
UUI (n=45) non-UUI (n=45)	TUC	EQUC, 16S rRNA amplicon sequencing	• Increase of *Gardnerella* and decrease of *Lactobacillus* in UUI group • Urine samples were dominated most frequently by *Lactobacillus* and *Gardnerella* • *Actinobaculum schaalii, Actinomyces neuii, Aerococcus urinae, Arthrobacter cumminsii, Corynebacterium coyleae, Gardnerella vaginalis, Oligella urethralis* and *Streptococcus anginosus* were more frequently identified in UUI group • *Lactobacillus* species-level differences were observed according with cohorts i.e., in healthy: *Lactobacillus crispatus*, in UUI: *Lactobacillus gasseri*	Pearce *et al.,* 2014 (26)
UUI (n=93) non-UUI (n=89)	TUC	16S rRNA amplicon sequencing	• DNA sequencing confirmed urinary bacterial DNA in UUI-women who had no signs of infection • Urine samples were dominated by *Lactobacillus*, followed by *Gardnerella*, *Gardnerella/Prevotella*, *Enterobacteriaceae*, *Staphylococcus*, *Aerococcus* and *Bifidobacterium*	Pearce *et al.,* 2015 (4)

UUI (n=10) non-UUI (n=10)	TUC	16S rRNA amplicon sequencing	• Polymicrobial community in the female bladder in both healthy controls and women with UUI • An increase in the severity of UUI symptoms is associated with a decrease in microbial diversity among UUI group • *Alteromonadaceae* sp., *Stenotrophomonas* sp., *Brevundimonas* sp., *Elizabethkingia* sp. and *Methylobacterium* sp. observed in higher RA in UUI • In bladder, at the genus level, *Gardnerella* stood out as one of those with mean abundances exceeding 2%, alongside *Prevotella, Arthrobacter, Escherichia, Shigella, Anoxybacillus,* and *Lactobacillus*.	Karstens *et al.,* 2016 (84)
rUTI (n=78) non-rUTI (n=18)	TUC	16S rRNA amplicon sequencing	• In recurrent UTI (rUTI) group, urine samples were dominated by *Gardnerella* (42.18%), *Lactobacillus* (24.87%), *Escherichia* (22.57%), and *Haemophilus* (6.52%) • *Atopobium, Megasphaera,* and *Ureaplasma*, known to be associated with BV, were detected only in the rUTI group • In non-rUTI group, urine samples were dominated by *Lactobacillus* (55.67%), *Gardnerella* (30.94%), *Haemophilus* (7.58%), and *Kocuria* (3.6%) • Risk factors such as host immunity, *Gardnerella* proliferation, and microbiome environment seem to influence the development of the phenotype rUTI or non-rUTI	Yoo *et al.,* 2022 (88)

			● After menopause, estrogen reduction and changes in *Lactobacillus* RA are thought to promote UTI	
mixed UI (n=126)	TUC, VS	16S rRNA amplicon sequencing	● Urinary microbiome community types were identified, characterized by varying levels of *Lactobacillus, Gardnerella, Prevotella, Tepidimonas, Acidovorax, Escherichia*, and others ● *Lactobacillus* genus was found to have the highest RA (76%) ● The composition of vaginal community types and urinary community types were similar but composed of slightly different bacterial taxa ● Vaginal community types were not associated with UI severity ● Fewer lactobacilli and more diverse bacteria was associated with more severe mixed UI	Carnes *et al.,* 2024 (85)

Abbreviations - BV: Bacterial Vaginosis, EQUC: Enhanced Quantitative Urine Culture, MALDI-TOF MS (MALDI-TOF VITEK MS): Matrix-Assisted Laser Desorption Ionization Time-Of-flight Mass Spectrometry, Metagenomic analysis: Genomic analysis of microorganisms in a community, by independent-culture analysis techniques, MSU: Midstream Urine, OAB: Overactive bladder, POP: Pelvic Organ Prolapse, RA: Relative Abundance, rRNA: Ribosomal Ribonucleic Acid, SUC: Standard Urine Culture, SPA: SPA: SupraPubic bladder Aspiratio, SUI: Stress Urinary Incontinence, TUC: TransUrethral Catheter, UI: Urinary Incontinence, VS: Vaginal Swab, UTI: Urinary Tract Infections, UUI: Urgency Urinary Incontinence, *Previously classified as groups or clades, these are now redefined and assigned to their respective species.

1.2.3.2 *Gardnerella* and Vaginal Health: Dysbiosis, Infections, and Reproductive Implications

Vaginal health is maintained by a dynamic balance of microbial communities, with *Lactobacillus* species playing a key role in preserving a protective environment. Disruptions to this balance, often referred to as dysbiosis, can lead to various vaginal conditions, including bacterial vaginosis (BV) (69), aerobic vaginitis (AV) (91), vulvovaginal candidiasis (92), and trichomoniasis (93). Among these, BV is the most associated with *Gardnerella* species, historically *Gardnerella vaginalis*, though recent research suggests that multiple *Gardnerella* species may be involved.

- ***Gardnerella* and Bacterial Vaginosis**

Gardnerella proliferation occurs when the healthy vaginal microbiota is disrupted, leading to BV. The term "bacterial vaginosis" is being reconsidered, as not all affected women experience symptoms, and its classification as an infection versus an imbalance remains debated (94,95). Worldwide, BV is the most common reproductive tract disorder among childbearing age women (23–29%; 23% in Europe and Central Asia) (96).

BV is a polymicrobial condition marked by increased abundance of *Gardnerella* spp., *F. vaginae, Prevotella bivia*, *Megasphaera, Bacteroides, Anaerococcus, Peptostreptococcus, Sneathia, Leptotrichia, L. iners,* and members of the class *Clostridia*, alongside a decrease in *L. crispatus* (97–100).

Recent studies indicate that BV is associated with high *Gardnerella* species diversity rather than a single dominant species. Munch *et al.,* found that *G. vaginalis* was the most prevalent species (69.7%), followed by *G. piotii/pickettii* (62.9%), *G. swidsinskii/greenwoodii* (58.6%), and *G. leopoldii* (41.0%), suggesting no single *Gardnerella* species acts as a definitive marker for BV (101). Similarly, Himschoot *et al.,* reported a widespread presence of multiple *Gardnerella* species in BV cases, further supporting the idea that BV results from complex bacterial interactions rather than a single pathogen (102).

A strong association between *Gardnerella* and *F. vaginae* has also been observed in the BV-associated biofilms (98,99,103). Some studies propose *F. vaginae* as a superior diagnostic marker for BV, emphasizing that BV is a structured polymicrobial condition rather

than a simple bacterial overgrowth (102). The mechanisms underlying BV remain unclear, partly due to the lack of a suitable animal model and the complex interplay between *Gardnerella*, other anaerobes, and host factors.

BV is characterized by the occurrence of an excessive thin white vaginal discharge, characteristic fishy odor, especially when a drop of 10% potassium hydroxide (KOH) is added to the vaginal fluid on a slide ("whiff test"), vaginal pH higher than normal (>4.5), and the presence of "clue cells" (epithelial cells covered with Gram-variable polymorphic rods, usually *Gardnerella*) on microscopic examination of vaginal fluid (44,104). These current BV outcomes are present in the Amsel criteria (104) (**BOX 2**). Additionally, the Nugent score, considered the gold standard for diagnosing BV, is based on microscopic examination of Gram-stained vaginal fluid smears (86) (**BOX 2**). This method identifies the presence of Gram-positive rods (such as *Lactobacillus*), Gram-negative and Gram-variable rods and cocci (such as *Gardnerella* spp. and *Prevotella*), and curved Gram-negative rods (like *Mobiluncus*) (44). While these methods remain widely used, they can be influenced by factors like menstruation, recent intercourse, douching, or antimicrobial use. Consequently, alternative diagnostic tools have emerged, including: *i)* a portable desktop ion mobility spectrometer (VGTest-IMS) to detect microbial biogenic amines indicative of BV (105), *ii)* phase-contrast wet mount microscopy (WMM) to evaluate vaginal secretions (106), *iii)* chromogenic tests like OSOM® BVBLUE®, which detects elevated vaginal fluid sialidase, an enzyme produced by some BV-associated bacteria, including *Gardnerella* spp. (107), *iv)* automated systems such as BD Affirm™ VPIII, designed to identification of microorganisms associated with vaginitis (108), and *v)* MALDI-TOF MS applied to vaginal swabs ("VAGI-TOF") (109). Despite promising results, these newer techniques have not yet been widely adopted in clinical practice, leaving BV diagnosis challenging and subjective (109).

- ***Gardnerella* and Adverse Reproductive Outcomes**

The vaginal microbiota plays a crucial role in fertility and pregnancy outcomes (110). *Gardnerella* overgrowth has been linked to infertility, preterm birth, and miscarriage (1,2,111). For instance, Wee and colleagues analyzed samples from 15 infertile and 16 fertile women, finding higher abundance of *Ureaplasma* in the vagina and *Gardnerella* in the cervix of infertile women (112). Recent findings suggest that a vaginal CST IV (*Gardnerella* as the most abundant), may increase the risk of spontaneous preterm birth (sPTB, before 34 weeks) (113). Of note, a statistically significant association between *G. leopoldii* and recurrent sPTB was observed (111). Elevated levels of *L. iners*, *G. vaginalis*, and *F. vaginae* have been linked to high-risk pregnancies, whereas *L. crispatus* dominance

appears protective (114,115). In addition, dysbiosis, characterized by increased microbial diversity and pathogen colonization, may lead to ascending infections that trigger immune and inflammatory pathways, increasing the risk of miscarriage (110).

- ***Gardnerella*, STIs, and Cervical Cancer**

The shift in vaginal microbiota associated with *Gardnerella* overgrowth can increase susceptibility to sexually transmitted infections (STIs), including HPV, HIV, HSV-2, *Chlamydia trachomatis*, and *Neisseria gonorrhoeae* (116–120). This probably occurs due to reduced lactic acid and hydrogen peroxide production, enabling the adhesion, survival, and proliferation of those pathogens (97), and sialidase activity that degrades vaginal mucus, exposing cervical epithelium to pathogens.

Gardnerella has also been linked to cervical cancer, with some studies identifying it as a biomarker for high-risk HPV progression (121). The BV-associated anaerobes also generate biological amines that induce oxidative stress, a key factor in carcinogenesis, potentially driving the development of cervical precancerous lesions (122). Nevertheless, while evidence suggests that BV increases the persistence of HPV and the risk of cervical intraepithelial neoplasia (CIN), other studies have not confirmed this association (123,124).

- ***Gardnerella* and Cervicovaginal Inflammation**

Recent findings highlight the species-specific effects of *Gardnerella* on immune responses. Shvartsman *et al.* reported that *G. leopoldii, G. swidsinskii,* and polymicrobial communities contribute to suppression of interferon-γ inducible protein 10 (IP-10), suggesting a potential immune modulation mechanism (125). Elevated levels of pro-inflammatory markers (IL-1α, IL-1β, IL-6, MIP-1α, MIP-1β) indicate cervicovaginal inflammation, which may compromise epithelial integrity and increase HIV susceptibility (126). Reduced IP-10 levels, alongside increased IL-1α and IL-1β, have been proposed as an immunological biomarker for BV when combined with elevated pH (127).

BOX 2. Common clinical diagnostic tools for BV

Amsel criteria:
The presence of at least three of the following precepts:

- ➢ Excessive thin white vaginal discharge,
- ➢ Characteristic fishy odor - positive “whiff test”,
- ➢ Vaginal pH >4.5,
- ➢ Presence of “clue cells” (**Fig. 1**)

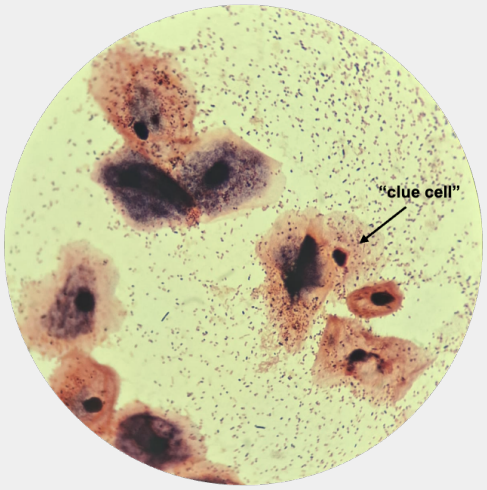


Figure 1: Gram staining of vaginal smears shows the vaginal microbiota (original magnification, ×1000). This photo reveals BV-associated microbiota, highlighting a vaginal “clue cell”.

Nugent Score:
The score evaluates the presence and proportion of different types of bacteria:

Gram-positive rods - *Lactobacillus* morphotypes: Beneficial bacteria.
Gram-negative and Gram-variable rods and cocci - *Gardnerella* and *Prevotella* morphotypes: Associated with BV.
Curved Gram-negative/Gram-variable rods - *Mobiluncus* morphotypes: Strongly associated with BV.

Scoring system (0–10):
0–3: Normal microbiota (predominantly *Lactobacillus*).
4–6: Intermediate microbiota (mixed flora; transition stage).
7–10: Bacterial vaginosis (dominated by *Gardnerella* and/or *Mobiluncus*).

Table 4: Summary of studies reporting the identification of *Gardnerella* and their main findings on vaginal microbiome in unhealthy conditions.

Population	Sample collection method	Methodology used for *Gardnerella* identification	Main findings	Reference
Bacterial Vaginosis (BV)				
BV (n=1) non-BV (n=1)	VS	Whole-genome sequencing	● *Gardnerella leopoldii** strain AMD, BV-associated ● *Gardnerella swidsinskii** strain 5-1, healthy associated ● *G. leopoldii** adheres more avidly to cultured vaginal epithelial cells than *G. swidsinskii** ● *G. swidsinskii** strain exhibited reduced cytotoxicity ● *G. leopoldii** encoded a different variant of a biofilm associated protein gene and demonstrated greater adherence, aggregation, and biofilm formation ● *G. swidsinskii** contained a number of putative phage proteins that were absent in *G. leopoldii** ● *G. leopoldii** contained an apparent toxin antitoxin gene cassette, with 100% identity to a protein from *Fannyhessea vaginae*, and 65% identity to a protein from *Mobiluncus mulieris* suggesting horizontal gene transfer with other BV-associated bacterial species ● It was observed a layer that resembled a polysaccharide capsule on the surface of *G. swidsinskii** that was absent on *G. leopoldii** ● Genomic analysis revealed significant differences between both strains, highlighting potential virulence genes and pathways for BV pathogenesis research	Harwich *et al.*, 2010 (13)

BV (n=7) non-BV (n=7)	VS	16S rRNA amplicon sequencing	• non-BV *Gardnerella vaginalis* strains were less virulent than BV (*Gardnerella vaginalis*) strains • BV strains were able to displace pre-coated vaginal protective lactobacilli • BV strains showed a greater ability to adhere to epithelial cells, and higher cytotoxicity score than non-BV strains • BV strains showed, on average, an expression of vaginolysin 2-fold higher than non-BV strains • No significant differences in expression of sialidases were detected between non-BV and BV strains • There were no significant differences in biofilm formation between non-BV and BV strains • No significant differences were found in antimicrobial resistance profiles between the 2 groups • *G. vaginalis* strains were more susceptible to clindamycin than to metronidazole or tinidazole	Castro *et al.,* 2015 (128)
BV (n=1; *Gardnerella vaginalis*), Other bacterial species associated with BV (n=15)	VS	qPCR targeting 16S rRNA*, vly, sld, HMPREF0424_0821, HMPREF0424_1122, HMPREF0424_0156* and *HMPREF0424_1196* genes	• Distinct biofilm structures were identified for each bacterial consortium, resulting in at least three unique dual-species biofilm morphotypes • Most of dual-species biofilms comprised slightly less than 50% of *Gardnerella vaginalis* • *G. vaginalis* was identified with higher concentration in the lower biofilm layers, suggesting a protective mode in adverse conditions • *Enterococcus faecalis* and *Actinomyces neuii* enhanced *G. vaginalis* vaginolysin expression,	Castro, Machado and Cerca, 2019 (98)

			whereas *Brevibacterium ravenspurgense* was the only species that suppressed it ● The interactions between co-infecting bacteria can significantly influence the progression and clinical outcome of BV, surpassing the impact of the mere presence of specific bacterial species	
BV (n=101) non-BV (n=150)	VS	qPCR targeting *cpn60* gene	● 91.1% of participants with symptomatic BV had 3 or more *Gardnerella* (*G.*) species detected ● *G. vaginalis* was detected in 98.0% of participants with symptomatic BV ● *G. vaginalis* was detected in 50.7% of non-BV although at lower concentrations ● Participants with BV had *G. vaginalis*, *G. piotii/pickettii*, and one or both *G. swidsinskii/greenwoodii* and *G. leopoldii* groups detected ● *G. swidsinskii/greenwoodii* and *G. leopoldii* were less frequently detected and negatively correlated in BV participants ● *G. swidsinskii/greenwoodii* and *G. leopoldii* were almost never exclusively detected together unless *G. vaginalis* or *G. piotii/pickettii* or both groups were also detected ● *Gardnerella* concentrations of all 4 species groups were significantly lower in the week after metronidazole prescription for BV	Munch *et al.*, 2024 (101)

BV (n=254) non-BV (n=271)	VS	qPCR targeting 16S rRNA gene and using *dnaG* primers designed for *Gardnerella* species	• *Gardnerella leopoldii* (54.2%), *Gardnerella piotii* (57.9%), *Gardnerella swidsinskii* (55.7%), *Gardnerella vaginalis* (50.9%) and *Fannyhessea vaginae* (55.0%) occurred in women with BV • *Lactobacillus crispatus* (66.2%) and *Lactobacillus iners* (56.2%) were most abundant among women with a healthy vaginal microbiome • *G. piotii* was associated with dysuria and laboratory markers of UTIs • *G. vaginalis, G. piotii, G. leopoldii, G. swidsinskii* and *F. vaginae* were positively associated with the Nugent score • *L. crispatus* was negatively associated with the Nugent score • *G. vaginalis, G. piotii, G. leopoldii, G. swidsinskii* are involved in BV, although none were associated with the clinically important BV symptoms and/or pregnancy outcomes	Himschoot *et al.*, 2025 (102)
Infertility, spontaneous Preterm Birth (sPTB) and Miscarriage				
9-24 weeks pregnant women (n=783)	VS	Gram's stain, categorization of vaginal microbiome in 3 grades (normal, intermediate, abnormal)	• Abnormal vaginal microbiome was characterized by increased number of *Gardnerella vaginalis*, other morphotypes, or both, while intermediate vaginal microbiome was characterized by reduced lactobacilli mixed with other morphotypes • Association between abnormal and intermediate vaginal microbiome colonisation with PTB and late miscarriage	Hay *et al.,* 1994 (1)

Infertile women (n=15) Fertile women (n=16)	VS, cervical and endometrial samples	16S rRNA amplicon sequencing	• Infertile women had prevalence of *Ureaplasma* in the vagina and *Gardnerella* in the cervix • Tenascin-C expression was correlated with case of miscarriage	Wee *et al.,* 2017 (112)
<34 and between 34-36 weeks pregnant women with PTB (n=94) ≥37 weeks pregnant women (term delivery, n=356)	VS	16S rRNA amplicon sequencing	• *Lactobacillus gasseri/Lactobacillus johnsonii, Lactobacillus crispatus, Lactobacillus acidophilus, Lactobacillus iners, Ralstonia solanacearum and Bifidobacterium longum/ Bifidobacterium breve* were associated with decreased risk of early (<34 weeks) but not late (34-36 weeks) preterm birth • Early sPTB was higher in CST IV associated with *Gardnerella vaginalis, Fannyhessea vaginae* and *Veillonellaceae* bacterium	Tabatabaei *et al.,* 2018 (113)
Pregnant women (n=25) Women with embryonic miscarriage (n=25)	VS	16S rRNA amplicon sequencing	• *Lactobacillus* was significantly decreased in women with embryonic miscarriage (56%), compared with women with normal pregnancy (88%) • Women with embryonic miscarriage showed a significantly higher level of diversity compared with women with normal pregnancy • *Gardnerella* genus was identified in both groups • *Gardnerella, Prevotella, Megastrobila,* and *Cyclospora*, and low *Lactobacillus* RA may increase the chance of miscarriage	Xu *et al.,* 2022 (114)

<24 weeks pregnant women (n=221)	VS	Whole-genome metagenomic sequencing	• Of 221 women, 192 (87%) delivered at term and 29 (13%) experienced sPTB • In 99% of samples, the most common species specimens collected was *Lactobacillus iners* (RA, 2–77%) followed by *Gardnerella* genus (RA, 2–52%) • Two groups dominated by *Gardnerella* and *L. iners* species predicted sPTB • *Gardnerella* metagenomic subspecies type 2, characterized by diverse mix of other subspecies of *Gardnerella*, was more abundant among women having sPTB and in those with HIV • *Gardnerella* metagenomic subspecies type 1, dominated by *Gardnerella swidsinkii/Gardnerella leopoldii* was less abundant in women having sPTB and in those with HIV	Price *et al.,* 2022 (129)
<16 weeks pregnant women at high risk for sPTB (n=154)	VS	16S rRNA amplicon sequencing, qPCR performed to differentiate *Gardnerella* spp.	• The majority of samples (86.2%) had at least one *Gardnerella* species detected • *Gardnerella vaginalis* (57.1%) and *Gardnerella piotii* (57.3%) had a similar prevalence, and were followed by *Gardnerella swidsinskii* (44.4%) and *Gardnerella leopoldii* (22.2%) • In 27.4% of all samples, a single *Gardnerella* spp. was present (*G. leopoldii* 2.3%, *G. piotii* 9.5%, *G. swidsinskii* 5.2% and *G. vaginalis* 10.4%) • In 58.8% of all samples, a combination of *Gardnerella* spp. was present • The distribution of *Gardnerella* spp. changed over time in 91.3% of women	Schuster *et al.,* 2024 (111)

			• Women treated for BV, demonstrated a change in microbiota profile from mainly CST IV (*Gardnerella* and other anaerobes dominated) to mainly CST III (*Lactobacillus iners* dominated) • *Gardnerella* spp. did not have statistically significant associations with sPTB, except *G. leopoldii* • No association was observed between the presence of STIs and sPTB	
Sexually Transmitted Infections (STIs)				
HIV-1–seronegative women (n=657)	VS, cervical samples	Gram's stain, Nugent Score	• Abnormal vaginal microbiome was significantly associated with HIV-1 infection and trichomoniasis, but not with vaginal candidiasis • For HIV-1 prevention, lactobacilli appeared to be beneficial, particularly H_2O_2-producing strains	Martin *et al.,* 1999 (130)
HSV-2–seronegative women (n=670)	VS	Gram's stain, Nugent Score	• BV increases the susceptibility to acquisition of HSV-2 in women • Behavioral factors, such as new sexual partners, may account for the association between BV and HSV-2 acquisition	Cherpes *et al.,* 2003 (117)
Reproductive-age women, non-pregnant (n=3620)	VS	Gram's stain, Nugent Score	• A Nugent score intermediate (4-6) or BV indicative (7-10) conferred a 1–2-fold increased risk for incident trichomonal, gonococcal, and chlamydial infection	Brotman *et al.,* 2010 (116)
Women with an incident high-risk HPV infection (n=273)	VS, cervical samples	16S rRNA amplicon sequencing	• *Gardnerella, Prevotella* and *Anaerococcus,* were the bacterial species with the highest positive correlation to high-risk HPV development and progression	Usyk *et al.,* 2020 (121)

			• Five fungal taxa associated with HPV progression • *Gardnerella* presence and high cervicovaginal microbiome diversity are key biomarker associated with cervical intraepithelial neoplasia grade 2 and 3 (CIN2+) • Presence of commensal bacteria, such as *Lactobacillus* have the ability to stimulate a local immune response, contributing factors to the clearance of incident high-risk HPV infections	
Women with low risk to HIV and seropositive to varicella zoster virus (n=45)	VS, cervical Samples	16S rRNA amplicon sequencing	• *Gardnerella leopoldii*, Gardnerella swidsinskii** appear to be the most common in VS regardless BV status • *Gardnerella* dominance and polymicrobial community structures are associated with increased mucosal pro-inflammatory cytokines • *G. leopoldii*, G. swidsinskii** and polymicrobial communities, contribute to IP-10 suppression, suggesting a potential mechanism of immune modulation • non-*Lactobacillus* dominant microbial communities promote mucosal inflammation which may in turn increase HIV susceptibility but are not necessarily associated with increased HIV target cells in this low-risk cohort	Shvartsman *et al.,* 2023 (125)

Abbreviations - BV: Bacterial Vaginosis, CIN: Cervical Intraepithelial Neoplasia, CST: Community State Types, EQUC: Enhanced Quantitative Urine Culture, HIV: Human Immunodeficiency Virus, HPV: Human Papillomavirus, HSV-2: Herpes Simplex Virus type 2, Metagenomic analysis: Genomic analysis of microorganisms in a community, by independent-culture analysis techniques, RA: Relative Abundance, rRNA: Ribosomal Ribonucleic Acid, sPTB: spontaneous Preterm Birth, STIs: Sexually Transmitted Infections, VS: Vaginal Swab, *Previously classified as groups or clades, these are now redefined and assigned to their respective species.

1.3 *Gardnerella*'s Genomic Landscape: Insights from Genome Mining

Genome mining, an advanced bioinformatics-driven approach, has emerged as a powerful tool for exploring *Gardnerella'*s evolutionary trajectory, niche adaptation mechanisms, and putative virulence factors.

Since the pioneering work of Vaneechoutte *et al.,* (2019) (10), only eight studies (11,12,131–136) have explored *Gardnerella* spp. using WGS, significantly limiting our ability to develop effective treatment strategies and mitigate recurrence rates in bacterial vaginosis (BV).

Genome mining using different bioinformatic tools and datasets resulted in a significant variability in what is considered *Gardnerella*'s core genome, which ranges from 138 to 608 core genes and includes approximately 176 core proteins (134,136). One study identified 2,399 protein families within *Gardnerella*, of which only 29% belonged to the core genome, while nearly half (49.4%) were classified as hypothetical proteins with unknown functions (9). Moreover, genomic analyses have uncovered a high degree of strain-level diversity in the presence of antibiotic resistance genes. Dillard *et al*. reported that most *Gardnerella* genomes (n=110) carry resistance genes for rifamycin (98%), streptogramins (85%), lincosamides (81%), and pleuromutilins (81%), while genes for resistance to tetracyclines (20%), macrolides (13%), and aminoglycosides (1%) were less common (136).

Compared to other members of the *Bifidobacteriaceae* family, *Gardnerella* exhibits a lower GC content (40.99–42.9 mol%) and a relatively small genome (1.47–2.01 Mbp) (9,10). Despite its compact genome size, *Gardnerella* harbors a substantial array of virulence-associated genes, suggesting a strong capacity for competitive exclusion of other vaginal microbiota members (137). As previously referred, the number of genomic species identified within the *Gardnerella* genus varies depending on the analytical tools used, with estimates ranging from 8 to 14 distinct genomic species (11,12,131,132,134). These species can be further distinguished based on differences in core and accessory genomes (11,131).

Recent studies have highlighted four core metabolic genes - *gpsA* (K00057), *fas* (K11533), *suhB* (K01092), and *psd* (K01613) - as essential to *Gardnerella*'s metabolic function (136). Additionally, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) spacer sequences have been used to differentiate *Gardnerella* strains at species level (138). Furthermore, genes encoding competence-promoting proteins ComEA, ComEC, and CinA,

typically found in naturally competent bacteria, have been identified within *Gardnerella* genomes, suggesting potential horizontal gene transfer mechanisms (137).

As a key member of the urogenital microbiome, *Gardnerella* exhibits a complex interplay between commensalism and pathogenicity, influenced by microbial interactions and environmental factors. Integrating bioinformatics-driven genomic analyses with phenotypic assays is crucial to elucidate *Gardnerella*'s role in vaginal health and disease. These combined approaches will aid in the development of precision-based diagnostics and therapeutic interventions, ultimately improving women's reproductive health outcomes.

1.3.1 Virulence Factors and Antimicrobial Resistance in *Gardnerella*

The ability of *Gardnerella* species to establish and persist within the urogenital microbiome is largely determined by their virulence factors and resistance mechanisms. While traditionally associated with bacterial vaginosis (BV), *Gardnerella* strains exhibit substantial genetic and functional diversity, influencing their role in both asymptomatic colonization and symptomatic infections.

Key virulence factors identified in *Gardnerella* include sialidases (NanH1, NanH2, and NanH3), which degrade mucosal barriers and promote biofilm formation; vaginolysin (VLY), a pore-forming toxin implicated in host cell lysis; and phospholipase (*pho*), a potential contributor to tissue damage (100,139–141). Additional factors, such as type I glycosyltransferases linked to biofilm development and type IV Flp pili associated with epithelial adhesion, further enhance the pathogenic potential of certain strains (132). However, not all *Gardnerella* strains possess or express these factors at levels that contribute to disease, highlighting the complexity of their virulence.

Biofilm formation plays a central role in the persistence of *Gardnerella* within the vaginal microbiome (142). These bacterial communities protect *Gardnerella* and other BV-associated species from host immune responses and antimicrobial treatment, contributing to recurrent infections (13,37,98,100,139). While *Gardnerella* has been described as a dominant species in BV-associated biofilms, evidence suggests that not all strains exhibit strong biofilm-forming capacity (98,99). The presence of other microorganisms, such as *Enterococcus faecalis* and *Actinomyces neuii*, may also influence BV progression more significantly than traditionally recognized species like *Mobiluncus mulieris* or *Fusobacterium vaginae* (98).

Antimicrobial resistance (AMR) in *Gardnerella* is a less explored factor affecting treatment outcomes. Nevertheless, resistance genes such as *lsaC* (macrolides, lincosamides, streptogramins A, and pleuromutilins resistance), *ermX* (macrolide resistance), and *tetL/tetM* (tetracycline resistance) have been already identified in various strains (133).

1.3.1.1 Sialidases: Key Enzymes in Pathogenicity?

Sialidases, also known as neuraminidases, are key enzymes that contribute to the pathogenicity of various microorganisms, including *Gardnerella* and *Prevotella* species (143). These enzymes cleave terminal sialic acid residues from glycoconjugates in the vaginal environment, providing *Gardnerella* spp. with a nutrient source through sialic acid catabolism. Beyond nutrient acquisition, sialidase activity disrupts the vaginal mucosal barrier, facilitates bacterial colonization, and promotes biofilm formation (90,98,144).

The role of sialidases in *Gardnerella*-associated infections has been further demonstrated in an animal model. In a mouse study, the *G. piotii* strain JCP8151B was found to contribute to vaginal sialidase activity, leading to epithelial exfoliation and enhancing the invasive potential of *Prevotella bivia* by facilitating its ascension into the uterus (145).

Initially, the *nanH1* gene (formerly *sialidase A*) was thought to be the primary driver of sialidase activity in *Gardnerella* spp., as it was present in sialidase-positive strains (146). However, its detection in sialidase-negative strains and the absence of a signal peptide suggest that NanH1 functions intracellularly, likely playing a role in nutrient metabolism rather than extracellular sialidase activity (90). Subsequent research identified two additional sialidase genes, *nanH2* and *nanH3*, which are unevenly distributed among *Gardnerella* species (90). Biochemical studies revealed that NanH2 and NanH3 efficiently cleave both α-2,3- and α-2,6-linked terminal sialic acids attached to N- and O-linked sialoglycans, making them the primary enzymes responsible for extracellular sialidase activity in *Gardnerella* (90). Unlike NanH1, these proteins contain signal peptides or C-terminal transmembrane domains, indicating that they are either anchored to the cell surface or secreted extracellularly via proteolytic processing.

Interestingly, a polycytosine homopolymer (8-14 cytosines) was discovered within the *nanH3* gene of *G. piotii* and *G. pickettii* strains (147). These homopolymeric sequences are prone to slipped-strand mispairing, a genetic mechanism that enables phase variation by switching sialidase expression "on" or "off." This dynamic regulation may enhance the

pathogenicity of certain *Gardnerella* strains by modulating bacterial adhesion, biofilm initiation and dispersal, adaptation to diverse host environments, and immune evasion (147). A more recent discovery added another layer of complexity to *Gardnerella* sialidases. The *nanH4* gene, identified in 12 *G. vaginalis* strains, encodes a NanH1-like protein, lacking a signal peptide and predicted to be intracellular. However, its enzymatic activity remains uncharacterized, and further biochemical studies are needed to determine its functional role (143).

Sialidase activity is commonly used as a diagnostic marker for BV, with rapid clinical detection facilitated by tests such as OSOM® BVBLUE® (107). The presence of sialidase-encoding genes in *Gardnerella* strains supports the idea of a potential distinction between pathogenic and commensal strains (132). However, species-level variation in sialidase expression has been observed. For instance, Bulavaitė *et al.,* reported that *G. leopoldii* and *G. swidsinskii* exhibit a sialidase-negative phenotype, emphasizing the functional diversity of *Gardnerella* species (48).

Given the role of sialidases in *Gardnerella* pathogenesis - compromising the protective vaginal mucus layer and promoting biofilm formation - sialidase inhibitors represent a promising avenue for novel BV therapies. Zanamivir, a known sialidase inhibitor, has been shown to reduce sialidase activity and inhibit the invasion of human cells by *G. piotii* strain JCP8066 and *G. pickettii* strain JCP8017A (148). Additionally, Zanamivir targets both bacterial and human sialidases, such as NEU1 and NEU3, which helps to reduce inflammation and preserve functional IgA, a key factor in preventing bacterial adhesion to mucosal surfaces (149,150). This dual action suggests potential benefits for both controlling bacterial virulence and mitigating host inflammation. However, the specificity of sialidase inhibitors varies. While the broad-spectrum inhibitor Neu5Ac2en effectively suppressed all tested vaginal bacterial sialidases, Zanamivir successfully inhibited *Prevotella timonensis* sialidases but failed to inhibit *Gardnerella* NanH3 (143). These findings highlight the need for the development of more targeted inhibitors to effectively counteract *Gardnerella* sialidases without disrupting beneficial host enzymes.

1.3.1.2 Insights into the role of vaginolysin

Vaginolysin (VLY), encoded by the *vly* gene, is a cholesterol-dependent cytolysin that plays a crucial role in *Gardnerella* pathogenicity. It targets the complement regulatory protein CD59 on human epithelial cells, neutrophils, and erythrocytes, leading to pore formation

and cytotoxicity (13,137,151). It is reported that the *vly* gene is present in approximately 83% of *Gardnerella* strains and is part of the accessory genome, suggesting acquisition through horizontal gene transfer (131). It is of note that *Gardnerella* strains isolated from women with bacterial vaginosis (BV) express *vly* at significantly higher levels than those from women without BV (128).

VLY is active within a pH range of 4.5 to 7.0, aligning with conditions observed in BV (>4.5) (152,153). Strain-dependent differences in VLY production have been observed, with toxin concentrations varying from 0.05 to 1.305 μg/mL (154). Based on sequence variations in the undecapeptide region and the CD59-binding domain, VLY has been classified into five types (1A, 1B, 1C, 2, and 3) (151). Among them, Type 1A and Type 2 exhibit the highest cytotoxicity, with Type 1A causing greater damage to the apical vaginal mucosa due to its strong CD59-binding affinity (151).

Mechanistically, VLY induces cytotoxicity via two distinct pathways: a cholesterol-dependent, CD59-independent mechanism and a CD59-assisted pathway, similar to intermedilysin (ILY) (155). Structural studies suggest that the undecapeptide loop can adopt different conformations, allowing for dual functionality. Moreover, a P480W mutation in VLY disrupts its ability to activate the p38 mitogen-activated protein kinase pathway and reduces IL-8 expression in HeLa cells, leading to lower hemolytic activity and highlighting its immunopathological role (156,157).

Given its significance in *Gardnerella* pathogenicity, VLY has become a target for improved diagnostic and therapeutic strategies. A recently developed rapid diagnostic test combines an enzyme-linked immunosorbent assay (ELISA) with a lateral flow assay (LFA) for VLY detection (158). Additionally, a smartphone-based microscopy kit for clue cell visualization provides a cost-effective, portable solution for BV diagnosis (158). These advancements underscore the importance of VLY in BV pathology and its potential as a therapeutic target for novel treatments aimed at mitigating its cytolytic effects.

1.3.1.3 Antibiotic Resistance Genes (ARGs): Distribution, Implications and Alternative Treatments

Antibiotic resistance is also a concern in *Gardnerella* species, potentially contributing to the high recurrence rates of bacterial vaginosis (BV). Several resistance genes have been

identified across different *Gardnerella* species. The *ermX* gene, linked to macrolide, lincosamide, and streptogramin resistance, has been found in *G. vaginalis*, *G. piotii*, and *G. swidsinskii*. The *lsaC* gene, conferring resistance to lincosamides, streptogramins A, and pleuromutilins, was detected in *G. vaginalis*, *G. piotii*, *G. pickettii*, and *G. greenwoodii* (135). Additionally, tetracycline resistance genes (*tetL* and *tetM*) were found exclusively in *G. vaginalis*, while the *nim* gene, associated with metronidazole resistance, was identified in *G. swidsinskii* (133,159).

Phenotypic studies confirm the presence of antibiotic-resistant *Gardnerella* strains. Some isolates exhibit high resistance to metronidazole (≥128 µg/mL), the primary treatment for BV. In one study, all *G. leopoldii* and G*. swidsinskii* strains were metronidazole-resistant, while resistance was observed in 35% of *G. vaginalis* and only 7.1% of *G. pickettii* and *G. piotii* (160). Additionally, authors proposed that proteins YadH and McrA might be implicated in metronidazole resistance (133).

Despite increasing resistance, standardized susceptibility testing guidelines for *Gardnerella* are lacking from major regulatory bodies like CLSI and EUCAST. The 2023 WHO guidelines recommend metronidazole (400–500 mg orally twice daily for 7 days) as the primary BV treatment (161). Alternative options include vaginal metronidazole gel (0.75% for 7 days), tinidazole (2 g single dose, except in pregnancy), clindamycin (300 mg twice daily for 7 days or 2% gel intravaginally), and secnidazole (2 g single dose) (161).

For severe infections, particularly septic gardnerellosis, combination therapies have been used, including metronidazole with ceftriaxone or piperacillin–tazobactam, as well as clindamycin with ampicillin or amoxicillin-clavulanate (162,163). Treatment duration varies, often requiring adjustments based on bacterial culture results and patient response. While most patients recover, adverse outcomes, including pregnancy complications, have been reported (70,162,163).

While metronidazole treatment significantly reduces BV-associated bacteria and increases *Lactobacillus* abundance, *Gardnerella* populations often rebound over time. *Gardnerella* biofilms, which protect bacteria from antibiotics, exhibit higher metronidazole resistance than planktonic cells (160). This contributes to BV recurrence and highlights the need for early detection of metronidazole-resistant strains.

Alternative approaches are being explored to improve treatment outcomes. Clindamycin, either alone or combined with probiotics such as *Lactobacillus acidophilus* KS400 or

Lactobacillus rhamnosus (a bacteriocin producer), has shown promise in inhibiting *Gardnerella* and other anaerobes (164). Additionally, emerging strategies such as antimicrobial peptides, bacteriophage therapy, probiotics, and vaginal microbiome transplantation (VMT) offer potential solutions for recurrent and antibiotic-resistant BV (164,165).

Further research into regional drug susceptibility patterns and *Gardnerella*'s role in BV pathogenesis is essential for optimizing treatment strategies and reducing recurrence rates.

1.3.1.4 Nutrient Competition and Microbial Dynamics

Nutrient competition plays a key role in shaping microbial communities, influencing their composition and function. In the vaginal microbiome, *Gardnerella* competes with other species for essential nutrients such as amino acids and carbon sources (e.g., glycogen) (134). This competition drives adaptations that impact *Gardnerella*'s growth, survival, and ability to outcompete a *Lactobacillus*-dominated community. *Gardnerella* has been proposed to follow a "scramble" competition strategy, where it rapidly depletes resources, limiting their availability to others (134). This may explain why *G. vaginalis* and other *Gardnerella* species can outcompete their counterparts *in vitro* by utilizing a broader range of nutrients (134).

Additionally, *Gardnerella* is thought to contribute to the vaginal microbiome by producing "public goods," enriching the shared nutrient pool and supporting the growth of less abundant microorganisms like *G. greenwoodii*. However, the role of microbial carbohydrate metabolism, particularly *Gardnerella*'s ability to utilize glycogen, remains poorly understood (134,166).

Glycogen, a key nutrient in the vaginal environment, is a linear polymer of α-1,4-linked glucose chains with periodic α-1,6 branches (167). It is stored in vaginal epithelial cells and can be released into the lumen upon cell lysis (167). Most glycogen-degrading bacteria produce amylases from the glycosyl hydrolase family 13 (GH13), a subgroup of carbohydrate-active enzymes (CAZymes) (167). Several *Gardnerella* species, including *G. vaginalis, G. pickettii, G. piotii, G. swidsinskii*, and *G. leopoldii*, produce extracellular amylases that break down glycogen into glucose, maltose, maltotriose, and maltotetraose (167). These glycogen-debranching enzymes, anchored to the cell envelope, include a single-domain α-amylase (EC 3.2.1.1) and an α-amylase-pullulanase (EC 3.2.1.41). These enzymes hydrolyze α-1,4 glycosidic bonds, with the α-amylase-pullulanase also targeting

α-1,6 glycosidic bonds (167). Interestingly, *G. vaginalis* appears to have a unique enrichment in amylase and pullulanase proteins compared to *Lactobacillus* species, suggesting distinct nutritional strategies between these bacterial groups (168).

Once glycogen is broken down, *Gardnerella* species use specialized ABC sugar transporters to take up the resulting sugars. These transporters differ among species, highlighting their metabolic diversity:

- **MalXFGK** (found in *G. vaginalis, G. piotii, G. leopoldii*, and *G. swidsinskii*)
- **MusE_2FGK_2I** (exclusive to G. swidsinskii)
- **MusEFGK_2I** (*G. vaginalis, G. piotii, G. leopoldii*)
- **RafEFGK** and **TMSP** (exclusive to *G. vaginalis*)
- **MalEFG** (exclusive to *G. leopoldii*)

This diversity in sugar transport mechanisms reflects *Gardnerella*'s ecological specialization, allowing different species to efficiently utilize glycogen-derived sugars. Such adaptations may provide a competitive advantage over other vaginal microorganisms like *Lactobacillus*, particularly under dysbiosis conditions like BV (166).

Despite growing insights into *Gardnerella*'s genomic diversity and metabolic adaptations, many questions remain regarding its role in urogenital health and disease, particularly its species-specific traits, pathogenic potential, and contribution to antimicrobial resistance. Addressing these gaps is essential for improving diagnostic and therapeutic strategies.

2. Objectives

This study explored the diversity of *Gardnerella* species and strains within the female urogenital microbiome, focusing on isolates from symptomatic and asymptomatic women. Using whole-genome sequencing, biochemical assays, and antibiotic susceptibility testing, we integrated participant-derived isolates with publicly available genomes to investigate their genetic, phenotypic, and ecological characteristics, providing a basis for understanding their clinical significance.

Hypothesis: The genetic diversity, virulence factors, and metabolic capabilities of *Gardnerella* species shape their adaptability and pathogenicity within the urogenital microbiome. Species-specific and strain-specific traits may influence clinical outcomes, including bacterial vaginosis (BV) and antimicrobial resistance, affecting host interactions and microbiome stability.

The **main goal** of this thesis was to assess the genomic and functional diversity of *Gardnerella* spp. and their roles within the urogenital microbiome by comparing isolates from symptomatic and asymptomatic women, with the goal of identifying species- and strain-specific traits relevant to health and disease.

The **specific goals** were the following:

i) Optimize DNA extraction protocols for *Gardnerella* species to ensure high-quality genomic data.
ii) Characterize the pangenome and species-specific traits, comparing strains from symptomatic *vs.* asymptomatic individuals.
iii) Integrate newly identified species (*G. pickettii* and *G. greenwoodii*) into the broader taxonomic and functional framework of the genus.
iv) Investigate sialidases in host-pathogen interactions, elucidating their functional properties and clinical significance.
v) Analyze vaginolysin (*vly*) diversity and regulation, to understand its contribution to virulence.
vi) Examine glycogen metabolism as a competitive advantage, assessing its impact on *Gardnerella*'s ecological role.

vii) Bridge phenotypic and genotypic antimicrobial resistance**,** addressing the implications for treatment strategies.
viii) Explore the potential clinical applications of the results, with an emphasis on improving diagnostics and targeted therapeutic strategies.

By identifying key genomic and functional traits, this thesis aimed to clarify the potential role of *Gardnerella* in health and disease. Ultimately, these findings will contribute to a better understanding of *Gardnerella*'s impact on the urogenital microbiome and inform future diagnostic and therapeutic strategies.

3. Results and Discussion

Unraveling the *Gardnerella* Pangenome: Insights into Genetic Diversity, Virulence, and Antibiotic Resistance

Pedro Teixeira,[1,2#§] Márcia Sousa,[1,2§] Teresa Gonçalves Ribeiro,[1,2,3] Filipa Grosso,[1,2] Luísa Peixe [1,2,3*]

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

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

[3]Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP).

[#]current affiliation: National Reference Laboratory of Antibiotic Resistances and Healthcare Associated Infections, Department of Infectious Diseases, National Institute of Health Dr. Ricardo Jorge, 1649-016 Lisbon, Portugal

[§]These authors contributed equally to this work.

*Luísa Peixe, lpeixe@ff.up.pt

Manuscript in preparation

Abstract

In recent years, the development of user-friendly and accessible bioinformatic tools has significantly advanced *Gardnerella* genome analysis. However, despite the taxonomic reclassification of the genus in 2019, many studies continue to rely on subgroups or clades for species identification. This practice leads to frequent misidentifications, hindering accurate assessments of species-specific features, such as antimicrobial resistance and pathogenic potential. This study aims to address these gaps by employing precise species-level identification to improve our understanding on *Gardnerella* and its clinical implications. We performed the most extensive pangenome analysis of *Gardnerella* to date, analyzing 153 genomes (120 from NCBI and 33 newly sequenced).

G. vaginalis, G. leopoldii, G. swidsinskii, and *G. pickettii* species exhibited open pangenomes. The *Gardnerella* pangenome is highly diverse, with 90% of its genes belonging to the accessory genome. Carbohydrate transport and metabolism were the features that contributed the most for the primary genomic differences observed among *Gardnerella* species. Additionally, the search for antibiotic resistance genes (ARGs) revealed that the *lsaC* gene was the most prevalent ARG, potentially conferring resistance to clindamycin, a commonly prescribed antibiotic for treating bacterial vaginosis (BV). The clustering of *Gardnerella* genomes mimics the core genome phylogeny, pointing to species specific virulence-associated factors. Furthermore, we observed species clustering based on their virulence factor profiles, highlighting important factors such as vaginolysin and sialidases.

Our comprehensive pangenome analysis shows that *Gardnerella* species are well-adapted within the urogenital tract, with identified virulence factors and metabolic capabilities potentially enhancing the pathogenicity of some strains. Future studies should explore *Gardnerella*'s ecological interactions and pathogenic mechanisms within the urogenital microbiome.

Keywords: *Gardnerella,* Pangenome analysis, female urogenital microbiome, pathogenicity

Background

Historically, *Gardnerella vaginalis* was considered the unique species within the genus. However, taxonomic revision in 2019 led to the emended description of *G. vaginalis* and the identification of three additional species: *Gardnerella leopoldii*, *Gardnerella piotii*, and *Gardnerella swidsinskii*, along with nine distinct *Gardnerella* genomic species (GG) (1, 2). Further studies expanded this diversity with the recognition of an additional GG (3). More recently, our research group formally described two GG, *Gardnerella pickettii* (formerly GG3) and *Gardnerella greenwoodii* (formerly GG8) (4).

Gardnerella, particularly *G. vaginalis*, has long been associated with bacterial vaginosis (BV), a common vaginal dysbiosis linked to adverse pregnancy outcomes and increased susceptibility to sexually transmitted infections (5–8). However, the discovery of new *Gardnerella* species challenges the long-held assumption that *G. vaginalis* is the only species involved in BV and other urogenital infections. Moreover, *Gardnerella* species, including *G. vaginalis*, have been isolated from the urogenital tract and midstream urine of asymptomatic women (9, 10), highlighting their potential role as commensals. It is believed that BV development is triggered by the formation of polymicrobial biofilms, facilitated by the action of sialidases and the secretion of vaginolysin (11, 12). Notably, not all *Gardnerella* strains, including some *G. vaginalis* strains are associated with BV, exhibit sialidase activity nor encode vaginolysin (1, 13–15). Recent studies have also identified novel sialidases and vaginolysin variants (13–15), underscoring the need for further investigation into their roles in BV pathogenesis.

Pangenome analysis provides valuable insights into the bacterial genetic diversity, evolution, and functional traits (3, 16, 17). This approach has been applied in *Gardnerella* research to explore species delineation, dynamics of lateral gene flow, metabolic capabilities, virulence factors (VFs), and evolutionary history (3, 16–21). However, some earlier studies relied on outdated taxonomic classifications (e.g., groups or clades encompassing different *Gardnerella* species) or misidentified isolates (22–24), emphasizing the need for caution when interpreting previous findings.

The increasing availability of *Gardnerella* genomes presents an opportunity to refine pangenome analysis by integrating updated taxonomic frameworks. A comprehensive genomic approach can provide a deeper understanding of the genetic factors and mechanisms that shape the diversity and pathogenic potential of *Gardnerella* species.

In this study, we present the most extensive *Gardnerella* genomic analyses to date, incorporating the latest taxonomic updates. Our research explores the phylogenetic relationships, resistome, and virulome of *Gardnerella*, providing novel insights into species-specific traits and potential pathogenicity. Additionally, we investigate the colonization patterns and emphasize the importance of host health status in understanding the complex interplay between *Gardnerella* and the urogenital tract.

Results

Overview of the *Gardnerella* Genomes Collection

A total of 153 *Gardnerella* genomes were analyzed, including 120 publicly available genomes (15 previously unexplored) and additional 33 newly sequenced genomes from our collection (Table S1). Most genomes were available in draft form, reflecting the current state of *Gardnerella* genome assembly. The majority of the *Gardnerella* strains analyzed (59.7%, n=89) originated from vaginal samples, with the remaining strains primarily sourced from urine samples (38.3%, n=57). We assessed the correlation between *Gardnerella* strains and human health conditions, identifying BV (38.3%, n=57) and overactive bladder syndrome (12.1%, n=18) as the most prevalent conditions, while 27.5% (n=41) of isolates originated from asymptomatic human. Unfortunately, a subset of *Gardnerella* isolates (18.1%, n=27) catalogued in the National Center for Biotechnology Information (NCBI) database lacked information regarding the human health status.

Genomic Analyses Reveal Frequent Misidentification of *Gardnerella* sp.

Approximately 57% of *Gardnerella* genomes in the NCBI Refseq database are misidentified or classified as *Gardnerella* sp. or *G. vaginalis*, likely due to their deposition before the formal description of new *Gardnerella* species (1, 3, 4). We employed a comprehensive genomic approach, combining Average Nucleotide Identity (ANI), digital DNA-DNA hybridization (dDDH), and core genome phylogeny to delineate the species and infer the phylogenetic relationships among the *Gardnerella* strains.

ANI and dDDH network analysis grouped *Gardnerella* strains into 16 clusters, adhering to established species delineation thresholds (70% for dDDH and 96% for ANI) (25–27) (Figure 1). Two previously unreported genomic clusters were identified: a potential new genomic species (GG15), comprising strains N165 and JCP7275, and GG16, represented

by strain JNFY17 (Figure 1). However, core-genome phylogenetic analysis of 92 single-copy genes repositioned these strains within GG2 and *G. pickettii*, respectively. Moreover, the phylogenomic analysis supported a classification of 6 species and 8 GG, consistent with previous research (Figure 2) (3, 18).

Among the identified species, *G. vaginalis* stands out as the most prevalent (38.5%, n=59), followed by *G. pickettii* (15.0%, n=23), *G. leopoldii* (13.1%, n=20), *G. swidsinskii* (12.4%, n=19), *G. piotii* (5.9%, n=9), and *G. greenwoodii* (3.3%, n=5). The remaining GG accounted for 11.8% (n=18) of the dataset (Figure 2, Table S1).

The genome size varied across *Gardnerella* species, ranging from 1.50 ± 0.02 Mbp (GG9) to 1.69 ± 0.06 Mbp (*G. vaginalis*) ($p < 2.2x10^{-16}$), with an overall average of 1.61 ± 0.09 Mbp (Figure S1). Guanine-Cytosine (GC) content also differed, with GG12 exhibiting the lowest (38.1 ± 0.1%) and GG14 the highest (45.4 ± 0.1%) ($p < 2.2x10^{-16}$) (Figure S2). Additionally, the number of coding sequences (CDS) ranged from 1,124 ± 0.7 (GG9) to 1,287 ± 61 (*G. vaginalis*) ($p < 6.8x10^{-10}$) (Figure S3).

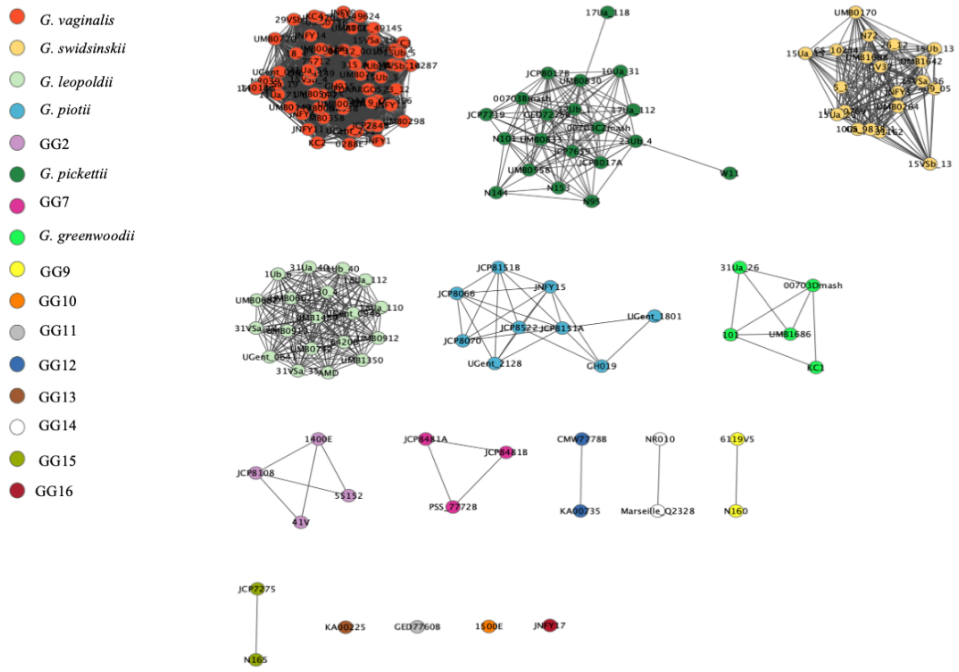


Figure 1. ANI- and dDDH-based network analysis of *Gardnerella* genomes. Each node represents a *Gardnerella* isolate, and node colors correspond to different species. Nodes connected by edges indicate clusters with ANI and dDDH values exceeding 96% and 70%, respectively.

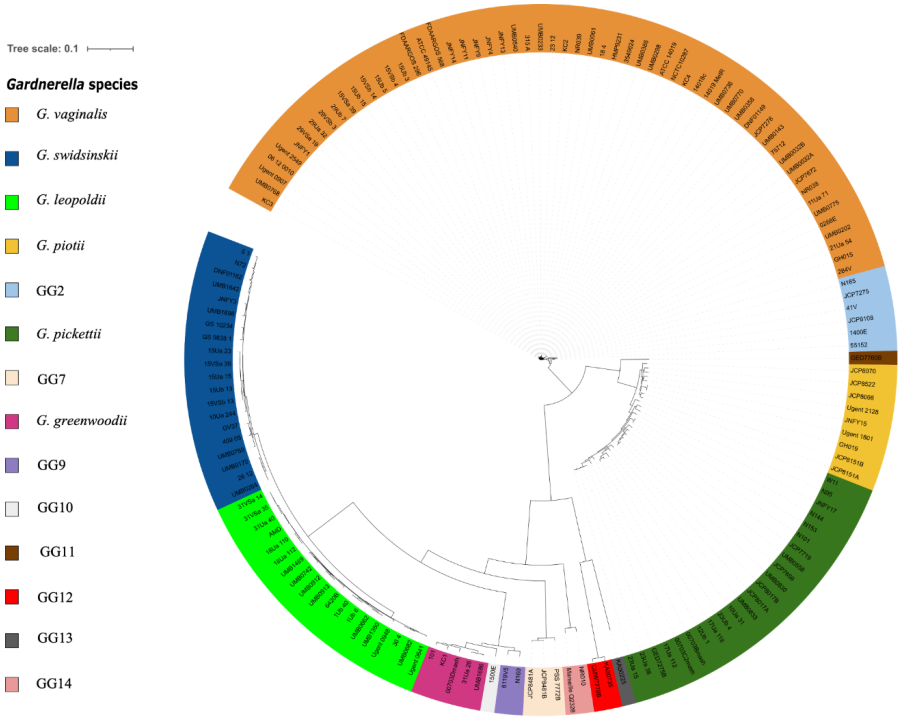


Figure 2. Core genome maximum likelihood phylogeny of *Gardnerella* genomes based on the alignment of the concatenation of amino acid sequences from the 92 single-copy core genes using anvi'o.

Genomic Plasticity in *Gardnerella*: Evidence of Open Pangenomes

The pangenome of *Gardnerella* genus encompasses a total of 4824 gene clusters, categorized into 505 core genes (present in 100% of the genomes), 316 soft core genes (present in > 95% of genomes), 2584 dispensable genes (present in more than two genomes but less than 95%), and 1419 strain-specific genes (singletons). The core genome specifically includes 92 single-copy core genes shared across all 153 assemblies (Table S2). Approximately 10% of the pangenome is represented by the core genome, while the remaining 90% comprises accessory genome, including soft core, dispensable and singleton genes.

Variations in sample sizes across *Gardnerella* species can introduce bias when determining whether a pangenome is open or closed. To address this, we conducted species-specific

pangenome analyses for *G. vaginalis*, *G. leopoldii*, *G. swidsinskii*, and *G. pickettii*. These analyses confirmed that all four species exhibit open pangenomes (Figure S4).

Based on core phylogeny and SNP matrix analysis, we identified identical clonal strains present in distinct biological samples from the same donor, specifically in urine and vaginal swabs, across *G. vaginalis*, *G. leopoldii, G. piotii,* and *G. swidsinskii* species (Table S3) (Figure S5A-D). Notably, a *G. pickettii* clone differing by only one SNP was identified in donors with different health statuses from separate studies. Strain GED7275B was isolated from the vaginal swab of an asymptomatic woman, while strain c17Ua_112^{T} was recovered from the urine of a woman diagnosed with overactive bladder.

Following the identification of the *Gardnerella* core and accessory genomes, gene clusters underwent functional annotation. Out of the total 4841 gene clusters, only 40% were associated with known Clusters of Orthologous Groups of proteins (COG functions), and 26% were linked with Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO).

Functional Insights into Core and Accessory Genomes of *Gardnerella* Species

Metabolic functions predominated within both the core and accessory genomes of *Gardnerella*. In the core genome, proteins were primarily linked to translation, ribosomal structure, and biogenesis (19.6%), followed by carbohydrate (9.8%) and amino acid (9.7%) transport and metabolism. In the accessory genome, the most prevalent COG category was carbohydrate transport and metabolism (G), accounting for 9.7%, with categories J (translation, ribosomal structure, and biogenesis) and L (replication, recombination, and repair) closely following at 7.3% each.

Further investigation into the functional categories of proteins encoded by isolates from different *Gardnerella* species revealed distinct clustering patterns. Non-metric multidimensional scaling (NMDS) analysis revealed a clear separation between *G. vaginalis* and GG2, as compared to other *Gardnerella* species (Figure 3).

The dissimilarity among species was explored through Similarity Percentage (SIMPER) analysis, which identifies the taxa that contribute most to the observed differences among sample groups. Along with *G. vaginalis* and GG2, *G. piotii* and *G. pickettii* also exhibited distinct grouping, separate from *G. leopoldii, G. swidsinskii*, and other GG (Figure 3). This differentiation was partly driven by the lower content of proteins linked to coenzyme transport and metabolism in *G. piotii* and *G. pickettii*, with a mean dissimilarity of 9.5 ± 1.4% (Figure 4A).

The most prominent distinction between *G. vaginalis* and GG2, and the other *Gardnerella* species and GG was the proportion of proteins related to carbohydrate transport and metabolism, accounting for a mean dissimilarity of 32.1 ± 3.2% (Figure 4A). Additionally, the proportion of proteins associated with translation, ribosomal structure, and biogenesis was notably lower in *G. vaginalis* and GG2 compared to the other species (Figure 4B).

Functional enrichment analysis revealed 13 KEGG modules and 3 COG pathways that were enriched in specific *Gardnerella* species (Table S4). Purine degradation, tryptophan, and asparagine biosynthesis were enriched in *G. leopoldii* and *G. swidsinskii*. Thiamine and molybdopterin biosynthesis were enriched in *G. leopoldii*, *G. swidsinskii*, and *G. vaginalis*, with only *G. leopoldii* and *G. swidsinskii* exhibiting molybdopterin adenylyltransferases (Table S4). The glucuronate and pentose phosphate pathways were enriched in *G. vaginalis*. Fatty acid biosynthesis, galactose degradation, and the non-phosphorylated Entner-Doudoroff pathway were enriched in *G. vaginalis*, *G. piotii*, and *G. pickettii*. These three species also shared enriched in functions related to lysine and menaquinone biosynthesis, keratan sulfate and methionine degradation, and glycolysis.

Notably, methionine degradation was enriched in *G. swidsinskii*, as well as *G. vaginalis*, *G. pickettii*, and *G. piotii*, due to the presence of 14 gene clusters encoding DNA-cytosine methylases. Of note, the functional enrichment analysis did not reveal significant differentiation based on isolation source or host disease associations of the different *Gardnerella* strains.

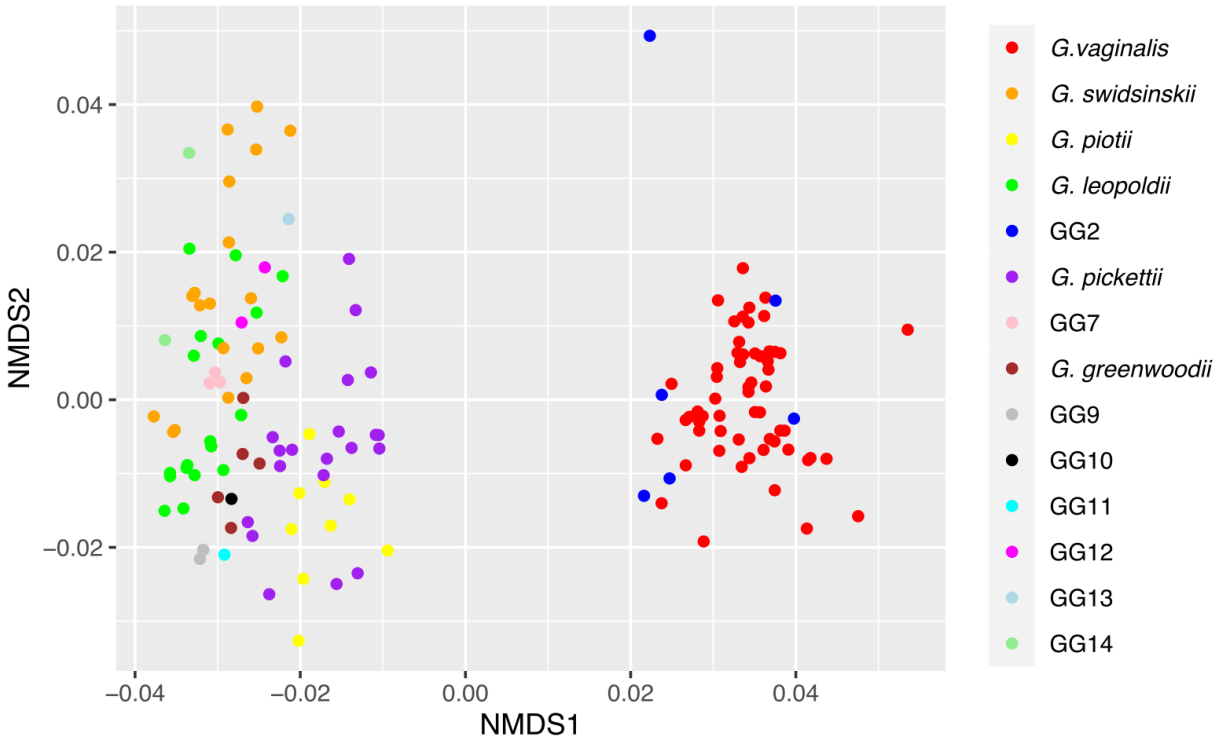


Figure 3. NMDS plot of *Gardnerella* spp. based on the distribution of COG categories.

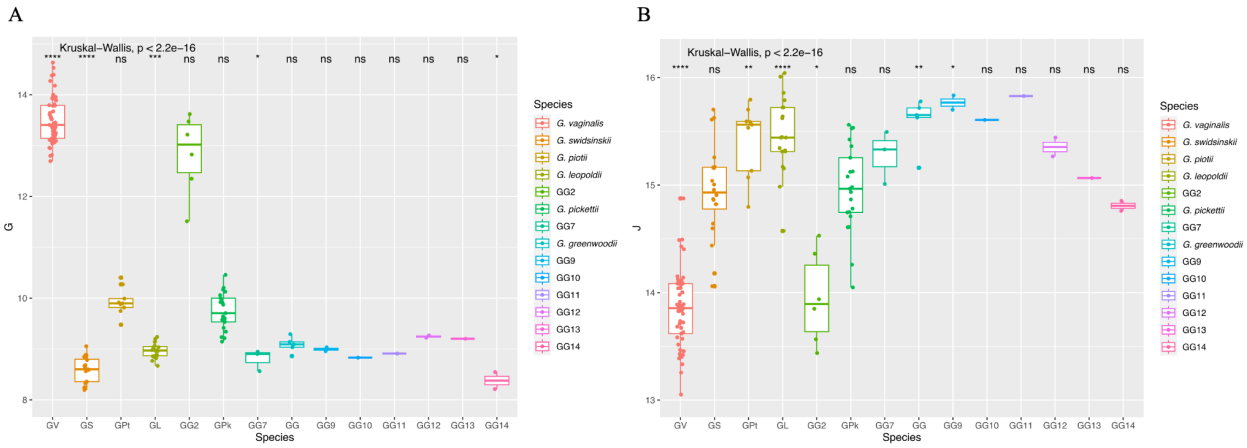


Figure 4. Differential abundance of key COG categories driving NMDS separation among *Gardnerella* spp. based on SIMPER analysis: **A)** Carbohydrate transport and metabolism (category G) and **B)** translation, ribosomal structure and biogenesis (category J).

Prevalence of *lsaC* Gene and Species Clustering based on Virulence Factor Profiles

We identified six antibiotic resistance genes (ARGs) in the pangenome: *aph(3')-Ia* (encoding resistance to aminoglycosides), *tetM* (tetracyclines), *mefA* (macrolides), *msrD* (macrolides and streptogramin B), and *ermX* (macrolides, lincosamides, and streptogramins) and *lsaC* (macrolides, lincosamides, streptogramins A, and pleuromutilins) (Table S5). Among these, *lsaC*-like gene was most prevalent, present in 120 out of 153 genomes across all *Gardnerella* species, except for GG7 and GG13. The *mefA*-like and *msrD*-like genes were detected in all *Gardnerella* species (17 and 27 genomes, respectively), except GG9, GG13, and GG14. The *ermX* gene was found in only six strains (4 *G. vaginalis,* 1 *G. piotii*, and 1 *G. swidsinskii*), all isolated from BV-positive samples. The *tetM*-like gene was identified in 19 *Gardnerella* strains (8 *G. vaginalis*, 4 *G. pickettii*, 2 *G. swidsinskii*, 2 GG2, 1 *G. piotii*, 1 GG12, and 1 GG14), isolated from humans with different health conditions (pregnant woman, BV-positive women, diagnosed with overactive bladder).

We detected 117 different virulence-associated functions distributed across 261 gene clusters in the *Gardnerella* pangenome (Table S6). The clustering of *Gardnerella* genomes based on the presence/absence of virulence gene clusters mirrored the core genome phylogeny, pointing that virulence functions are species-specific within the genus (Figure S6). Two main clades emerged in the species clustering analysis: one composed of *G. vaginalis*, *G. piotii*, GG2, *G. pickettii* and GG11, and the other of the remaining *Gardnerella* species/GG. The differences between these two clusters were largely attributed to distinct virulence factor profiles, particularly those associated to adhesion, mucin degradation and immune evasion (Figure S6).

A more detailed representation of these virulence factors is provided in Figure 5. Related to immune evasion, the protein Lysophospholipase L1 (COG2755), was detected exclusively in *G. piotii*, *G. pickettii* and GG14. Similarly, dTDP-4-dehydrorhamnose reductase was detected only in *G. vaginalis, G. piotii, G. pickettii,* and GG2. Moreover, allantoin utilization, mediated by allantoinase AIIB (K01466) and allantoin permease (K10975), was identified only in *G. leopoldii*, *G. swidsinskii*, GG13, and GG14. The vaginolysin (K11031), a thiol-activated cytolysin and member of the pore-forming toxin family, was detected in all isolates of *G. leopoldii*, *G. swidsinskii*, *G. greenwoodii*, GG9, GG10, and GG12-GG14, and in most of *G. vaginalis* (98.3%) and *G. pickettii* (82.61%), and more than half of GG2 (66.7%) isolates (Figure 5). The distribution was observed across strains isolated from humans with different health conditions. Proteins associated with mucin degradation, such as the α-/β-

galactosidases LacZ (COG3250), GalA (COG3345), and endo-1,4-β-mannosidase (COG3934) were exclusive to *G. vaginalis* and GG2, while α-L-fucosidases were detected in *G. vaginalis*, GG2 and GG12. The α-mannosidase MngB protein was detected in all *G. vaginalis*, *G. pickettii*, *G. piotii*, GG2, and GG11 isolates, as well as in 66.7% of *G. swidsinskii* isolates. Sialidase NanH1 was detected in all *G. vaginalis*, *G. piotii*, *G. pickettii, G. greenwoodii,* GG2, GG9-GG11 and GG14 isolates, while NanH2 was present in 66.7% of *G. piotii* and 8.7% of *G. pickettii* (n=2/23) isolates, and in the unique GG11 isolate. The NanH3 was detected in all *G. piotii* and *G. pickettii*, and 18.9% of *G. vaginalis* isolates. Additionally, an uncharacterized sialidase was detected in three *G. vaginalis* isolates and no sialidases were identified in *G.leopoldii, G. swidsinskii,* GG7, GG12 and GG13.

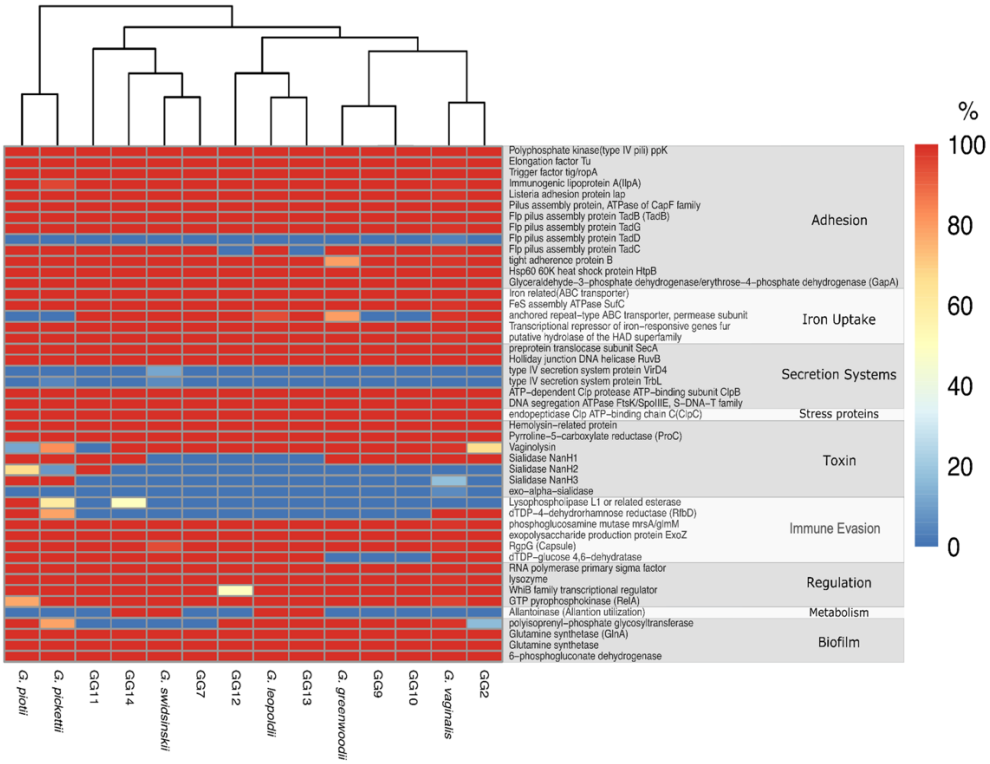


Figure 5. Heatmap showing the percentage of the most relevant VFs among *Gardnerella* genomes. The dendrogram represents the clustering of isolates based on Euclidean distance. The percentage of virulence genes is color-coded, with red (100%) indicating full presence and blue (0%) indicating absence. Intermediate prevalence is represented by shades of orange, yellow, and light blue.

Finally, four secondary metabolite biosynthetic gene clusters (BGCs) were identified across the 153 *Gardnerella* genomes (Table S7). All 18 *G. swidsinskii* isolates harbored a BGC encoding a ribosomally synthesized and post-translationally modified peptide product (RiPP), and approximately 83% (n=15/18) carried a BGC encoding a class II lanthipeptide. A BGC encoding a type III polyketide synthase was detected in all GG12 (n=2) and in most of *G. vaginalis* and *G. pickettii* isolates (93.1% and 82.6%, n=54/58 and n=19/23, respectively). Moreover, a BGC encoding a linear azol(in)-containing peptide was detected in *G. vaginalis* (3.4%, n=2/58) and *G. swidsinskii* (5.6%, n=1/18).

Discussion

Genomic Diversity and Species Delineation

The pathogenic potential of *Gardnerella* spp. has long been debated due to a lack of taxonomic resolution, previously grouping distinct species together. Advances in genomic analyses, including whole-genome sequencing and phylogenomics, now enabled a more detailed understanding of *Gardnerella*, revealing key differences in host-adaptations and pathogenic traits (19, 20, 24). Our findings highlight that while ANI and dDDH offer insight into genetic relatedness, they are insufficient for accurate delineation of *Gardnerella* species, underscoring the importance of core genome phylogenetic analysis.

Instead, core genome phylogeny provides a more robust framework for classification, while pangenome analysis highlights high intraspecies variability in genome size, coding sequence content, and metabolic capabilities.

The detection of the same *Gardnerella* clones in different biological samples from the same donor suggests that these species can colonize multiple anatomical sites within the urogenital tract and persist over time (28). The presence of *G. pickettii* clones in both asymptomatic women and those diagnosed with overactive bladder suggests a complex relationship between *Gardnerella* colonization and urogenital health. Similarly, the detection of specific *Gardnerella* species (*G. vaginalis, G. piotii, G. leopoldii* and *G. swidsinskii)* in vaginal swabs of women with BV (29) should not be interpreted as a direct causation, as other species, such as *Fannyhessea vaginae*, *Prevotella bivia*, are also associated with BV pathogenesis, emphasizing the polymicrobial nature of this condition (30).

Genome features analysis reveals high intraspecies variability in genome size and CDS, underscoring the genetic diversity within this genus (31). The overall *Gardnerella*

pangenome metrics align with previous findings (3, 19), though minor discrepancies exist due to different pangenome computation tools, parameters, and the use of a larger genome collection, which tends to increase pangenome size while decreasing core genome size in species with open pangenomes (32).

The extensive accessory genome of *Gardnerella*, comprising 89.5% of its pangenome, reflects its high genetic diversity, a characteristic also observed in *Bifidobacterium* species (20). This diversity is largely driven by metabolism-associated genes, which vary significantly among *Gardnerella* species.

Metabolic Specialization and Host AdaptationS

Distinct metabolic adaptations among *Gardnerella* species suggest niche specialization within the human urogenital tract. *G. vaginalis* and GG2 exhibit a high proportion of genes involved in carbohydrates transport and metabolism, particularly for glucuronate, pentose phosphate, and galactose degradation, key energy sources in the vaginal microbiome (33). The ability to metabolize pentose sugars and degrade galactose, a mucin component, may provide an advantage for colonization and persistence. Similarly, *G. piotii* and *G. pickettii* exhibit enrichment in carbohydrate metabolism genes, including those for keratan sulfate degradation, a glycosaminoglycan component of the urothelial glycocalyx (34). This capacity may facilitate bacterial persistence in the urinary tract, potentially linking these species to urogenital disorders such as overactive bladder.

Other features such as the lower abundance of translation, ribosomal structure and biogenesis associated proteins in *G. vaginalis* as well as the higher proportion of coenzyme transport and metabolism associated proteins in *G. leopoldii* and *G. swidsinskii.* These findings corroborate previous findings (21) and suggest that *G. vaginalis* may employ a different regulatory mechanism, while *G. leopoldii* and *G. swidsinskii* exhibit enhanced metabolic capabilities, potentially reflecting adaptations to the urogenital niche or host interactions. Moreover, the GG7, GG12, GG13, and GG14 appear to have a similar proportion of coenzyme transport and metabolism-associated proteins as *G. leopoldii* and *G. swidsinskii*. However, further genomes sequencing is required to validate these findings.

Beyond carbohydrate metabolism, differences in amino acid utilization may further drive species-specific adaptations. The identification of genes involved in L-tryptophan biosynthesis from indole in *G. leopoldii, G. swidsinskii,* GG7, and GG10 suggests a potential role in modulating vaginal health. Recent studies have shown that women with BV exhibit

higher vaginal indole levels (35), indicating that metabolic shifts in the microbiota influence host-microbe interactions. If *Gardnerella* species with tryptophan biosynthesis genes contribute to indole metabolism, they may impact microbial balance, inflammation, or immune modulation in the vaginal ecosystem. Additionally, allantoin utilization genes detected in *G. leopoldii, G. swidsinskii,* GG13, and GG14 suggest an alternative carbon and nitrogen source that may provide a competitive advantage to obtain carbon and nitrogen from the urogenital environment (18, 36, 37).

Virulence and Antibiotic Resistance

Gardnerella species harbor genes that may confer resistance to clinically relevant antibiotics, including those used to treat urogenital infections (36, 38, 39). This is particularly concerning because some antibiotic resistance genes were associated with mobile genetic elements (40), further suggesting the potential for horizontal gene transfer, necessitating closer monitoring of *Gardnerella* resistance mechanisms. Of note is the high prevalence of the *lsaC*-like gene, which may encode a protein able to confer resistance to clindamycin, a commonly prescribed antibiotic for BV (36, 39).

The VF analysis was conducted using an *in-house* database based on previous studies on *Gardnerella* (13, 36, 41) (Table S6). The content in VFs appears to be closely associated with *Gardnerella* speciation. The lysophospholipase L1 (COG2755) was detected exclusively in *G. piotii*, *G. pickettii* and GG14, warranting further investigation in the context of *Gardnerella* pathogenicity particularly because phospholipase, namely phospholipase C, have been associated with BV (42). The presence of this VF suggests a possible mechanism for tissue invasion and colonization in the vaginal environment, potentially contributing to the disruption of the vaginal epithelium and the development of BV-associated symptoms (44). Additionally, the dTDP-4-dehydrorhamnose reductase (K23987) and 6-phosphogluconate dehydrogenase (K00033) were also found in *G. vaginalis*, *G. piotii*, *G. pickettii*, and GG2. The dTDP-4-dehydrorhamnose reductase, a key enzyme in the dTDP-L-rhamnose biosynthesis pathway, has been shown to impact either the viability or virulence of numerous bacterial species (44). Similarly, 6-phosphogluconate dehydrogenase (K00033), a surface-located immunogenic lectin protein, can function as an adhesin and has demonstrated significance in bacterial pathogenesis (45, 46).

Previously identified as an accessory gene (19), vaginolysin was widely distributed across *Gardnerella* species in this study, including strains from humans with diverse health

conditions. Vaginolysin, a human-specific cytotoxin, binds to the complement regulatory protein CD59, leading to pore formation in host cell membranes (47). Furthermore, its cytolytic activity releases intracellular contents, potentially providing nutrients that promote the growth of *Gardnerella* and other BV-associated microorganisms (48). However, some strains lacking vaginolysin likely compensate for its absence through alternative mechanisms, suggesting functional redundancy and adaptability within the genus.

Proteins associated with mucin degradation, particularly NanH2 and NanH3, which degrade sialic acids on epithelial cells, were predominantly found in *G. vaginalis, G. piotii,* and *G. pickettii* (13). NanH2 and NanH3, have been previously recognized as key contributors to sialidase activity, directly involved in breaking down sialic acids (13). This enzymatic activity may lead to glycosaminoglycan layer depletion, compromising the vaginal mucosal barrier and potentially increasing the risk of bacterial adherence and infection (13, 49). However, these proteins have also been identified in healthy women, suggesting their presence alone may not be sufficient to drive pathogenesis (13). Although previously BV-associated, NanH1 appears to exhibit basal sialidase activity (13, 49), likely contributing more to general metabolic processes rather than active mucin degradation. Additionally, the uncharacterized sialidase identified in this study, likely corresponding to the recently identified NanH4, may be inactive, as it is present in *G. vaginalis* ATCC 49145, a strain that does not exhibit sialidase activity (15). However, functional sialidase activity assays are necessary to confirm this prediction and determine its potential role in *Gardnerella* physiology. These results suggest that sialidases may provide a competitive advantage in the urogenital environment by enabling *Gardnerella* to utilize a broader range of energy sources.

Moreover, the utilization of allantoin as an intermediate metabolite of purine metabolism is a trait shared by some bacterial species, which can contribute to their virulence (37, 50).

Secondary Metabolite Biosynthetic Gene Clusters

The identification of BGCs across *Gardnerella* genomes suggests that these species may produce bioactive compounds with potential ecological or pathogenic significance. All *G. swidsinskii* isolates harbored a BGC encoding a ribosomally synthesized and post-translationally modified peptide product (RiPP), while most *G. vaginalis* and *G. pickettii* isolates carried a BGC for a type III polyketide synthase. Additionally, a linear azol(in)-containing peptide BGC was detected in a small subset of *G. vaginalis* and *G. swidsinskii*

isolates. The role of these secondary metabolites in *Gardnerella* physiology remains unclear, but their presence suggests potential antimicrobial properties or involvement in host interactions.

Conclusions

This study provides the most detailed genomic characterization of *Gardnerella* genus, uncovering extensive genetic diversity, metabolic versatility, and species-specific adaptations. Approximately 40% of the genes in the *Gardnerella* pangenome are associated with known COG functions, while the genus's extensive accessory genome (comprising 90% of the total pangenome) highlights its significant metabolic versatility, which may contribute to its dual roles in both health and disease. NMDS analysis further differentiated *G. vaginalis* and GG2 from other *Gardnerella* species, highlighting their unique genetic traits, particularly the higher proportion of proteins involved in carbohydrate transport and metabolism. This suggests that these species may have specialized metabolic pathways that enable them to exploit specific carbohydrate sources in the urogenital environment, potentially influencing their survival and colonization, which could play a role in their persistence and pathogenicity. Understanding the metabolic profiles of these species could provide insights into their ecological niches and offer new avenues for targeting species-specific therapeutic strategies. Our analysis identified species-specific virulence factors, such as allantoin utilization in *G. leopoldii* and *G. swidsinskii*, which may enable these species to thrive in the urogenital tract. Importantly, they have the potential to offer novel therapeutic or diagnostic targets for conditions like bacterial vaginosis. The presence of antimicrobial resistance genes (ARGs), such as *lsa*(C), raises concerns about the effectiveness of current treatments and the risk of resistance spread. Moreover, the detection of identical *Gardnerella* clones in both urine and vaginal samples suggests the possibility of persistent colonization across different urogenital sites. This may indicate the capacity for *Gardnerella* species to establish stable, long-term associations within the host microbiota, which could influence the dynamics of *Gardnerella* in both health and disease. However, there is currently no direct evidence linking these specific clones to the onset or recurrence of bacterial vaginosis (BV), warranting further investigation into their potential role in disease persistence or recurrent infections.

The lack of comprehensive host health data impacts the interpretation of pangenome findings. Host health information is crucial for understanding the intricate interactions between *Gardnerella* and the urogenital tract. Recent research has shown that Gardn*erella*

exposures can alter bladder gene expression, affecting inflammation, immunity, and epithelial turnover (51, 52). These host responses may determine whether *Gardnerella* colonization leads to symptomatic conditions or persists in a commensal state.

Future research should prioritize the complex ecological dynamics between *Gardnerella* species and other microorganisms within the urogenital microbiome. A deeper understanding of the pathogenic mechanisms of different *Gardnerella* species across various urogenital conditions will be essential for the development of targeted, species-specific diagnostic and therapeutic strategies. Additionally, sequencing more *Gardnerella* genomes, particularly from under-studied species, and investigating host-pathogen interactions at the molecular level. Furthermore, exploring the role of *Gardnerella* in increasing susceptibility to other urogenital pathogens will provide deeper insights into its clinical significance.

Materials and Methods

Bacterial Collection

Thirty-three *Gardnerella* strains were isolated from urine and vaginal swabs of eight asymptomatic women (one postmenopausal) and three women diagnosed with OAB (one postmenopausal). Participants provided midstream urine (0.1 mL) and vaginal swabs (28). Samples were processed as previously described, and preliminary identification was based on *cpn60* sequencing (28), resulting in *G. vaginalis* (n=12), *G. pickettii* (n=7), *G. greenwoodii* (n=1), *G. leopoldii* (n=7), *G. swidsinskii* (n=6). Genomic DNA was extracted as previously described (2), and the strains were subsequently subjected to whole-genome sequencing.

DNA Extraction and Whole-Genome Sequencing

Isolates were grown on 30 ml of Brain Heart Infusion broth (Frilabo, Portugal) supplemented with 1% glucose, 1% starch, 0.5% yeast extract and 1% gelatin, and 5% sheep blood (Thermo Fisher Scientific, USA). Cultures were incubated at 37 °C in 5% CO_2 for 48 h. Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega, USA) with the following modifications: 30 ml of culture was centrifuged for 10 min at 5000 g to *pellet* cells, washed with PBS, and treated with lysozyme for 20 min at 37°C, followed

by addition of 10% SDS and proteinase K under the same conditions. DNA concentration was measured using the Qubit™ dsDNA High Sensitivity Assay Kit (Invitrogen™, Thermo Fisher Scientific, USA) and integrity confirmed on a 0.8% agarose gel.

Whole-genome sequencing was performed by Eurofins Genomics (Constance, Germany) using the NEBNext Ultra II FS DNA library prep kit (England Biolabs, San Diego, CA, USA) and sequencing on a NovaSeq 6000 instrument (Illumina) producing 2x150-bp paired-end reads. Reads were trimmed by Trim Galore 0.6.6 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) in Bacterial and Viral Bioinformatics Resource Center (https://www.bv-brc.org/) (53) and quality checked using FastQC 0.11.9 (54). *De novo* assembly was performed using Unicycler v0.4.8 (55) and assembly quality assessed by QUAST v5.0.2 (56). Genome annotation was generated through the NCBI Prokaryotic Genome Annotation Pipeline (57).

Gardnerella Genomes Dataset

Genome sequences of *Gardnerella* strains were downloaded from the RefSeq NCBI public database (as of May of 2023). After performing quality control to remove low-quality, incomplete assemblies, and duplicated genomes, a total of 120 high-quality genomes were retained for downstream analysis (58). These genomes were annotated with Prokka v1.13 to ensure comparable quality standards (59). The 33 genomes sequenced in this study were added to this collection, resulting in a final dataset of 153 genomes.

Pangenome Analysis

The pangenome of the *Gardnerella* genus was generated by identifying homologous genes through Markov clustering with an inflation parameter of 6 in anvi'o v7.1 (Distance: Euclidean; Linkage: Ward) (60). Predicted protein sequences were annotated against the Clusters of Orthologous Groups of proteins (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (61, 62). For each genome, the percentage of proteins in each COG category was calculated, and the distribution of the proportional abundance of each COG category was compared between *Gardnerella* species using the Kruskal-Wallis test. Multiple comparisons between two independent *Gardnerella* species/GG were performed using the unpaired two-sample Wilcoxon test. The structural patterns of COG categories and *Gardnerella* species/GG were visualized using Non-metric Multidimensional Scaling (NMDS) based on Bray-Curtis dissimilarity matrices. Similarity Percentages

(SIMPER) were calculated with 999 permutations to identify which COG categories were responsible for differences between *Gardnerella* spp. (63). Functional enrichment analysis was performed to infer the enrichment scores of metabolic modules and pathways associated within specific genomes. A function or metabolic module was enriched in each group if the adjusted q-value was less than 0.05 and present in at least 50% of the *Gardnerella* strains.

Pangenome analysis of particular *Gardnerella* species (minimum of 10 genomes), namely *G. vaginalis*, *G. leopoldii*, *G. swidsinskii* and *G. pickettii*, was generated with Roary (64). Single Nucleotide Polymorphisms (SNPs) were extracted from the core genome alignment of each *Gardnerella* spp. using SNP-sites, and the SNP-alignment was used to generate a phylogenetic tree with RAxML using the general time-reversible model and gamma correction with bootstrap support of 1000 replicates (65, 66). The resulting core genome SNP tree was combined with a presence/ absence matrix of core and accessory genes in each pangenome. Core genome alignments were also used to calculate pairwise SNP distance matrices using snp-dists (https://github.com/tseemann/snp-dists). The α value of Heap's Law was calculated for each *Gardnerella* spp. based to assess genomic plasticity, based on pangenome results from Roary.

Taxonomic and Core Genome Analysis

Average Nucleotide Identity (ANI) and digital DNA-DNA hybridization (dDDH) values were calculated between the 153 *Gardnerella* genomes using fastANI and the Type Strain Genome Server platform, respectively (67, 68). Network analysis was conducted using Cytoscape v3.9.1 with the recommended thresholds for species delineation (70% for dDDH and 96% for ANI) (25–27). A core genome phylogeny was computed by aligning the concatenated amino acid sequences of 92 single-copy core genes using anvi'o (60). FastTree and iTOL v5 were used to compute and visualize the maximum-likelihood core genome tree, respectively (69, 70).

Genome Mining of Antibiotic Resistance, Virulence and Secondary Metabolites

Antibiotic resistance genes were detected using the Comprehensive Antimicrobial Resistance Database (CARD) and the AMRFinder v3.10.45, applying a minimum identity threshold of 90% and 100% coverage (71, 72). A database of virulence factors previously described in *Gardnerella* was used for genome mining against the *Gardnerella* genus

pangenome (13, 36, 41). The proportion of virulence factors represented in each *Gardnerella* species/GG was visualized as a heatmap, generated using the heatmap library with Euclidean distances and complete clustering (73). Secondary metabolite biosynthetic gene clusters were identified using antiSMASH v5 (74).

Author Contributions

PT designed the study, performed the genomic and pangenome analysis, and wrote part of the original draft, MS performed the isolates isolation, genomic DNA extraction, annotation, and species identification, and wrote part of the original draft of the manuscript. FG and TGR contributed to study design, reviewing the manuscript and supervision. LP contributed to reviewing the manuscript, project design, administration, and funding.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Funding Information

This work is 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. P.T. M.S., and F.G. were supported by FCT (project NORTE-01-0145-FEDER-000052, “Healthy&ValorFOOD, SFRH/BD/05038/2020, DL57/2016/CP1346/CT0034, respectively), T.G.R. by Unidade de Ciências Biomoleculares Aplicadas (UIDP/QUI/04378/2020), with the financial support of the FCT/ MCTES through national funds.

Ethical Approval

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.

Supplementary Material

Fig. S1 – S6 and **Table S1 – S7**: Supplementary information regarding *Gardnerella* genome data used in the current study.

References

1. 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.

2. Gardner HL, Dukes CD. 1955. *Haemophilus vaginalis* vaginitis. A newly defined specific infection previously classified “nonspecific” vaginitis. Am J Obstet Gynecol 69:962–976.

3. 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 12;6(3):e00154-21.

4. Sousa M, Ksiezarek M, Perovic SU, Antunes-Lopes T, Grosso F, Ribeiro TG, Peixe L. 2023. *Gardnerella pickettii* sp. nov. (formerly *Gardnerella* genomic species 3) and *Gardnerella greenwoodii* sp. nov. (formerly *Gardnerella* genomic species 8) isolated from female urinary microbiome. Int J Syst Evol Microbiol 73:1–10.

5. Javed A, Parvaiz F, Manzoor S. 2019. Bacterial Vaginosis: An insight into the prevalence, alternative regimen treatments and it’s associated resistance patterns. Microb Pathog 127:21–30.

6. Bilardi JE, Walker S, Temple-Smith M, McNair R, Mooney-Somers J, Bellhouse C, Fairley CK, Chen MY, Bradshaw C. 2013. The burden of bacterial vaginosis: Women’s experience of the physical, emotional, sexual and social impact of living with recurrent bacterial vaginosis. PLoS One 11;8(9):e74378.

7. Giakoumelou S, Wheelhouse N, Cuschieri K, Entrican G, Howie SEM, Horne AW. 2016. The role of infection in miscarriage. Hum Reprod Update 22:116–133.

8. Rosca AS, Castro J, Sousa LG V, Cerca N. 2020. *Gardnerella* and vaginal health: the truth is out there. FEMS Microbiol Rev 44:73–105.

9. 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 9:1557.

10. Morrill S, Gilbert NM, Lewis AL. 2020. *Gardnerella vaginalis* as a Cause of Bacterial Vaginosis: Appraisal of the Evidence From in vivo Models. Front Cell Infect Microbiol 24:10:168.

11. Castro J, Alves P, Sousa C, Cereija T, França Â, Jefferson KK, Cerca N. 2015. Using an *in-vitro* biofilm model to assess the virulence potential of Bacterial Vaginosis or non-Bacterial Vaginosis *Gardnerella vaginalis* isolates. Sci Rep 5:1–10.

12. Jung HS, Ehlers MM, Lombaard H, Redelinghuys MJ, Kock MM. 2017. Etiology of bacterial vaginosis and polymicrobial biofilm formation. Crit Rev Microbiol 43:651–667.

13. Robinson LS, Schwebke J, Lewis WG, Lewis AL. 2019. Identification and characterization of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium *Gardnerella vaginalis*. J Biol Chem 294:5230–5245.

14. Garcia EM, Serrano MG, Edupuganti L, Edwards DJ, Buck GA, Jefferson KK. 2021. Sequence comparison of vaginolysin from different *Gardnerella* species. Pathogens 10:1–15.

15. Pelayo P, Hussain FA, Werlang CA, Wu CM, Woolston BM, Xiang CM, Rutt L, France MT, Tavel J, Ribbeck K, *et al.* 2024. *Prevotella* are major contributors of sialidases in the human vaginal microbiome. Proc Natl Acad Sci 121:e2400341121.

16. Dillard LR, Glass EM, Lewis AL, Thomas-White K, Papin JA. 2022. Metabolic network models of *Gardnerella* pangenome identify interactions in the vaginal environment. mSystems 8:e00689-22.

17. Ksiezarek M, Grosso F, Ribeiro TG, Peixe L. 2022. Genomic diversity of genus *Limosilactobacillus*. Microb Genomics 8(7), mgen000847.

18. Potter RF, Burnham CAD, Dantas G. 2019. *In silico* analysis of *Gardnerella* genomospecies detected in the setting of bacterial vaginosis. Clin Chem 65:1375–1387.

19. Bohr LL, Mortimer TD, Pepperell CS. 2020. Lateral Gene Transfer Shapes Diversity of *Gardnerella* spp. Front Cell Infect Microbiol 10:1–17.

20. Tarracchine C, Lugli GA, Mancabelli L, Milani C, Turroni F, Ventura M. 2021. Assessing the Genomic Variability of *Gardnerella* vaginalis through Comparative Genomic Analyses: Evolutionary and Ecological Implications. Appl Environ Microbiol 87:1–16.

21. Khan S, Vancuren SJ, Hill JE. 2021. A Generalist Lifestyle Allows Rare *Gardnerella* spp. to Persist at Low Levels in the Vaginal Microbiome. Microb Ecol 82:1048–1060.

22. Khan S, Voordouw MJ, Hill JE. 2019. Competition Among *Gardnerella* Subgroups From the Human Vaginal Microbiome. Front Cell Infect Microbiol 9:1–10.

23. Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, Powell E, Janto B, Eutsey R, Luisa Hiller N, *et al.* 2012. Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. J Bacteriol 194:3922–3937.

24. Janulaitiene M, Paliulyte V, Grinceviciene S, Zakareviciene J, Vladisauskiene A, Marcinkute A, Pleckaityte M. 2017. Phenotypic characterization of *Gardnerella vaginalis* subgroups suggests differences in their virulence potential Migle. BMC Infect Dis 17:1–20.

25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski N, Ideker T, 山本隆久, *et al.* 2003. Cytoscape: A Software Environment for Integrated Models. Genome Res 13:426.

26. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60.

27. Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int J Syst Evol Microbiol 68:2386–2392.

28. 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(1):64.

29. Himschoot L, Mulinganya G, Rogier T, Bisimwa G, Kampara F, Kujirakwinja Y, Mongane J, Mubalama I, Callens S, Vaneechoutte M, Cools P. 2024. Prevalence and clinical correlates of *Gardnerella* spp., *Fannyhessea vaginae, Lactobacillus crispatus* and *L. iners* in pregnant women in Bukavu, Democratic Republic of the Congo. Front Cell Infect Microbiol 14:1–16.

30. Castro J, Machado D, Cerca N. 2019. 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.

31. Castro J, Jefferson KK, Cerca N. 2020. Genetic Heterogeneity and Taxonomic Diversity among *Gardnerella* Species. Trends Microbiol 28:202–211.

32. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. 2019. The Ecology and Evolution of Pangenomes. Curr Biol 29:R1094–R1103.

33. Cornejo OE, Hickey RJ, Suzuki H, Forney LJ. 2018. Focusing the diversity of *Gardnerella* vaginalis through the lens of ecotypes. Evol Appl 11:312–324.

34. Jafari N V., Rohn JL. 2022. The urothelium: a multi-faceted barrier against a harsh environment. Mucosal Immunol 15:1127–1142.

35. Sharma J, Gupta K, Schwebke JR, Geisler WM, Muzny CA. 2022. Association of Bacterial Vaginosis with Higher Vaginal Indole Levels. Sex Transm Dis 49:E53–E55.

36. Zhang K, Lu M, Zhu X, Wang K, Jie X, Li T, Dong H, Li R, Zhang F, Gu L. 2022. Antibiotic resistance and pathogenicity assessment of various *Gardnerella* sp. strains in local China. Front Microbiol Sep 26;13:1009798.

37. Yu WL, Ko WC, Cheng KC, Lee CC, Lai CC, Chuang YC. 2008. Comparison of prevalence of virulence factors for *Klebsiella pneumoniae* liver abscesses between isolates with capsular K1/K2 and non-K1/K2 serotypes. Diagn Microbiol Infect Dis 62:1–6.

38. Austin MN, Beigi RH, Meyn LA, Hillier SL. 2005. Microbiologic response to treatment of bacterial vaginosis with topical clindamycin or metronidazole. J Clin Microbiol

43:4492–4497.

39. Beigi RH, Austin MN, Meyn LA, Krohn MA, Hillier SL. 2004. Antimicrobial resistance associated with the treatment of bacterial vaginosis. Am J Obstet Gynecol 191:1124–1129.

40. Forster SC, Liu J, Kumar N, Gulliver EL, Gould JA, Escobar-Zepeda A, Mkandawire T, Pike LJ, Shao Y, Stares MD, *et al.* 2022. Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome. Nat Commun 17;13(1):1445.

41. Yeoman CJ, Yildirim S, Thomas SM, Durkin AS, Torralba M, Sutton G, *et al.* 2010. Comparative genomics of *Gardnerella vaginalis* strains reveals substantial differences in metabolic and virulence potential. PLoS One 26;5(8):e12411.

42. Estrada-Moreno A, Cortés-Sarabia K, Romero LA, Flores-Alfaro E, Castro-Alarcón N, Vences-Velázquez A. 2017. Biotypes, biofilm and phospholipase c production of *Gardnerella vaginalis* associated with normal flora and bacterial vaginosis. Sex Transm Infect 93:A97.1-A97.

43. McGregor JA, French JI, Jones W, Parker R, Patterson E, Draper D. 1992. Association of cervicovaginal infections with increased vaginal fluid phospholipase A2 activity. Am J Obstet Gynecol 167:1588–1594.

44. van der Beek SL, Le Breton Y, Ferenbach AT, Chapman RN, van Aalten DMF, Navratilova I, Boons GJ, Mciver KS, van Sorge NM, Dorfmueller HC. 2015. GacA is essential for Group A *Streptococcus* and defines a new class of monomeric dTDP-4-dehydrorhamnose reductases (RmlD). Mol Microbiol 98:946–962.

45. Fu S, Yuan F, Zhang M, Tan C, Chen H, Bei W. 2012. Cloning, expression and characterization of a cell wall surface protein, 6-phosphogluconate dehydrogenase, of *Haemophilus parasuis*. Res Vet Sci 93:57–62.

46. Daniely D, Portnoi M, Shagan M, Porgador A, Givon-Lavi N, Ling E, Dagan R, Nebenzahl YM. 2006. Pneumococcal 6-phosphogluconate-dehydrogenase, a putative adhesin, induces protective immune response in mice. Clin Exp Immunol 144:254–263.

47. Gelber SE, Aguilar JL, Lewis KLT, Ratner AJ. 2008. Functional and phylogenetic characterization of vaginolysin, the human-specific cytolysin from *Gardnerella*

vaginalis. J Bacteriol 190:3896–3903.

48. LaRocca TJ, Stivison EA, Hod EA, Spitalnik SL, Cowan PJ, Randis TM, Ratner AJ. 2014. Human-specific bacterial pore-forming toxins induce programmed necrosis in erythrocytes. MBio 26;5(5):e01251-14.

49. Lewis WG, Robinson LS, Gilbert NM, Perry JC, Lewis AL. 2013. Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted *Actinobacterium Gardnerella vaginalis*. J Biol Chem 288:12067–12079.

50. Matiasovicova J, Havlickova H, Sisak F, Pilousova L, Rychlik I. 2011. allB, allantoin utilisation and *Salmonella enterica* serovar *Enteritidis* and *Typhimurium* colonisation of poultry and mice. Folia Microbiol (Praha) 56:264–269.

51. Gilbert NM, O'Brien VP, Waller C, Batourina E, Mendelsohn CL, Lewis AL. 2022. *Gardnerella* Exposures Alter Bladder Gene Expression and Augment Uropathogenic *Escherichia coli* Urinary Tract Infection in Mice. Front Cell Infect Microbiol 12:1–14.

52. O'Brien VP, Lewis AL, Gilbert NM. 2021. Bladder Exposure to *Gardnerella* Activates Host Pathways Necessary for *Escherichia coli* Recurrent UTI. Front Cell Infect Microbiol 11:1–13.

53. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, *et al.* 2023. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res 51:D678–D689.

54. Babraham Bioinformatics. 2016. FastQC: A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Retrieved 24 April 2016.

55. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

56. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

57. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, *et al.* 2018. RefSeq: An update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860.

58. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

59. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

60. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, *et al.* 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3–6.

61. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded Microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269.

62. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res 51:D587–D592.

63. Warton DI, Wright ST, Wang Y. 2012. Distance-based multivariate analyses confound location and dispersion effects. Methods Ecol Evol 3:89–101.

64. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693.

65. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb genomics 2:e000056.

66. Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

67. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.

68. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 16;10(1):2182.

69. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-

likelihood trees for large alignments. PLoS One 5:e9490.

70. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.

71. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, *et al.* 2013. The Comprehensive Antibiotic Resistance Database. Antimicrob Agents Chemother 57:3348–3357.

72. Sherry NL, Horan KA, Ballard SA, Gonçalves da Silva A, Gorrie CL, Schultz MB, Stevens K, Valcanis M, Sait ML, Stinear TP, *et al.* 2023. An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. Nat Commun 14:60.

73. Kolde R. 2019. Package “pheatmap”: pretty heatmaps. 1.0.12.

74. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87.

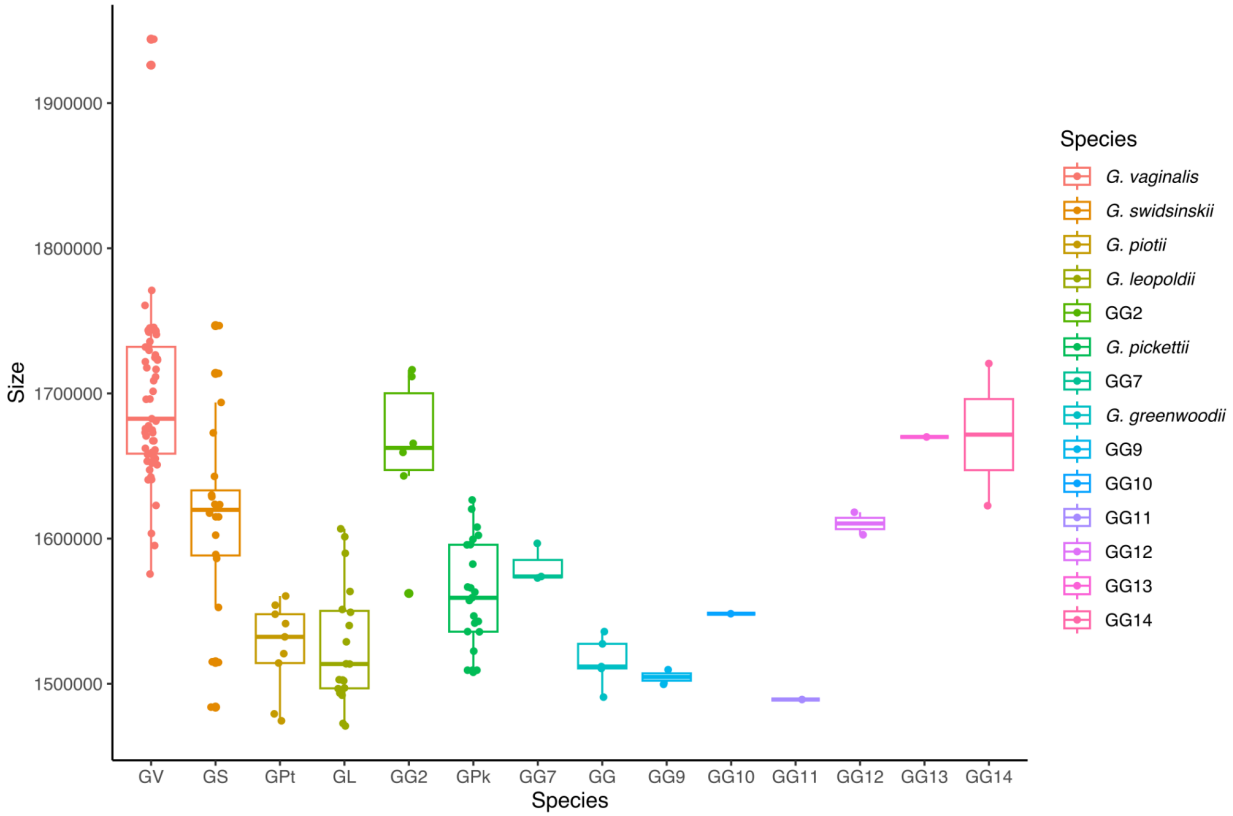


Figure S1. Average genome size across *Gardnerella* species and genomic species, ranging from 1.49 Mbp (*Gardnerella* genomic species 11 – GG11) to 1.69 ± 0.06 Mbp (*G. vaginalis*) (p value < 2.2×10^{-16}), with an overall average of 1.61 ± 0.09 Mbp.

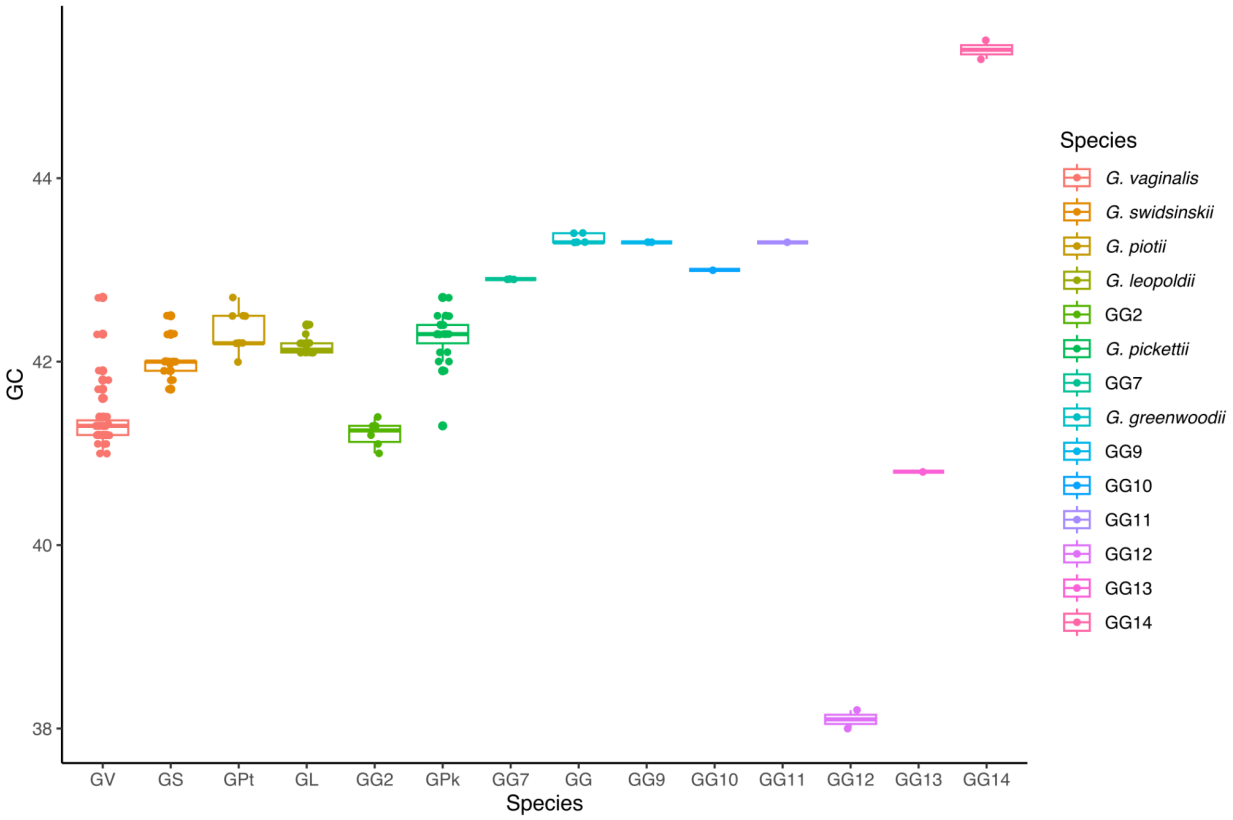


Figure S2. Guanine-Cytosine (GC) content across *Gardnerella* species and genomic species. *Gardnerella* genomic species 12 (GG12) genomes exhibit the lowest GC content at 38.1 ± 0.1%, while *Gardnerella* genomic species 14 (GG14) genomes display the highest at 45.4 ± 0.1% (p value < 2.2×10^{-16}).

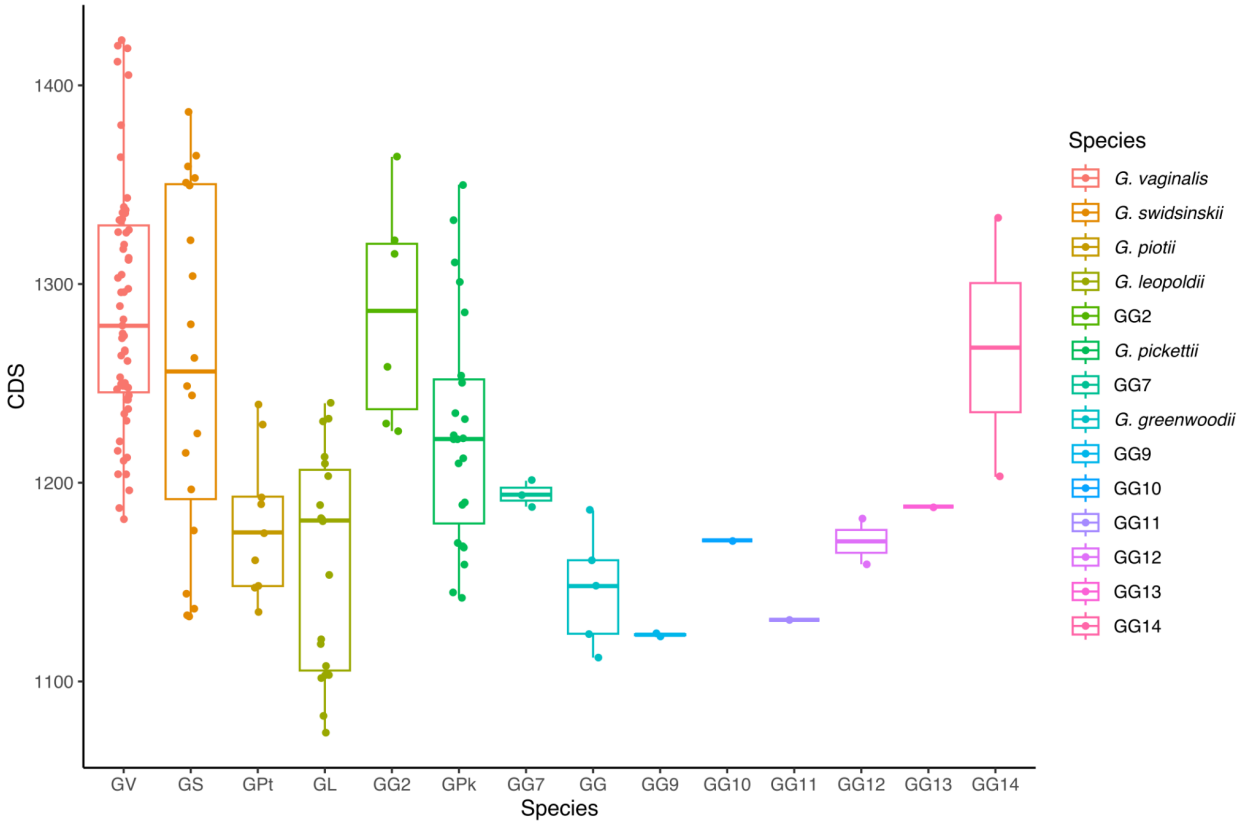


Figure S3. Number of coding sequences (CDS) across *Gardnerella* species and genomic species, ranging from 1124 ± 0.7 (*Gardnerella* genomic species 9 - GG9) to 1287 ± 61 (*G. vaginalis*) (p value < 6.8×10^{-10}).

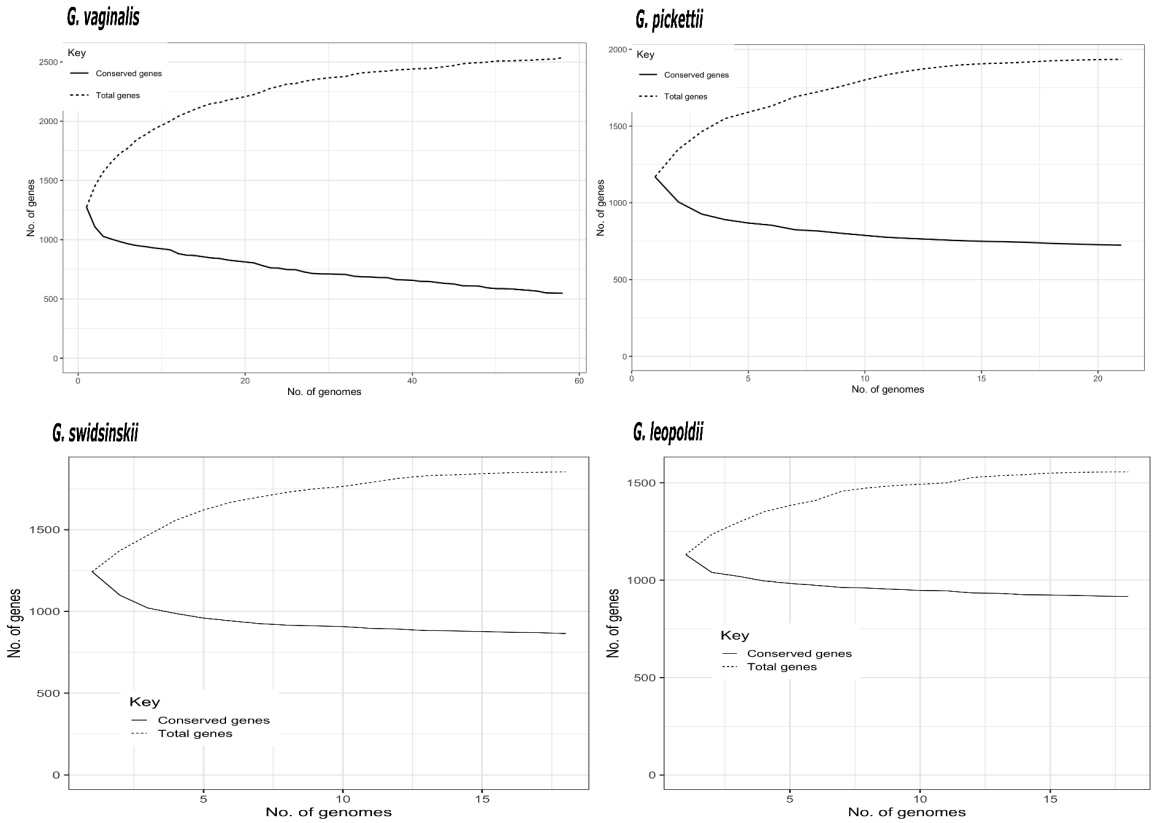
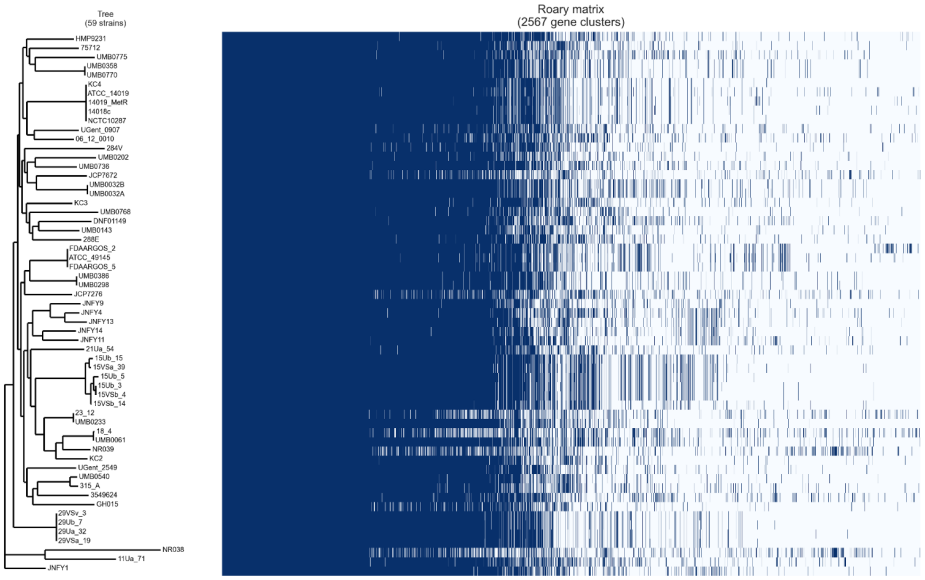
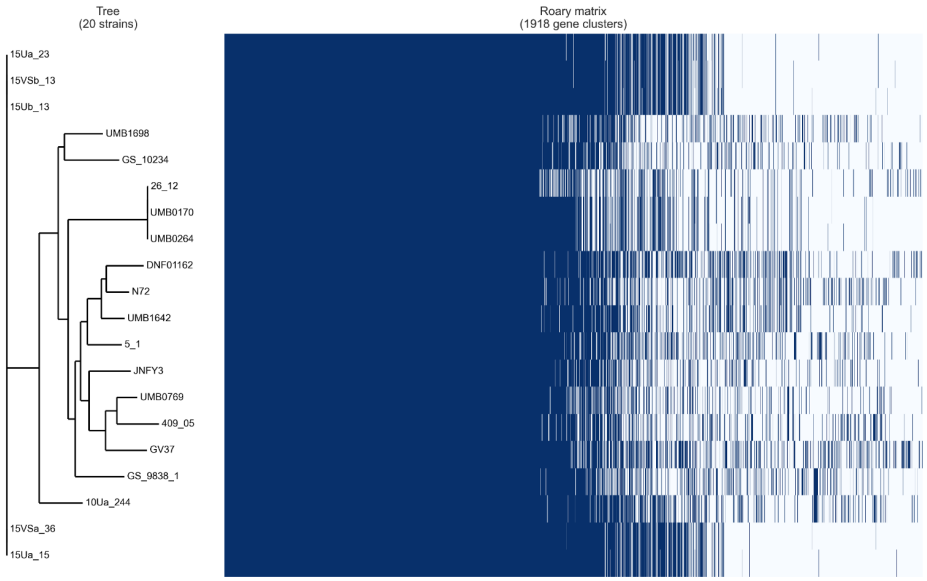


Figure S4. Pangenome analyses for *G. vaginalis, G. leopoldii, G. swidsinskii,* and *G. pickettii* revealed that all these species exhibit open pangenomes.

A


Tree
(59 strains)
HMP9231
75712
UMB0775
UMB0358
UMB0770
KC4
ATCC_14019
14019_MetR
14016c
NCTC10287
UGent_0907
06_12_0010
284V
UMB0202
UMB0736
JCP7672
UMB0032B
UMB0032A
KC3
UMB0766
DNF01149
UMB0143
288E
FDAARGOS_2
ATCC_49145
FDAARGOS_5
UMB0386
UMB0298
JCP7278
JNFY9
JNFY4
JNFY13
JNFY14
JNFY11
21Ua_54
15Ub_15
15VSa_39
15Ub_5
15Ub_3
15VSb_4
15VSb_14
23_12
UMB0233
15_4
UMB0061
NR039
KC2
UGent_2549
UMB0540
315_A
3549624
GH015
29VSv_3
29Ub_7
29Ua_32
29VSa_19
NR038
11Ua_71
JNFY1
Roary matrix
(2567 gene clusters)

B


Tree
(20 strains)
15Ua_23
15VSb_13
15Ub_13
UMB1698
GS_10234
26_12
UMB0170
UMB0264
DNF01162
N72
UMB1642
5_1
JNFY3
UMB0769
409_05
GV37
GS_9838_1
10Ua_244
15VSa_36
15Ua_15
Roary matrix
(1918 gene clusters)

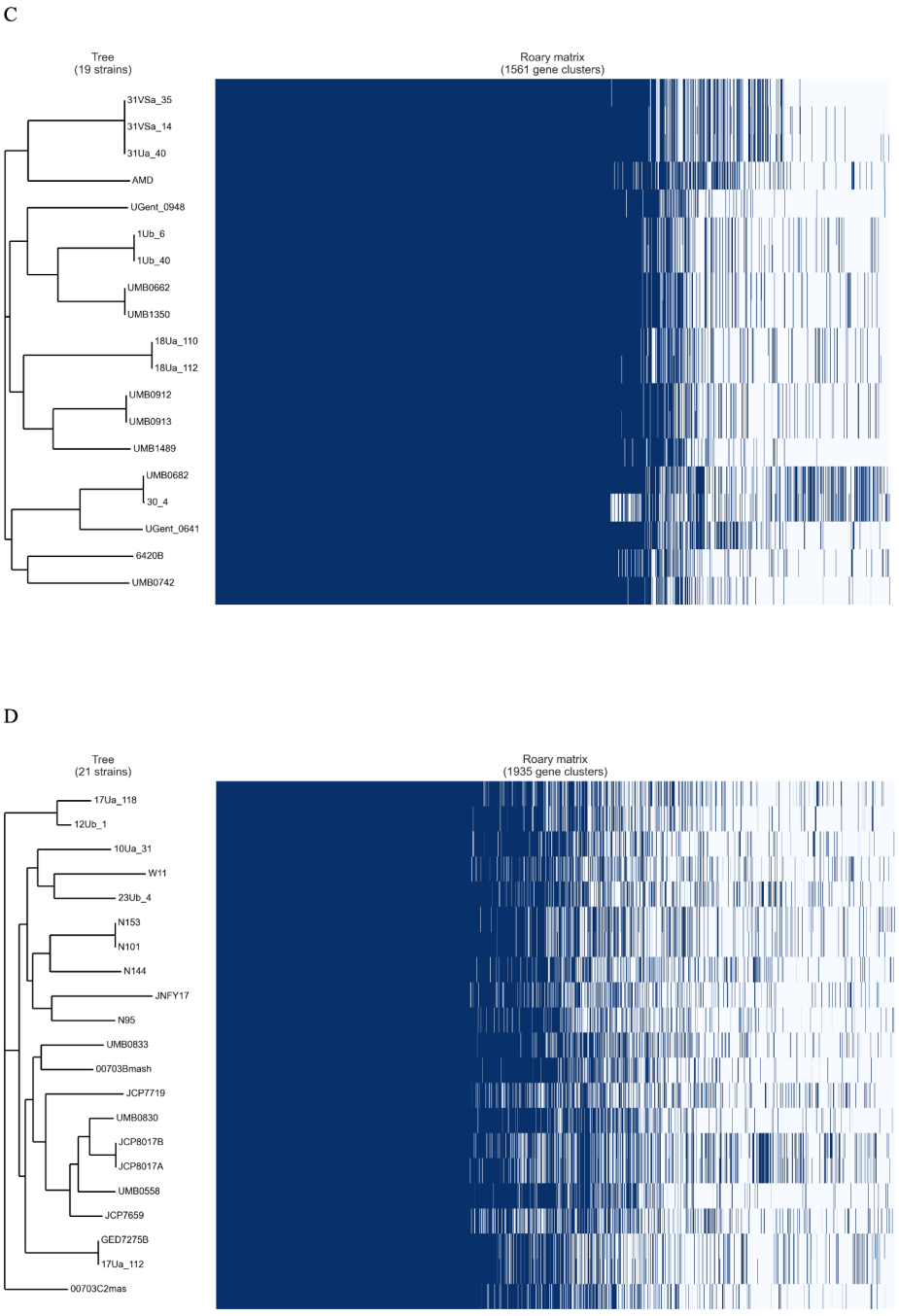


Figure S5. Core phylogeny and SNP matrix analysis, we identified identical clonal strains present in distinct biological samples from the same donor, specifically urine and vaginal swabs, across *G. vaginalis* (A)*, G. swidsinskii* (B)*, G. leopoldii* (C) and *G. pickettii* (D)*,* species.

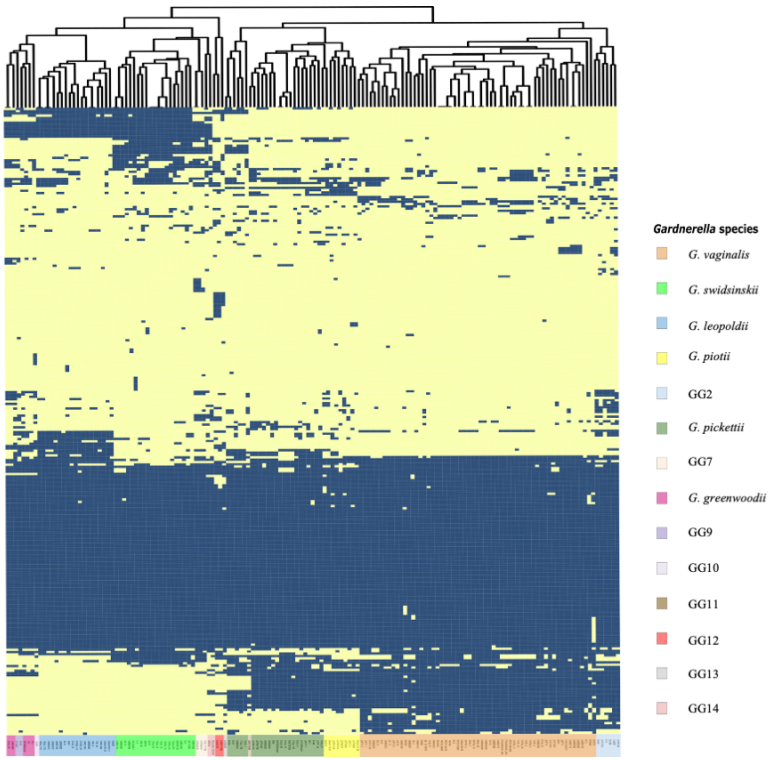


Figure S6. Heatmap representing presence/absence of putative VFs among *Gardnerella* genomes. The dendrogram results from clustering of isolates based on Euclidean distance. Presence or absence of virulence genes is represented as blue and yellow, respectively.

Table S1: Genomic and metadata overview of reclassified *Gardnerella* strains: *Gardnerella* strains from our collection (boldtype) and genomes retrieved from NCBI (available in may 2023).

Species from NCBI	Species after reclassification	Strain name	Assembly Accession	Assembly Level	BioSample Accession	Genome size	Genome GC%	Completeness	Contamination	rRNA	tRNA	CDS	Host	Host disease	Isolation Source
Gardnerella vaginalis	*Gardnerella vaginalis*	**c11Ua_71**	JARFVR000000000	Contig	SAMN33551187	1761830	41.33	100	0.0	1	45	1381	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	**c15Ub_15**	JANUYY000000000	Contig	SAMN27671227	1735765	41.2	100	0.0	3	45	1419	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	**c15Ub_3**	JAODXZ000000000	Contig	SAMN27671224	1744502	41.3	100	0.0	4	45	1420	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	**c15Ub_5**	JANUZA000000000	Contig	SAMN27671225	1740572	41.2	100	0.0	4	45	1412	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	**c15VSa_39**	JANUYX000000000	Contig	SAMN27671228	1743306	41.2	100	0.0	5	45	1423	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	**c15VSb_14**	JANUYU000000000	Contig	SAMN27671231	1722989	41.1	100	0.0	5	45	1380	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	**c15VSb_4**	JANUYW000000000	Contig	SAMN27671229	1731950	41.3	100	0.0	4	45	1405	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	**c21Ua_54**	JANUYR000000000	Contig	SAMN27671235	1677670	41.3	100	0.0	6	45	1337	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	**c29Ua_32**	JANUYN000000000	Contig	SAMN27671239	1673049	41.3	100	0.0	3	45	1336	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	**c29Ub_7**	JANUYM000000000	Contig	SAMN27671240	1662167	41.3	100	0.0	3	45	1326	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	**c29VSa_19**	JANUYL000000000	Contig	SAMN27671241	1670683	41.3	100	0.0	3	45	1332	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	**c29VSb_3**	JANUYK000000000	Contig	SAMN27671242	1658403	41.3	100	0.0	3	45	1318	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	75712	GCF_000263535.1	Contig	SAMN02393774	1672968	41.3	100	0.0	3	45	1249	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	3549624	GCF_001049785.1	Contig	SAMN03801593	1732251	41.4	99.71	0.0	6	45	1320	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	0288E	GCF_000263555.1	Contig	SAMN02393775	1708773	41.2	100	0.0	3	45	1296	Homo Sapiens	BV	VS/ Endometrium
Gardnerella vaginalis	*Gardnerella vaginalis*	06-12-0010	GCF_014857145.1	Contig	SAMN16294983	1717703	41.2	100	0.0	13	45	1248	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	14018c	GCF_004336715.1	Contig	SAMN11037839	1657819	41.3	99.55	0.0	3	45	1237	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	14019_MetR	GCF_001278345.1	Contig	SAMN04014465	1661101	41.3	99.39	0.0	3	45	1231	NA	BV	NA

Gardnerella vaginalis	*Gardnerella vaginalis*	18-4	GCF_001660755.1	Scaffold	SAMN04625602	1622829	42.3	96.2	0.0	5	41	1216	Homo Sapiens	NA	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	23-12	GCF_001660735.1	Scaffold	SAMN04625558	1595201	41.9	98.68	0.45	3	43	1187	Homo Sapiens	NA	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	284V	GCF_000263435.1	Contig	SAMN02393773	1650838	41.2	100	0.0	7	45	1235	Homo Sapiens	BV and *Chlamydia trachomatis*	VS/ Endometrium
Gardnerella vaginalis	*Gardnerella vaginalis*	315-A	GCF_000214315.1	Contig	SAMN00138210	1653275	41.4	100	0.0	3	45	1266	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	ATCC 14019	GCF_000159155.2	Complete Genome	SAMN00001462	1667350	41.4	99.24	0.0	6	45	1211	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	ATCC 49145	GCF_003034925.1	Contig	SAMN08644262	1701439	41.2	100	0.0	5	45	1273	NA	Intermediate BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	DNF01149	GCF_002894105.1	Contig	SAMN05578253	1724694	41.2	100	0.09	6	45	1312	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	FDAARGOS_296	GCF_002206225.1	Contig	SAMN06173309	1770983	41.3	100	1.67	6	45	1343	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	FDAARGOS_568	GCF_003812765.1	Complete Genome	SAMN10163192	1716590	41.3	100	0.0	6	45	1313	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	GH015	GCF_003408745.1	Scaffold	SAMN04446401	1575593	41	93.14	0.0	1	43	1182	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	HMP9231	GCF_000213955.1	Complete Genome	SAMN00100736	1726519	41.2	100	0.0	6	45	1303	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JCP7276	GCF_000414685.1	Scaffold	SAMN02436904	1659589	41	96.83	0.61	6	44	1289	Homo Sapiens	Intermediate BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JCP7672	GCF_000414645.1	Scaffold	SAMN02436831	1603533	41.2	98.18	0.68	6	43	1213	Homo Sapiens	Asymptomatic (pregnant woman)	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JNFY1	GCF_023277725.1	Complete Genome	SAMN21246400	1711437	41.6	100	0.15	6	45	1298	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JNFY11	GCF_023277645.1	Complete Genome	SAMN21246404	1743456	41.8	100	0.15	6	45	1339	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JNFY13	GCF_023277625.1	Complete Genome	SAMN21246405	1682566	41.6	100	0.0	6	45	1282	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JNFY14	GCF_023277605.1	Complete Genome	SAMN21246406	1680963	41.7	100	0.0	6	45	1274	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JNFY4	GCF_023277685.1	Complete Genome	SAMN21246402	1742450	41.6	100	0.15	6	45	1332	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	JNFY9	GCF_023277665.1	Complete Genome	SAMN21246403	1640523	41.4	100	0.15	6	45	1221	Homo Sapiens	BV	VS

Gardnerella vaginalis	*Gardnerella vaginalis*	KC2	GCF_023016185.1	Scaffold	SAMN23424279	1651935	41.3	100	0.0	3	45	1248	Homo Sapiens	BV (pregnant woman)	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	KC3	GCF_023016245.1	Scaffold	SAMN23424280	1721847	41.2	100	0.0	3	45	1305	Homo Sapiens	BV (pregnant woman)	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	KC4	GCF_023016225.1	Contig	SAMN23424281	1657976	41.3	99.55	0.0	3	45	1250	Homo Sapiens	BV (pregnant woman)	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	NCTC10287	GCF_900637625.1	Complete Genome	SAMEA4535760	1667383	41.4	99.55	0.0	6	45	1250	Homo Sapiens	BV (pregnant woman)	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	NR038	GCF_003585655.1	Scaffold	SAMN07490630	1944040	42.7	91.26	0.0	3	42	1267	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	NR039	GCF_003585755.1	Scaffold	SAMN07490631	1926148	42.7	90.87	0.0	2	43	1244	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	UGent 09.07	GCF_003397665.1	Scaffold	SAMN09373175	1723772	41.1	99.57	2.32	3	47	1296	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	UGent 25.49	GCF_003397605.1	Scaffold	SAMN09373179	1658566	41.2	100	0.0	4	45	1242	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0032A	GCF_002862015.1	Scaffold	SAMN08193673	1745481	41.2	100	0.0	7	45	1327	Homo Sapiens	Asymptomatic	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0032B	GCF_002862005.1	Scaffold	SAMN08193675	1745062	41.2	100	0.0	8	45	1326	Homo Sapiens	Asymptomatic (pregnant woman)	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0061	GCF_002861165.1	Scaffold	SAMN08193668	1742219	41.2	100	0.0	6	45	1336	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0143	GCF_013315005.1	Scaffold	SAMN15064064	1647311	41.3	99.55	0.0	3	45	1196	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0202	GCF_013315075.1	Scaffold	SAMN15064060	1655074	41.4	100	0.0	6	45	1204	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0233	GCF_002862045.1	Contig	SAMN08193672	1642442	41.2	100	0.0	3	45	1242	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0298	GCF_002861975.1	Scaffold	SAMN08193676	1676015	41.2	100	0.0	6	45	1261	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0358	GCF_013315085.1	Scaffold	SAMN15064061	1696164	41.2	100	0.0	6	45	1247	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0386	GCF_002861965.1	Scaffold	SAMN08193674	1675681	41.2	100	0.0	6	45	1253	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0540	GCF_013315045.1	Scaffold	SAMN15064062	1640659	41.3	100	0.0	4	45	1204	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0736	GCF_013315025.1	Scaffold	SAMN15064063	1729722	41.1	100	0.0	5	45	1275	Homo Sapiens	Kidney Stone	MSU
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0768	GCF_002884835.1	Contig	SAMN07511408	1674824	41.2	100	0.0	6	45	1264	Homo Sapiens	NA	TUC
Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0770	GCF_002861945.1	Scaffold	SAMN08193677	1695986	41.2	100	0.0	6	45	1279	Homo Sapiens	OAB	MSU

Gardnerella vaginalis	*Gardnerella vaginalis*	UMB0775	GCF_002861925.1	Contig	SAMN08193678	1743554	41.2	100	0.0	5	45	1333	Homo Sapiens	Stress Urinary Incontinence (SUI)	MSU
Gardnerella vaginalis	*Gardnerella* genomic species 2	JCP7275	GCF_000414705.1	Scaffold	SAMN02436832	1562234	41	97.08	1.12	3	38	1230	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella* genomic species 2	N165	GCF_003408785.1	Scaffold	SAMN04446402	1711624	41.4	97.74	0.0	3	44	1364	Homo Sapiens	NA	VS (sex workers)
Gardnerella vaginalis	*Gardnerella* genomic species 2	55152	GCF_000263475.1	Contig	SAMN02393778	1643189	41.3	99.85	0.0	3	45	1226	Homo Sapiens	BV	VS/ Endometrium
Gardnerella vaginalis	*Gardnerella* genomic species 2	1400E	GCF_000263495.1	Contig	SAMN02393779	1716325	41.2	100	0.0	4	44	1315	Homo Sapiens	BV	VS/ Endometrium
Gardnerella vaginalis	*Gardnerella* genomic species 2	41V	GCF_000165635.1	Contig	SAMN02472074	1659370	41.3	100	0.0	3	45	1258	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella* genomic species 2	JCP8108	GCF_000414525.1	Scaffold	SAMN02436830	1665594	41.1	98.11	0.95	6	43	1322	Homo Sapiens	BV	VS
Gardnerella genomic species 3	*Gardnerella pickettii*	**c10Ua_31**	JANUYG010000000	Contig	SAMN27671220	1626615	42.3	100	0.0	3	45	1350	Homo Sapiens	Assymptomatic	MSU
Gardnerella genomic species 3	*Gardnerella pickettii*	**c12Ub_1**	JANUZC000000000	Contig	SAMN27671221	1535783	42.4	100	0.0	6	45	1235	Homo Sapiens	Assymptomatic	MSU
Gardnerella genomic species 3	*Gardnerella pickettii*	**c17Ua_112**	GCF_029226345.1	Contig	SAMN25554339	1509345	42.5	100	0.0	3	45	1222	Homo Sapiens	OAB	MSU
Gardnerella genomic species 3	*Gardnerella pickettii*	**c17Ua_118**	JANUYT000000000	Contig	SAMN27671232	1595696	42.3	100	0.0	6	45	1311	Homo Sapiens	OAB	MSU
Gardnerella genomic species 3	*Gardnerella pickettii*	**c23Ua_15**	JANUYQ000000000	Contig	SAMN27671236	1509345	42.5	100	0.0	3	45	1222	Homo Sapiens	Assymptomatic	MSU
Gardnerella genomic species 3	*Gardnerella pickettii*	**c23Ua_38**	JANUYP000000000	Contig	SAMN27671237	1509345	42.5	100	0.0	3	45	1222	Homo Sapiens	Assymptomatic	MSU
Gardnerella genomic species 3	*Gardnerella pickettii*	**c23Ub_4**	JANUYO000000000	Contig	SAMN27671238	1599514	42.4	100	0.0	6	45	1332	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella pickettii*	JNFY17	GCF_023277565.1	Complete Genome	SAMN21246408	1595814	42.7	100	0.0	5	45	1212	Homo Sapiens	BV	VS

Gardnerella vaginalis	*Gardnerella pickettii*	00703Bmash	GCF_000263615.1	Contig	SAMN02393781	1566055	42.3	100	0.0	3	45	1210	Homo Sapiens	BV and HSV-2	VS
Gardnerella vaginalis	*Gardnerella pickettii*	00703C2mash	GCF_000263515.1	Contig	SAMN02393782	1546682	42.3	100	0.0	4	45	1167	Homo Sapiens	BV and HSV-2	VS
Gardnerella vaginalis	*Gardnerella pickettii*	GED7275B	GCF_001546445.1	Scaffold	SAMN03851014	1507903	42.5	99.55	0.0	3	45	1142	Homo Sapiens	Asymptomatic	VS
Gardnerella vaginalis	*Gardnerella pickettii*	JCP7659	GCF_000414665.1	Scaffold	SAMN02436712	1535941	41.9	97.12	1.8	4	38	1224	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella pickettii*	JCP7719	GCF_000414625.1	Scaffold	SAMN02436711	1563149	42	97.27	0.38	5	35	1254	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella pickettii*	JCP8017A	GCF_000414605.1	Scaffold	SAMN02436912	1607921	42.1	98.71	0.23	5	43	1301	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella pickettii*	JCP8017B	GCF_000414585.1	Scaffold	SAMN02436773	1602251	42	98.56	0.38	5	37	1286	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella pickettii*	N101	GCF_003369895.1	Contig	SAMN03145579	1542981	42.4	100	0.91	3	45	1168	Homo Sapiens	NA	VS (sex workers)
Gardnerella vaginalis	*Gardnerella pickettii*	N144	GCF_003408835.1	Scaffold	SAMN04446400	1582426	41.3	100	0.0	3	45	1232	Homo Sapiens	NA	VS (sex workers)
Gardnerella vaginalis	*Gardnerella pickettii*	N153	GCF_003369935.1	Contig	SAMN03145603	1541756	42.4	99.72	0.0	3	45	1145	Homo Sapiens	NA	VS (sex workers)
Gardnerella vaginalis	*Gardnerella pickettii*	N95	GCF_003369965.1	Contig	SAMN03145504	1522480	42.4	100	0.0	4	45	1159	Homo Sapiens	NA	VS (sex workers)
Gardnerella vaginalis	*Gardnerella pickettii*	UMB0558	GCF_013315115.1	Scaffold	SAMN15064059	1557387	42.3	99.55	0.0	4	45	1170	Homo Sapiens	SUI	MSU
Gardnerella vaginalis	*Gardnerella pickettii*	UMB0830	GCF_002861905.1	Scaffold	SAMN08193679	1559223	42.3	99.92	0.0	3	43	1190	Homo Sapiens	Asymptomatic (pregnant woman)	MSU
Gardnerella vaginalis	*Gardnerella pickettii*	UMB0833	GCF_002861885.1	Scaffold	SAMN08193680	1620322	42.1	100	0.0	5	45	1250	Homo Sapiens	Asymptomatic (pregnant woman)	MSU
Gardnerella vaginalis	*Gardnerella pickettii*	W11	GCF_003369875.1	Contig	SAMN03145604	1566657	42.3	100	0.0	3	45	1189	Homo Sapiens	NA	VS (sex workers)
Gardnerella vaginalis	*Gardnerella piotii*	GH019	GCF_003426565.1	Scaffold	SAMN07358667	1532315	42.7	86.2	0.0	2	45	1193	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella piotii*	JCP8066	GCF_000414565.1	Scaffold	SAMN02436710	1520733	42.2	97.23	0.83	5	39	1175	Homo Sapiens	Asymptomatic	VS
Gardnerella vaginalis	*Gardnerella piotii*	JCP8070	GCF_000414545.1	Scaffold	SAMN02436911	1479254	42.2	98.71	1.98	5	44	1161	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella piotii*	JCP8151A	GCF_000414505.1	Scaffold	SAMN02436772	1560453	42	98.15	1.19	5	39	1229	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella piotii*	JCP8151B	GCF_000414485.1	Scaffold	SAMN02436771	1554137	42.2	98.71	0.76	5	44	1239	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella piotii*	JCP8522	GCF_000414425.1	Scaffold	SAMN02436909	1474487	42.2	96.82	1.53	4	40	1135	Homo Sapiens	BV	VS
Gardnerella piotii	*Gardnerella piotii*	JNFY15	GCF_023277585.1	Complete Genome	SAMN21246407	1541442	42.5	100	0.15	6	45	1148	Homo Sapiens	BV	VS
Gardnerella piotii	*Gardnerella piotii*	UGent 18.01	GCF_003397585.1	Contig	SAMN09373177	1514270	42.5	100	0.0	5	45	1147	Homo Sapiens	BV	VS

Gardnerella vaginalis	*Gardnerella piotii*	UGent 21.28	GCF_003397615.1	Scaffold	SAMN09373178	1547915	42.5	100	0.0	3	45	1189	Homo Sapiens	BV	VS
Gardnerella leopoldii	*Gardnerella leopoldii*	**c1Ub_6**	JANUUT010000000	Contig	SAMN27671218	1502056	42.2	100	0.0	6	45	1182	Homo Sapiens	Assymptomatic	MSU
Gardnerella leopoldii	*Gardnerella leopoldii*	**c1Ub_40**	JAOPFO000000000	Contig	SAMN27671219	1496617	42.2	100	0.0	6	45	1181	Homo Sapiens	Assymptomatic	MSU
Gardnerella leopoldii	*Gardnerella leopoldii*	**c18Ua_110**	JANUYS000000000	Contig	SAMN27671233	1497084	42.2	100	0.0	6	45	1189	Homo Sapiens	OAB	MSU
Gardnerella leopoldii	*Gardnerella leopoldii*	**c18Ua_112**	JAPDSJ000000000	Contig	SAMN27671234	1528899	42.3	100	0.0	6	45	1203	Homo Sapiens	OAB	MSU
Gardnerella leopoldii	*Gardnerella leopoldii*	**c31Ua_40**	JANUYJ000000000	Contig	SAMN27671243	1551185	42.1	100	0.0	6	45	1232	Homo Sapiens	Assymptomatic	MSU
Gardnerella leopoldii	*Gardnerella leopoldii*	**c31VSa_14**	JANUYI000000000	Contig	SAMN27671244	1549240	42.1	100	0.0	6	45	1231	Homo Sapiens	Assymptomatic	VS
Gardnerella leopoldii	*Gardnerella leopoldii*	**c31VSa_35**	JARFVP000000000	Contig	SAMN33551189	1540890	42.13	100	0.0	2	45	1167	Homo Sapiens	Assymptomatic	VS
Gardnerella sp.	*Gardnerella leopoldii*	30-4	GCF_001641215.1	Scaffold	SAMN04625622	1589896	42.4	99.44	0	3	45	1213	Homo Sapiens	NA	U
Gardnerella vaginalis	*Gardnerella leopoldii*	6420B	GCF_000263575.1	Contig	SAMN02393777	1493594	42.2	100	0	3	45	1108	Homo Sapiens	Assymptomatic	VS/ Endometrium
Gardnerella vaginalis	*Gardnerella leopoldii*	AMD	GCF_000176475.1	Contig	SAMN02472072	1606758	42.1	100	0.61	3	45	1240	Homo Sapiens	BV	VS
Gardnerella leopoldii	*Gardnerella leopoldii*	UGent 06.41	GCF_003293675.1	Chromosome	SAMN09373173	1563545	42.1	100	0.0	6	45	1182	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella leopoldii*	UGent 09.48	GCF_003397635.1	Scaffold	SAMN09373176	1470925	42.2	100	0.0	3	45	1103	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella leopoldii*	UMB0662	GCF_013315255.1	Scaffold	SAMN15064053	1502587	42.1	100	0.0	3	45	1102	Homo Sapiens	NA	TUC
Gardnerella vaginalis	*Gardnerella leopoldii*	UMB0682	GCF_002862065.1	Contig	SAMN08193671	1601254	42.1	100	0.0	3	45	1210	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella leopoldii*	UMB0742	GCF_013315135.1	Contig	SAMN15064058	1472679	42.2	100	0.0	3	45	1074	Homo Sapiens	OAB	MSU
Gardnerella vaginalis	*Gardnerella leopoldii*	UMB0912	GCF_002861125.1	Scaffold	SAMN08193669	1513749	42.1	100	0.0	3	45	1121	Homo Sapiens	Asymptomatic (pregnant woman)	MSU
Gardnerella vaginalis	*Gardnerella leopoldii*	UMB0913	GCF_002861145.1	Scaffold	SAMN08193670	1513612	42.1	100	0.0	3	45	1119	Homo Sapiens	Asymptomatic (pregnant woman)	MSU
Gardnerella vaginalis	*Gardnerella leopoldii*	UMB1350	GCF_013315125.1	Scaffold	SAMN15064055	1502805	42.1	100	0.0	3	45	1103	Homo Sapiens	OAB	MSU

Gardnerella vaginalis	*Gardnerella leopoldii*	UMB1489	GCF_013315195.1	Scaffold	SAMN15064056	1492090	42.2	100	0.0	3	45	1083	Homo Sapiens	NA	TUC
Gardnerella swidsinskii	*Gardnerella swidsinskii*	**c10Ua_244**	JARFVS000000000	Contig	SAMN33551186	1719964	42.00	100	0.09	2	45	1380	Homo Sapiens	Assymptomatic	MSU
Gardnerella swidsinskii	*Gardnerella swidsinskii*	**c15Ua_15**	JAPDSI000000000	Contig	SAMN27671222	1630003	42	100	0.09	6	45	1359	Homo Sapiens	Assymptomatic	MSU
Gardnerella swidsinskii	*Gardnerella swidsinskii*	**c15Ua_23**	JANUZB000000000	Contig	SAMN27671223	1623408	42	100	0.09	6	45	1353	Homo Sapiens	Assymptomatic	MSU
Gardnerella swidsinskii	*Gardnerella swidsinskii*	**c15Ub_13**	JANUYZ000000000	Contig	SAMN27671226	1614983	42	100	0.09	6	45	1350	Homo Sapiens	Assymptomatic	MSU
Gardnerella swidsinskii	*Gardnerella swidsinskii*	**c15VSb_13**	JANUYV000000000	Contig	SAMN27671230	1615001	42	100	0.09	6	45	1351	Homo Sapiens	Assymptomatic	VS
Gardnerella swidsinskii	*Gardnerella swidsinskii*	**c15VSa_36**	JARFVQ000000000	Contig	SAMN33551188	1623600	41.97	100	0.09	2	45	1269	Homo Sapiens	Assymptomatic	VS
Gardnerella sp.	*Gardnerella swidsinskii*	26-12	GCF_001660745.1	Scaffold	SAMN04625631	1483907	42.5	99.85	0.91	3	45	1133	Homo Sapiens	NA	MSU
Gardnerella vaginalis	*Gardnerella swidsinskii*	409-05	GCF_000025205.1	Complete Genome	SAMN02299422	1617545	42	99.94	0.0	6	45	1249	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella swidsinskii*	5-1	GCF_000176495.1	Contig	SAMN02472071	1672842	42	99.94	0.0	3	45	1304	Homo Sapiens	Assymptomatic	VS
Gardnerella sp.	*Gardnerella swidsinskii*	DNF01162	GCF_002894125.1	Contig	SAMN05578537	1693779	41.7	100	0.0	5	45	1322	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella swidsinskii*	GS 10234	GCF_003397745.1	Scaffold	SAMN09373171	1589042	41.9	99.94	0.0	3	45	1197	Homo Sapiens	BV	VS
Gardnerella swidsinskii	*Gardnerella swidsinskii*	GS 9838-1	GCF_003397705.1	Contig	SAMN09373170	1622089	41.9	100	0.17	4	45	1244	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella swidsinskii*	GV37	GCF_001953155.1	Complete Genome	SAMN06199463	1746718	41.8	99.94	0.0	6	45	1387	Homo Sapiens	Bacteremia	B
Gardnerella swidsinskii	*Gardnerella swidsinskii*	JNFY3	GCF_023277705.1	Complete Genome	SAMN21246401	1602355	42	100	0.15	6	45	1215	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella swidsinskii*	N72	GCF_003408815.1	Scaffold	SAMN04446399	1642851	41.9	99.55	0.0	3	45	1280	Homo Sapiens	BV	VS (sex workers)
Gardnerella vaginalis	*Gardnerella swidsinskii*	UMB0170	GCF_002884855.1	Scaffold	SAMN07511406	1514739	42.3	100	0.0	6	45	1137	Homo Sapiens	NA	TUC
Gardnerella vaginalis	*Gardnerella swidsinskii*	UMB0264	GCF_002884875.1	Scaffold	SAMN07511407	1515133	42.3	100	0.0	6	45	1133	Homo Sapiens	NA	TUC
Gardnerella vaginalis	*Gardnerella swidsinskii*	UMB0769	GCF_013315215.1	Contig	SAMN15064054	1552587	42	99.94	0.0	3	45	1144	Homo Sapiens	OAB	MSU

Gardnerella vaginalis	*Gardnerella swidsinskii*	UMB1642	GCF_002884795.1	Contig	SAMN07511411	1628789	41.8	100	0.0	6	45	1225	Homo Sapiens	Asymptomatic	MSU
Gardnerella vaginalis	*Gardnerella swidsinskii*	UMB1698	GCF_013315145.1	Scaffold	SAMN15064057	1586334	41.9	100	0.0	3	45	1176	Homo Sapiens	Diabetes	MSU
Gardnerella vaginalis	*Gardnerella* genomic species 7	JCP8481A	GCF_000414465.1	Scaffold	SAMN02436910	1572875	42.9	97.84	0.61	5	44	1194	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella* genomic species 7	JCP8481B	GCF_000414445.1	Scaffold	SAMN02436829	1573879	42.9	99.04	0.45	5	43	1201	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella* genomic species 7	PSS_7772B	GCF_001546485.1	Scaffold	SAMN03851016	1596709	42.9	99.39	0.45	3	45	1188	Homo Sapiens	Asymptomatic (pregnant woman)	MSU
Gardnerella genomic species 8	*Gardnerella greenwoodii*	**c31Ua_26**	GCF_029207615.1	Contig	SAMN25596848	1511834	43.3	98.94	0.61	3	45	1186	Homo Sapiens	Assymptomatic	MSU
Gardnerella vaginalis	*Gardnerella greenwoodii*	101	GCF_000165615.1	Contig	SAMN02472073	1527495	43.4	99.39	0.45	24	47	1161	Homo Sapiens	BV	VS
Gardnerella vaginalis	*Gardnerella greenwoodii*	00703Dmash	GCF_000263635.1	Contig	SAMN02393783	1490797	43.4	99.39	0.45	3	45	1124	Homo Sapiens	BV and HSV-2	VS
Gardnerella vaginalis	*Gardnerella greenwoodii*	KC1	GCF_023016205.1	Scaffold	SAMN23424278	1535938	43.3	98.94	0.61	3	45	1148	Homo Sapiens	BV (pregnant woman)	VS
Gardnerella vaginalis	*Gardnerella greenwoodii*	UMB1686	GCF_002884775.1	Scaffold	SAMN07511412	1510549	43.3	98.94	0.61	11	45	1112	Homo Sapiens	Diabetes	MSU
Gardnerella vaginalis	*Gardnerella* genomic species 9	6119V5	GCF_000263655.1	Contig	SAMN02393784	1499602	43.3	99.39	0.45	3	45	1124	Homo Sapiens	Assymptomatic	VS
Gardnerella vaginalis	*Gardnerella* genomic species 9	N160	GCF_003408775.1	Scaffold	SAMN04446403	1509687	43.3	99.39	0.45	9	43	1123	Homo Sapiens	NA	VS (sex workers)
Gardnerella vaginalis	*Gardnerella* genomic species 10	1500E	GCF_000263595.1	Contig	SAMN02393780	1548244	43	99.39	0.0	3	45	1171	Homo Sapiens	BV	VS/ Endometrium
Gardnerella vaginalis	*Gardnerella* genomic species 11	GED7760B	GCF_001546455.1	Scaffold	SAMN03851015	1489151	43.3	99.55	0.0	3	45	1131	Homo Sapiens	NA	VS
Gardnerella vaginalis	*Gardnerella* genomic species 12	CMW7778B	GCF_001563665.1	Scaffold	SAMN03851013	1602630	38	98.22	0.55	3	44	1159	Homo Sapiens	Asymptomatic (pregnant woman)	VS
Gardnerella sp.	*Gardnerella* genomic species 12	KA00735	GCF_002894085.1	Contig	SAMN05578249	1618183	38.2	98.22	1.0	6	45	1182	Homo Sapiens	NA	VS
Gardnerella vaginalis	Gardnerella genomic species 13	KA00225	GCF_002896555.1	Contig	SAMN05578087	1670019	40.8	98.57	0.15	8	45	1188	Homo Sapiens	NA	VS
Gardnerella sp.	*Gardnerella* genomic species 14	Marseille-Q2328	GCF_940796375.1	Contig	SAMEA1430585 9	1720610	45.3	99.76	0.0	3	45	1333	NA	NA	NA

Gardnerella vaginalis	*Gardnerella* genomic species 14	NR010	GCF_003408845.1	Scaffold	SAMN04446404	1622653	45.5	98.14	0.0	7	45	1203	Homo Sapiens	BV	VS (sex workers)

Abbreviations - B: Blood culture, CDS: Coding Sequence, GC%: Percentage of guanine (G) and cytosine (C) content, MSU: Midstream Urine, rRNA: Ribosomal RNA, tRNA: Transfer RNA, TUC: TransUrethral Catheter, VS: Vaginal Swab, NA: Not Available. Strains from our collection are represented in bold.

Table S2: Comprehensive analysis of the *Gardnerella* pangenome: core, soft core, dispensable, and strain-specific genes.

	Gene clusters	**Gene Calls**	**%**
Core	505	77421	10,5
Soft Core	316	47316	6,6
Singletons	1419	1437	29,4
Dispensable	2584	65109	53,6
Accessory	4319	113862	89,5
Single copy genes in Core	92	13708	1,9
Total	4824	191283	100

Table S3: Core phylogeny and SNP matrix analysis of *G. vaginalis, G. swidsinskii, G. leopoldii* and *G. pickettii*.

snp-dists 0.8.2													
G. vaginalis	**c11Ua_71**	**c15Ub_15**	**c15Ub_3**	**c15Ub_5**	**c15VSa_39**	**c15VSb_14**	**c15VSb_4**	**c21Ua_54**	**c29Ua_32**	**c29Ub_7**	**c29VSa_19**	**c29VSb_3**	
c11Ua_71	0	9922	9823	9977	9733	9806	9835	10119	9800	9804	9800	9804	**Vaginal swab**
c15Ub_15	9922	0	1580	1275	518	1195	1572	7322	6972	6976	6972	6976	**Urine**
c15Ub_3	9823	1580	0	665	1378	529	10	7374	7134	7138	7134	7138	
c15Ub_5	9977	1275	665	0	1379	786	673	7362	7181	7185	7181	7185	
c15VSa_39	9733	518	1378	1379	0	1053	1370	7255	6962	6966	6962	6966	
c15VSb_14	9806	1195	529	786	1053	0	523	7355	7077	7081	7077	7081	
c15VSb_4	9835	1572	10	673	1370	523	0	7379	7139	7143	7139	7143	
c21Ua_54	10119	7322	7374	7362	7255	7355	7379	0	6436	6440	6436	6440	
c29Ua_32	9800	6972	7134	7181	6962	7077	7139	6436	0	4	0	4	
c29Ub_7	9804	6976	7138	7185	6966	7081	7143	6440	4	0	4	0	
c29VSa_19	9800	6972	7134	7181	6962	7077	7139	6436	0	4	0	4	
c29VSb_3	9804	6976	7138	7185	6966	7081	7143	6440	4	0	4	0	

G. swidsinskii		c15Ua_15	c15Ua_23	c15Ub_13	c15VSa_36	c15VSb_13							
c10Ua_244	0	9196	9201	9201	9197	9201							
c15Ua_15	9196	0	11	11	3	11							
c15Ua_23	9201	11	0	2	12	2							
c15Ub_13	9201	11	2	0	12	0							
c15VSa_36	9197	3	12	12	0	12							
c15VSb_13	9201	11	2	0	12	0							

G. leopoldii	c18Ua_110	c18Ua_112	c1Ub_40	c1Ub_6	c31Ua_40	c31VSa_14	c31VSa_35						
c18Ua_110	0	0	8006	8006	7737	7737	7735						
c18Ua_112	0	0	8006	8006	7737	7737	7735						
c1Ub_40	8006	8006	0	3	7257	7257	7261						
c1Ub_6	8006	8006	3	0	7256	7256	7259						
c31Ua_40	7737	7737	7257	7256	0	0	0						
c31VSa_14	7737	7737	7257	7256	0	0	0						
c31VSa_35	7735	7735	7261	7259	0	0	0						

G. pickettii	10Ua_31	12Ub_1	17Ua_112	23Ua_15	23Ua_38	17Ua_118	23Ub_4	GED7275B		Asymptomatic			
c10Ua_31	0	19186	18348	18348	18348	19875	18538	18347		OAB			
c12Ub_1	19186	0	18704	18704	18704	7707	18318	18702					
c17Ua_112	18348	18704	0	1	1	20162	18683	1					
c23Ua_15	18348	18704	1	0	1	20162	18683	1					
c23Ua_38	18348	18704	1	1	0	20162	18683	1					
c17Ua_118	19875	7707	20162	20162	20162	0	19902	20163					
c23Ub_4	18538	18318	18683	18683	18683	19902	0	18682					

Table S4: Functional enrichment analysis through KEGG modules and COG pathways enriched in specific *Gardnerella* species.

KEGG Class	COG Category	KEGG Module	adjusted q value	accession	gene_clusters IDs	KOfam	COG20 FUNCTION	adjusted q value	Associated groups
Nucleotide Metabolism	-	Purine degradation, xanthine => urea	1.671437693 8587114e-24	M00546	GC_00001432	allantoinase [EC:3.5.2.5]	-	3.180950255 367414e-24	GG13, **GL**, GG14, **GS**
Amino acid Metabolism	Amino acid transport and metabolism (E)	Tryptophan biosynthesis, chorismate => tryptophan	1.671437693 8587114e-24	M00023	GC_00001417	tryptophan synthase beta chain [EC:4.2.1.20]	tryptophan synthase beta chain TrpB	3.180950255 367414e-24	GG10, **GL**, GG7, **GS**
Carbohydrate Metabolism	Carbohydrate transport and metabolism (G)	Glucuronate pathway (uronate pathway)	5.035626164 389603e-23	M00014	GC_00001430, GC_00001619	Glucuronate pathway (uronate pathway)+4:5	-	8.817689842 132147e-23	GG2, **GV**
						-	Sugar (pentulose or hexulose) kinase (XylB)	4.082517001 894295e-24	GG2, **GV**
Lipid Metabolism	Lipid transport and metabolism (I)	Fatty acid biosynthesis, elongation	3.788724738 425022e-13	M00083	GC_00001345	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	-	5.276467058 772674e-13	GG2, **GV**
						-	NAD(P)-dependent dehydrogenase, short-chain alcohol	4.082517001 894295e-24	**GV, GP**, GG2, GG11, GG3

							dehydrogenase family (FabG)		
Carbohydrate Metabolism	Carbohydrate transport and metabolism (G)	Pentose phosphate pathway, archaea, fructose 6P => ribose 5P	8.701806646586102e-10	M00580	GC_00001429	6-phospho-3-hexuloisomerase [EC:5.3.1.27]		1.1866776334904953e-9	GG2, **GV**
							D-arabinose 5-phosphate isomerase GutQ (GutQ)	1.300161173844769e-9	**GV**, GG2
Metabolism of cofactors and vitamins	Amino acid transport and metabolism (E)	Thiamine biosynthesis, prokaryotes, AIR (+ DXP/glycine) => TMP/TPP	1.6714376938587114e-24	M00895; M00127	GC_00000883	glycine oxidase [EC:1.4.3.19]		3.180950255367414e-24	GG13, GG10, **GL**, GG9, GG2, **GV**, GG14, GG7, GG12, GG8, **GS**
							Glycine/D-amino acid oxidase (deaminating) (DadA)	4.082517001894295e-24	**GV**, GG12, GG9, **GS**, GG2, GG13, GG14, **GL**, GG10, GG7, GG8
					GC_00000886; GC_00001422	thiazole synthase [EC:2.8.1.10]		3.180950255367414e-24	GG13, GG10, **GL**, GG9, GG2, **GV**, GG14, GG7, GG12, GG8, **GS**
							Thiazole synthase ThiGH, ThiG subunit	4.082517001894295e-24	**GV**, GG12, GG9, **GS**, GG2, GG13, GG14, **GL**, GG10, GG7, GG8
					GC_00000888; GC_00001423	sulfur carrier protein		3.180950255367414e-24	GG13, GG10, **GL**, GG9, GG2, **GV**, GG14, GG7, GG12, GG8, **GS**
							Sulfur carrier protein ThiS	4.082517001894295e-24	**GV**, GG12, GG9, **GS**, GG2, GG13, GG14, **GL**, GG10, GG7, GG8

					GC_00000894	phosphomethylpyrimidine synthase [EC:4.1.99.17]		3.180950255367414e-24	GG13, GG10, **GL**, GG9, GG2, **GV**, GG14, GG7, GG8, **GS**
							Thiamine monophosphate synthase (ThiE)	4.082517001894295e-24	**GV**, GG9, **GS**, GG2, GG13, GG14, **GL**, GG10, GG7, GG8
Carbohydrate Metabolism	Carbohydrate transport and metabolism (G)	Galactose degradation, Leloir pathway, galactose => alpha-D-glucose-1P	1.6714376938587114e-24	M00632	GC_00000024	UDPglucose--hexose-1-phosphate uridylyltransferase [EC:2.7.7.12]		3.180950255367414e-24	**GP**, GG11, GG13, GG2, **GG3**, **GV**, GG12
							Galactose-1-phosphate uridylyltransferase (GalT)	4.082517001894295e-24	**GV**, GG12, **GP**, GG2, GG11, GG13, **GG3**
					GC_00001149	galactokinase [EC:2.7.1.6]		3.180950255367414e-24	GG2, **GV**
							Galactokinase (GalK)	4.082517001894295e-24	**GV**, GG2
Amino acid Metabolism	Amino acid transport and metabolism (E)	Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine	1.6714376938587114e-24	M00016	GC_00000912	4-hydroxy-tetrahydrodipicolinate synthase [EC:4.3.3.7]		3.180950255367414e-24	**GP**, GG11, GG10, GG9, GG2, **GG3**, **GV**, GG14, GG8
					GC_00000996	2-dehydro-3-deoxy-D-pentonate aldolase [EC:4.1.2.28]		3.180950255367414e-24	**GP**, GG11, GG2, **GG3**, **GV**
Metabolism of cofactors and vitamins	Nucleotide transport and metabolism (F)	Menaquinone biosynthesis, futalosine pathway	1.6714376938587114e-24	M00930	GC_00000993	aminodeoxyfutalosine deaminase [EC:3.5.4.40]		3.180950255367414e-24	**GP**, GG11, GG2, **GG3**, **GV**
							Cytosine/adenosine deaminase or related metal-	4.082517001894295e-24	**GV**, **GP**, GG2, GG11, **GG3**

							dependent hydrolase (SsnA)		
Signal Transduction	-	Vancomycin resistance, D-Ala-D-Lac type	1.671437693 8587114e-24	M00651	GC_00000946	vancomycin resistance protein VanJ		3.180950255 367414e-24	**GP**, GG11, GG2, **GG3**, **GV**
Carbohydrate Metabolism	Carbohydrate transport and metabolism (G)	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	1.671437693 8587114e-24	M00001	GC_00000913; GC_00001055; GC_00001129; GC_00002713	glucokinase [EC:2.7.1.2]	glucokinase	3.180950255 367414e-24	**GP**, GG11, GG13, GG10, GG9, GG2, **GG3**, **GV**, GG14, GG12, GG8
Carbohydrate Metabolism	Carbohydrate transport and metabolism (G)	Keratan sulfate degradation	3.896400665 739811e-17	M00079	GC_00001063	hexosaminidase [EC:3.2.1.52]		5.662852530 056376e-17	**GP**, GG2, GG3, **GV**
							N-acetyl-beta-hexosaminidase (Chb)	6.665470311 796228e-17	GV, GP, GG2, GG3
Amino acid Metabolism	Coenzyme transport and metabolism H)	Methionine degradation	2.557643454 8830647e-20	M00035	GC_00000969	adenosylhomocysteinase [EC:3.3.1.1]		3.180950255 367414e-24	GP, GG11, GG2, GG3, GV
							S-adenosylhomocysteine hydrolase (SAM1)	4.082517001 894295e-24	GV, GP, GG2, GG11, GG3
	Replication, recombination and repair (L)				GC_00000969, GC_00001488, GC_00001806,	DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]		5.961711533 423468e-6	GG9, **GG3**, **GV**, GG14, GG12, **GS**

					GC_00001834, GC_00002012, GC_00002021, GC_00002238, GC_00002258, GC_00002500, GC_00003469, GC_00003573, GC_00004166, GC_00004346, GC_00004520		DNA-cytosine methylase (Dcm)	2.908565561 3419612e-5	**GV**, GG12, GG9, **GS**, GG14, **GG3**
Amino acid Metabolism	Amino acid transport and metabolism (E)	Asparagine biosynthesis	6.663165365 58072e-24	COG0367	GC_00001351	asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]		3.180950255 367414e-24	GG10, **GL**, GG9, GG7, GG8, **GS**
							Asparagine synthetase B (glutamine-hydrolyzing) (AsnB) (PDB:2ETD) (PUBMED:2595 1518)	4.082517001 894295e-24	GG9, **GS**, **GL**, GG10, GG7, GG8
Carbohydrate Metabolism	Cell wall/membrane/envelope biogenesis (M); Carbohydrate transport and metabolism (G)	Non-phosphorylated Entner-Doudoroff pathway	3.481264639 0704204e-23	COG4948	GC_00000987	L-fuconate dehydratase [EC:4.2.1.68]		3.180950255 367414e-24	**GP**, GG11, GG2, **GG3**, **GV**
							L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	4.082517001 894295e-24	GV, GP, GG2, GG11, GG3

							(RspA) (PDB:1BKH)		
					GC_00001024	gluconokinase [EC:2.7.1.12]		1.784026362 46593e-23	GP, GG2, GG3, GV
							Gluconate kinase (GntK)	2.257156864 8628508e-23	GV, GP, GG2, GG3
					GC_00001159; GC_00001430; GC_00001619		Sugar (pentulose or hexulose) kinase (XylB)	4.082517001 894295e-24	GV, GG2
					GC_00001430; GC_00001619	xylulokinase [EC:2.7.1.17]		8.817689842 132147e-23	GG2, GV
Genetic Information processing	Coenzyme transport and metabolism (H); Inorganic ion transport and metabolism (P)	Molybdopterin biosynthesis	6.663165365 58072e-24	COG0476	GC_00000889, GC_00001427, GC_00002300, GC_00004470	molybdopterin-synthase adenylyltransferase [EC:2.7.7.80]		3.180950255 367414e-24	GG13, GG10, **GL**, GG9, GG2, **GV**, GG14, GG7, GG12, GG8, **GS**
							Molybdopterin or thiamine biosynthesis adenylyltransferase (ThiF)	4.082517001 894295e-24	**GV**, GG12, GG9, **GS**, GG2, GG13, GG14, **GL**, GG10, GG7, GG8
					GC_00001469		ABC-type molybdate transport system, periplasmic Mo-binding protein ModA (ModA) (PDB:1AMF)	3.610607236 381702e-18	GG12, GG9, **GS**, GG13, **GL**

Abbreviations - GV: *G. vaginalis,* GL: *G. leopoldii,* GP: *G. piotii,* GS: *G. swidsinskii,* GPICK: *G. pickettii,* GG2, GG7 - GG14: *Gardnerella* genomic species 2, 7 - 14.

Table S5: Antibiotic Resistance Genes identified in *Gardnerella* pangenome.

		AMINOGLYCOSIDE		MACROLIDE						TETRACYCLINE		LINCOSAMIDE; STREPTOGRAMIN	
		Aminoglycoside O-phosphotransferase APH(3')-Ia		Macrolide efflux MFS transporter Mef(A)		ABC-F type ribosomal protection protein Msr(D)		23S rRNA (adenine(2058)-N(6))-methyltransferase Erm(X)		Tetracycline resistance ribosomal protection protein Tet(M)		ABC-F type ribosomal protection protein Lsa(C)	
		% Coverage	% Identity	% Coverage	% Identity	% Coverage	% Identity	% Coverage	% Identity	% Coverage	% Identity	% Coverage	% Identity
G. leopoldii	UMB0912			100.00	100.00	100.00	100.00					100.00	91.48
	UMB0913			100.00	100.00	100.00	100.00					100.00	91.48
	UGent 06.41					100.00	100.00					100.00	91.55
	AMD											100.00	91.75
	UGent 09.48											100.00	91.55
	6420B											100.00	91.55
	30-4												
	UMB1489					100.00	100.00					100.00	91.55
	UMB0662											100.00	91.55
	UMB0682					100.00	100.00					100.00	91.55
	UMB0742											100.00	91.55
	UMB1350											100.00	91.55
	18Ua_110											100.00	91.55
	18Ua_112											100.00	91.55

	c1Ub_40												
	c1Ub_6												
	c31VSa_14												
	c31Ua_40												
G. piotii	UGent 18.01											100.00	91.28
	UGent 21.28					100.00	100.00					100.00	91.55
	JNFY15							100.00	100.00			100.00	91.55
	JCP8151A									100.00	99.84		
	JCP8151B											100.00	91.41
	JCP8070											100.00	91.55
	JCP8066			100.00	99.51	100.00	99.79					100.00	91.55
	JCP8522											100.00	91.89
	GH019											100.00	91.28
G. swidsinskii	UMB1642											100.00	91.82
	JNFY3							100.00	100.00				
	UMB1698					100.00	100.00						
	GV37												
	409-05			100.00	100.00	100.00	100.00					100.00	91.55
	5-1	93.36	100.00									100.00	91.55
	26-12												

	UMB0264											100.00	91.35
	UMB0170											100.00	91.35
	GS 9838-1												
	GS 10234											100.00	91.55
	DNF01162												
	N72									100.00	99.84	100.00	91.82
	UMB0769									100.00	100.00	100.00	91.55
	c15Ua_15											100.00	91.35
	c15Ua_23											100.00	91.35
	c15Ub_13											100.00	91.35
	c15VSb_13											100.00	91.35
GG2	55152			100.00	100.00	100.00	100.00						
	JCP8108											100.00	91.82
	1400E											100.00	91.55
	41V												
	JCP7275	100.00	100.00	100.00	100.00	100.00	100.00			100.00	99.79	100.00	91.21
	N165									100.00	99.84		
G. pickettii	UMB0830											100.00	91.08
	UMB0833			100.00	100.00	100.00	100.00			100.00	100.00	100.00	91.48
	00703Bmash											100.00	91.55

	00703C2mash											99.93	91.55
	JCP7659											100.00	91.01
	JCP8017B											100.00	91.08
	GED7275B					100.00	100.00					100.00	91.48
	JCP8017A											100.00	91.08
	JCP7719											100.00	91.55
	N101									100.00	99.84	100.00	91.62
	N153											100.00	91.62
	N144									100.00	99.84	100.00	91.41
	N95											100.00	91.62
	W11									100.00	99.84	100.00	91.48
	UMB0558											100.00	91.62
	c10Ua_31											100.00	91.28
	12Ub_1											100.00	91.62
	17Ua_112					100.00	100.00					100.00	91.48
	17Ua_118											100.00	91.62
	23Ua_15					100.00	100.00					100.00	91.48
	23Ua_38					100.00	100.00					100.00	91.48
	23Ub_4											100.00	91.62
	JNFY17											100.00	91.82

GG7	JCP8481A			100.00	100.00	100.00	100.00						
	JCP8481B			100.00	100.00	100.00	100.00						
	PSS_7772B			100.00	100.00	100.00	100.00						
G. greenwoodii	101											100.00	91.48
	00703Dmash											100.00	91.48
	UMB1686			100.00	99.75	100.00	100.00					100.00	91.41
	KC1											100.00	91.48
	c31Ua_26			100.00	99.75	100.00	100.00					100.00	91.41
GG9	6119V5											100.00	91.28
	N160											100.00	91.21
GG10	1500E			100.00	100.00	100.00	100.00					100.00	91.55
GG11	GED7760B			100.00	100.00	100.00	100.00					100.00	91.28
GG12	CMW7778B											100.00	91.62
	KA00735			100.00	100.00	100.00	100.00			100.00	100.00	100.00	91.41
GG13	KA00225												
GG14	Marseille-Q2328									100.00	99.84	100.00	91.21
	NR010											100.00	91.55
G. vaginalis	c15Ub_15											100.00	91.55
	c15Ub_3											100.00	91.55
	c15Ub_5											100.00	91.55

	c15VSa_39											100.00	91.55
	c15VSb_14											100.00	91.55
	c15VSb_4											100.00	91.55
	c21Ua_54											100.00	91.62
	c29Ua_32												
	c29Ub_7												
	c29VSa_19												
	c29VSb_3												
	UMB0386											100.00	91.55
	18-4												
	284V											100.00	91.55
	23-12												
	UMB0775											100.00	91.28
	UMB0032A					100.00	100.00					100.00	91.62
	UMB0770			100.00	100.00	100.00	100.00					100.00	91.41
	UMB0032B					100.00	100.00					100.00	91.62
	315-A									100.00	99.69	100.00	91.08
	UGent 25.49											100.00	91.75
	DNF01149											100.00	90.94
	UMB0768											100.00	91.35

	3549624												
	JCP7672												
	ATCC 14019											100.00	91.55
	14019_MetR											100.00	91.55
	0288E											100.00	91.55
	UGent 09.07											100.00	91.75
	HMP9231											100.00	91.48
	JCP7276											100.00	91.55
	ATCC 49145B									100.00	99.69	100.00	91.55
	75712											100.00	91.48
	FDAARGOS_296									100.00	100.00	100.00	91.55
	UMB0233												
	UMB0061											100.00	91.48
	UMB0298											100.00	91.55
	NR038											100.00	91.41
	NR039											100.00	91.48
	UMB0202											100.00	91.55
	GH015											100.00	91.55
	UMB0143									100.00	100.00	100.00	90.94

	UMB0358			100.00	100.00	100.00	100.00					100.00	91.41
	UMB0540											100.00	91.08
	JNFY11							100.00	100.00			100.00	91.28
	14018c											100.00	91.48
	KC4												
	JNFY9							100.00	100.00	100.00	99.69	100.00	91.55
	NCTC10287											100.00	91.55
	UMB0736											100.00	91.55
	JNFY4							100.00	100.00	100.00	99.22	100.00	91.55
	KC2											100.00	91.01
	KC3											100.00	91.55
	06-12-0010											100.00	91.48
	JNFY14											100.00	91.28
	JNFY1							100.00	100.00			100.00	91.55
	JNFY13									100.00	99.53	100.00	91.55
	FDAARGOS_568									100.00	100.00	100.00	91.55

Table S6: Virulence-associated functions and gene clusters in the *Gardnerella* genus.

				GV (*n* = 58)		GL (*n* = 18)		GP (*n* = 9)		GS (*n* = 18)		GG2 (*n* = 6)		GPICK (*n* = 23)		GG7 (*n* = 3)		GG (*n* = 5)		GG9 (*n* = 2)		GG10 (*n* = 1)		GG11 (*n* = 1)		GG12 (*n* = 2)		GG13 (n = 1)		GG14 (n = 2)	
	COG/KEGG no.	**COG/KEGG name**	Gene Cluster	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
ADHESION	COG0050	Elongation factor Tu	GC_00000050	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG1704	lipoprotein antigen LemA family	GC_00000868	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	22	95,7	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG3942	Surface antigen	GC_00000986	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
			GC_00001331	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
			GC_00002447	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	1	100,0	0	0,0
			GC_00003018	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	66,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00000306	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG1464	ABC-type metal ion transport system, periplasmic component/surface antigen (NlpA)	GC_00000754	56	96,6	18	100,0	9	100,0	18	100,0	6	100,0	22	95,7	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K14956	6 kDa early secretory antigenic target	GC_00000686	56	96,6	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG3764	Sortase (surface protein transpeptidase) (SrtA)	GC_00000001	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000664	56	96,6	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00003434	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00003320	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	66,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001484	0	0,0	11	61,1	0	0,0	17	94,4	0	0,0	0	0,0	0	0,0	3	60,0	0	0,0	1	100,0	0	0,0	0	0,0	1	100,0	0	0,0
			GC_00000937	56	96,6	5	27,8	1	11,1	17	94,4	6	100,0	0	0,0	2	66,7	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG4962	Pilus assembly protein, ATPase of CapF family	GC_00000523	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG4965	Flp pilus assembly protein TadB (TadB)	GC_00002896	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00001309	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0
			GC_00002840	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	2	100,0	1	100,0	0	0,0

		GC_00000966	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	2	100,0
		GC_00000990	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
COG4961	Flp pilus assembly protein TadG	GC_00000475	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	0	0,0	2	100,0
COG5010	Flp pilus assembly protein TadD	GC_00003220	2	3,4	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
COG2064	Flp pilus assembly protein TadC	GC_00001352	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
K02279	pilus assembly protein CpaB	GC_00002853	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	1	100,0	0	0,0
		GC_00001328	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
		GC_00000992	58	100,0	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
COG0459	Hsp60 60K heat shock protein HtpB	GC_00000133	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
		GC_00003686	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	4,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
COG0057	Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase (GapA)	GC_00000227	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
COG0707	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase (MurG)	GC_00000556	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
		GC_00004430	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
COG1835	Peptidoglycan/LPS O-acetylase OafA/YrhL, contains acyltransferase and SGNH-hydrolase domains (OafA)	GC_00004380	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
		GC_00000941	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	1	100,0	0	0,0
		GC_00001037	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	18	78,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
		GC_00001332	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
COG1559	Endolytic transglycosylase MltG, terminates peptidoglycan polymerization (MltG)	GC_00000613	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
COG0744	Penicillin-binding protein 1B/1F, peptidoglycan transglycosylase/transpeptidase (MrcB)	GC_00000315	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
		GC_00000337	58	100,0	18	100,0	7	77,8	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0

	COG0728	Lipid II flippase MurJ/MviN (peptidoglycan biosynthesis) (MurJ)	GC_00000268	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000117	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	0	0,0	0	0,0	2	100,0
			GC_00002839	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	1	100,0	0	0,0
	COG0768	Cell division protein FtsI, peptidoglycan transpeptidase (Penicillin-binding protein 2) (FtsI)	GC_00000171	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000726	58	100,0	18	100,0	8	88,9	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG3409	Peptidoglycan-binding (PGRP) domain of peptidoglycan hydrolases (PGRP)	GC_00004795	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001595	9	15,5	3	16,7	0	0,0	6	33,3	0	0,0	4	17,4	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0
			GC_00001570	9	15,5	3	16,7	0	0,0	6	33,3	2	33,3	2	8,7	0	0,0	1	20,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
	COG2885	Outer membrane protein OmpA and related peptidoglycan-associated (lipo)proteins (OmpA)	GC_00001637	0	0,0	5	27,8	9	100,0	3	16,7	0	0,0	4	17,4	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0
			GC_00001049	55	94,8	0	0,0	0	0,0	0	0,0	4	66,7	23	100,0	0	0,0	2	40,0	2	100,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0
	COG2348	Lipid II:glycine glycyltransferase (Peptidoglycan interpeptide bridge formation enzyme) (FmhB)	GC_00000940	58	100,0	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0	1	100,0	2	100,0
	K14195	surface protein G	GC_00002206	0	0,0	0	0,0	1	11,1	0	0,0	0	0,0	0	0,0	0	0,0	2	40,0	1	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG3274	Surface polysaccharide O-acyltransferase WecH	GC_00002784	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	3	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001583	0	0,0	0	0,0	0	0,0	18	100,0	0	0,0	5	21,7	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	2	100,0
IRON UPTAKE	COG0396	FeS assembly ATPase SufC	GC_00000656	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG0672	High-affinity Fe2+/Pb2+ permease (FTR1)	GC_00000622	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K07230	periplasmic iron binding protein	GC_00000506	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K09819	manganese/iron transport system permease protein	GC_00000505	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K02012	iron(III) transport system substrate-binding protein	GC_00002120	6	10,3	0	0,0	0	0,0	0	0,0	2	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K02010	iron(III) transport system ATP-binding protein	GC_00002118	6	10,3	0	0,0	0	0,0	0	0,0	2	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0

	K02013	iron complex transport system ATP-binding protein	GC_00000086	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K02011	iron(III) transport system permease protein	GC_00002115	6	10,3	0	0,0	0	0,0	0	0,0	2	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K03711	Fur family transcriptional regulator, ferric uptake regulator	GC_00000371	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K07025	putative hydrolase of the HAD superfamily	GC_00000730	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K01626	3-deoxy-7-phosphoheptulonate synthase	GC_00000054	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG0783	DNA-binding ferritin-like protein (oxidative damage protectant) (Dps)	GC_00000015	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG2906	Bacterioferritin-associated ferredoxin (Bfd)	GC_00003870	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG0635	Coproporphyrinogen-III oxidase HemN (oxygen-independent) or related Fe-S oxidoreductase (HemN)	GC_00000156	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
SECRETION SYSTEMS	K03070	preprotein translocase subunit SecA	GC_00000234	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00002272	0	0,0	0	0,0	0	0,0	5	27,8	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0
	K03551	Holliday junction DNA helicase RuvB	GC_00000409	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG3505	type IV secretion system protein VirD4	GC_00004263	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00004090	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00003077	0	0,0	0	0,0	2	22,2	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K07344	type IV secretion system protein TrbL	GC_00003099	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	1	4,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K03695	ATP-dependent Clp protease ATP-binding subunit ClpB	GC_00000078	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00002208	0	0,0	0	0,0	1	11,1	0	0,0	2	33,3	2	8,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0
	K03466	DNA segregation ATPase FtsK/SpoIIIE, S-DNA-T family	GC_00000273	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0

			GC_00000620	58	100,0	18	100,0	8	88,9	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	1	50,0	1	100,0	2	100,0
			GC_00003668	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00003428	1	1,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00002625	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
	K03628	transcription termination factor Rho	GC_00000466	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
STRESS PROTEIN	K03696	endopeptidase Clp ATP-binding chain C(ClpC)	GC_00000168	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
TOXINS	COG1253	Hemolysin-related protein	GC_00000508	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000087	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG0345	Pyrroline-5-carboxylate reductase (ProC)	GC_00000598	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K11031	thiol-activated cytolysin (Vaginolysin)	GC_00000853	57	98,3	18	100,0	1	11,1	18	100,0	4	66,7	19	82,6	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	2	100,0	1	100,0	2	100,0
	COG2755	Lysophospholipase L1 or related esterase	GC_00001629	0	0,0	0	0,0	9	100,0	0	0,0	0	0,0	14	60,9	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
	COG1598	Antitoxin component HicB of the HicAB toxin-antitoxin system (HicB)	GC_00001038	44	75,9	18	100,0	1	11,1	9	50,0	2	33,3	9	39,1	3	100,0	0	0,0	1	50,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0
			GC_00001087	38	65,5	0	0,0	5	55,6	6	33,3	0	0,0	18	78,3	0	0,0	3	60,0	1	50,0	1	100,0	0	0,0	1	50,0	1	100,0	1	50,0
			GC_00000840	58	100,0	18	100,0	9	100,0	6	33,3	6	100,0	19	82,6	3	100,0	5	100,0	1	50,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
	COG3077	Antitoxin component of the RelBE or YafQ-DinJ toxin-antitoxin module (RelB)	GC_00001044	34	58,6	1	5,6	2	22,2	17	94,4	3	50,0	23	100,0	3	100,0	0	0,0	0	0,0	1	100,0	1	100,0	2	100,0	1	100,0	1	50,0
			GC_00001059	22	37,9	8	44,4	7	77,8	9	50,0	6	100,0	19	82,6	3	100,0	5	100,0	2	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
			GC_00001061	46	79,3	0	0,0	5	55,6	14	77,8	0	0,0	11	47,8	3	100,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0	0	0,0	1	50,0
			GC_00000017	57	98,3	11	61,1	6	66,7	8	44,4	6	100,0	22	95,7	3	100,0	3	60,0	2	100,0	1	100,0	1	100,0	2	100,0	0	0,0	2	100,0
			GC_00001365	1	1,7	15	83,3	2	22,2	2	11,1	3	50,0	15	65,2	3	100,0	3	60,0	1	50,0	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0
			GC_00001399	0	0,0	18	100,0	2	22,2	0	0,0	0	0,0	19	82,6	0	0,0	0	0,0	2	100,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
			GC_00001409	0	0,0	18	100,0	3	33,3	2	11,1	5	83,3	6	26,1	0	0,0	2	40,0	2	100,0	0	0,0	1	100,0	0	0,0	0	0,0	1	50,0
			GC_00001127	43	74,1	0	0,0	3	33,3	0	0,0	3	50,0	17	73,9	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0

			GC_00001130	56	96,6	0	0,0	0	0,0	0	0,0	6	100,0	2	8,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0
			GC_00001515	0	0,0	0	0,0	3	33,3	14	77,8	0	0,0	1	4,3	3	100,0	2	40,0	2	100,0	1	100,0	1	100,0	2	100,0	0	0,0	2	100,0
			GC_00001856	2	3,4	0	0,0	3	33,3	0	0,0	5	83,3	3	13,0	1	33,3	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	1	50,0
			GC_00001969	1	1,7	0	0,0	3	33,3	0	0,0	4	66,7	1	4,3	0	0,0	2	40,0	1	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00002921	0	0,0	0	0,0	1	11,1	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG2161	Antitoxin component YafN of the YafNO toxin-antitoxin module, PHD/YefM family (StbD)	GC_00001074	15	25,9	9	50,0	7	77,8	12	66,7	2	33,3	16	69,6	3	100,0	5	100,0	2	100,0	0	0,0	0	0,0	2	100,0	0	0,0	2	100,0
			GC_00001993	1	1,7	0	0,0	1	11,1	0	0,0	5	83,3	0	0,0	3	100,0	0	0,0	2	100,0	1	100,0	0	0,0	2	100,0	0	0,0	2	100,0
	K19092	toxin ParE1/3/4	GC_00001089	33	56,9	1	5,6	2	22,2	17	94,4	0	0,0	14	60,9	3	100,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0	1	100,0	0	0,0
			GC_00001462	18	31,0	1	5,6	3	33,3	2	11,1	0	0,0	6	26,1	0	0,0	0	0,0	0	0,0	1	100,0	1	100,0	0	0,0	0	0,0	1	50,0
			GC_00001117	17	29,3	8	44,4	5	55,6	9	50,0	5	83,3	16	69,6	0	0,0	5	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
			GC_00001123	0	0,0	18	100,0	3	33,3	18	100,0	0	0,0	13	56,5	0	0,0	1	20,0	1	50,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
			GC_00001523	8	13,8	6	33,3	5	55,6	0	0,0	2	33,3	0	0,0	3	100,0	2	40,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	1	50,0
			GC_00002071	0	0,0	0	0,0	1	11,1	0	0,0	0	0,0	0	0,0	3	100,0	0	0,0	2	100,0	0	0,0	0	0,0	2	100,0	0	0,0	1	50,0
	K13655	HTH-type transcriptional regulator / antitoxin MqsA	GC_00001068	54	93,1	0	0,0	0	0,0	0	0,0	5	83,3	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	0	0,0	1	100,0	2	100,0
	K19165	antitoxin Phd	GC_00000900	53	91,4	17	94,4	4	44,4	4	22,2	5	83,3	9	39,1	3	100,0	2	40,0	1	50,0	1	100,0	1	100,0	0	0,0	0	0,0	1	50,0
			GC_00002395	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	4	17,4	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K18918	RHH-type transcriptional regulator, rel operon repressor / antitoxin RelB	GC_00001374	1	1,7	0	0,0	6	66,7	18	100,0	0	0,0	12	52,2	2	66,7	2	40,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	2	100,0
	COG4115	Toxin component of the Txe-Axe toxin-antitoxin module, Txe/YoeB family	GC_00001299	42	72,4	0	0,0	2	22,2	0	0,0	3	50,0	2	8,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001451	0	0,0	18	100,0	1	11,1	2	11,1	5	83,3	6	26,1	0	0,0	2	40,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
	COG3657	Putative component of the toxin-antitoxin plasmid stabilization module	GC_00001415	0	0,0	2	11,1	4	44,4	16	88,9	0	0,0	0	0,0	3	100,0	3	60,0	2	100,0	1	100,0	1	100,0	0	0,0	1	100,0	1	50,0
	COG3041	mRNA-degrading endonuclease YafQ (mRNA interferase), toxin component of the YafQ-DinJ toxin-antitoxin module (YafQ)	GC_00001459	0	0,0	13	72,2	2	22,2	0	0,0	0	0,0	19	82,6	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001428	2	3,4	0	0,0	5	55,6	0	0,0	0	0,0	11	47,8	3	100,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0	0	0,0	1	50,0
			GC_00001696	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	17	73,9	0	0,0	1	20,0	2	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0

			GC_00001724	0	0,0	0	0,0	0	0,0	10	55,6	0	0,0	0	0,0	2	66,7	2	40,0	0	0,0	1	100,0	1	100,0	2	100,0	0	0,0	1	50,0
			GC_00001977	0	0,0	4	22,2	3	33,3	0	0,0	1	16,7	0	0,0	0	0,0	2	40,0	1	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001901	1	1,7	0	0,0	2	22,2	0	0,0	0	0,0	3	13,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	1	50,0
			GC_00001934	1	1,7	0	0,0	5	55,6	0	0,0	5	83,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0
			GC_00002085	0	0,0	0	0,0	2	22,2	1	5,6	1	16,7	3	13,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0
			GC_00002039	0	0,0	0	0,0	3	33,3	3	16,7	0	0,0	1	4,3	1	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
	K18830	HTH-type transcriptional regulator / antitoxin PezA	GC_00001182	56	96,6	0	0,0	0	0,0	0	0,0	3	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG5340	Abortive phage infection protein AbiEi, antitoxin conponent of an AbiEi-AbiEii TA system (AbiEi)	GC_00001716	2	3,4	0	0,0	6	66,7	0	0,0	3	50,0	4	17,4	0	0,0	2	40,0	0	0,0	0	0,0	0	0,0	2	100,0	1	100,0	0	0,0
			GC_00001641	0	0,0	0	0,0	4	44,4	0	0,0	6	100,0	0	0,0	3	100,0	3	60,0	1	50,0	1	100,0	1	100,0	1	50,0	1	100,0	1	50,0
			GC_00001514	0	0,0	18	100,0	3	33,3	7	38,9	0	0,0	1	4,3	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0
			GC_00001908	1	1,7	0	0,0	0	0,0	0	0,0	6	100,0	0	0,0	0	0,0	3	60,0	2	100,0	1	100,0	1	100,0	0	0,0	0	0,0	0	0,0
			GC_00002652	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	3	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00004667	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0
	COG3657	Putative component of the toxin-antitoxin plasmid stabilization module	GC_00001723	10	17,2	0	0,0	0	0,0	7	38,9	1	16,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
			GC_00002666	0	0,0	0	0,0	1	11,1	0	0,0	0	0,0	0	0,0	2	66,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K21498	antitoxin HigA-1	GC_00001967	2	3,4	0	0,0	0	0,0	0	0,0	0	0,0	8	34,8	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
	K07334	toxin HigB-1	GC_00002548	0	0,0	0	0,0	0	0,0	0	0,0	1	16,7	0	0,0	0	0,0	2	40,0	1	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K18831	HTH-type transcriptional regulator / antitoxin HigA	GC_00002306	3	5,2	0	0,0	0	0,0	0	0,0	0	0,0	1	4,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
			GC_00003028	0	0,0	0	0,0	0	0,0	2	11,1	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00003098	1	1,7	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00003106	0	0,0	0	0,0	0	0,0	0	0,0	2	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG5444	Predicted ribonuclease, toxin component of the YeeF-YezG toxin-antitoxin module (YeeF)	GC_00002005	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	7	30,4	3	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00003791	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG2931	Ca2+-binding protein, RTX toxin-related	GC_00004223	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0

IMMUNE EVASION	K23987	dTDP-4-dehydrorhamnose reductase (RfbD)	GC_00001045	58	100,0	0	0,0	9	100,0	0	0,0	6	100,0	18	78,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K03431	phosphoglucosamine mutase mrsA/glmM	GC_00000701	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
	K01000	phospho-N-acetylmuramoyl-pentapeptide-transferase RgpG (Capsule)	GC_00000691	57	98,3	18	100,0	9	100,0	17	94,4	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
	K01710	dTDP-glucose 4,6-dehydratase	GC_00000837	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0	1	100,0	2	100,0				
	COG1136	ABC-type lipoprotein export system, ATPase component (LolD)	GC_00000829	56	96,6	18	100,0	8	88,9	15	83,3	6	100,0	19	82,6	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00000487	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00000408	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00000264	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00000245	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00000547	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00000742	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	22	95,7	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00004132	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0		
			GC_00001753	2	3,4	0	0,0	2	22,2	0	0,0	5	83,3	0	0,0	3	100,0	5	100,0	1	50,0	1	100,0	0	0,0	0	0,0	0	0,0	1	50,0				
	COG1376	Lipoprotein-anchoring transpeptidase ErfK/SrfK	GC_00000822	57	98,3	18	100,0	8	88,9	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	0	0,0	1	100,0	0	0,0				
			GC_00000198	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00003047	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0		
			GC_00003055	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0		
	COG4591	ABC-type transport system involved in lipoprotein release, permease component LolC	GC_00000614	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				
			GC_00000550	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0				

			GC_00000522	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000048	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_0000289	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	1	100,0	0	0,0
			GC_00001315	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
			GC_00001001	58	100,0	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
	COG0597	Lipoprotein signal peptidase (LspA)	GC_00000557	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG0682	Prolipoprotein diacylglyceryltransferase (Lgt)	GC_00000511	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG3127	Predicted ABC-type transport system involved in lysophospholipase L1 biosynthesis	GC_00000200	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K06078	murein lipoprotein	GC_00000890	45	77,6	0	0,0	7	77,8	0	0,0	6	100,0	20	87,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
	COG0739	Murein DD-endopeptidase MepM and murein hydrolase activator NlpD	GC_00001022	56	96,6	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	1	100,0	0	0,0
			GC_00001317	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	3	100,0	5	100,0	2	100,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
			GC_00003180	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
	COG3863	Uncharacterized conserved protein YycO, NlpC/P60 family (YycO)	GC_00001563	12	20,7	0	0,0	0	0,0	9	50,0	2	33,3	2	8,7	0	0,0	1	20,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
			GC_00001900	0	0,0	0	0,0	0	0,0	13	72,2	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
	COG0791	Cell wall-associated hydrolase, NlpC_P60 family (NlpC)	GC_00004152	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000588	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG0577	ABC-type antimicrobial peptide transport system, permease component (SalY)	GC_00000814	56	96,6	18	100,0	8	88,9	15	83,3	6	100,0	19	82,6	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000322	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00003790	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0
	K03387	NADH-dependent peroxiredoxin subunit F	GC_00000712	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	0	0,0	0	0,0	2	100,0

	COG0450	Alkyl hydroperoxide reductase subunit AhpC (peroxiredoxin) (AhpC)	GC_00000812	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	0	0,0	0	0,0	2	100,0
	COG4932	Clumping factor A-related surface protein, MSCRAMM (microbial surface components recognizing adhesive matrix molecules) family, DEv-IgG fold (ClfA)	GC_00003450	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00003242	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
			GC_00002166	0	0,0	0	0,0	2	22,2	0	0,0	0	0,0	5	21,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00002354	5	8,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00002304	0	0,0	1	5,6	1	11,1	0	0,0	4	66,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00002243	6	10,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00002303	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	3	60,0	2	100,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00002549	0	0,0	0	0,0	0	0,0	0	0,0	3	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
			GC_00002796	0	0,0	3	16,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001918	0	0,0	0	0,0	3	33,3	0	0,0	0	0,0	9	39,1	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
			GC_00001906	0	0,0	11	61,1	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
			GC_00001999	0	0,0	0	0,0	0	0,0	7	38,9	0	0,0	0	0,0	0	0,0	2	40,0	0	0,0	1	100,0	0	0,0	1	50,0	0	0,0	0	0,0
			GC_00001966	0	0,0	5	27,8	0	0,0	2	11,1	1	16,7	1	4,3	1	33,3	1	20,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
			GC_00001787	3	5,2	2	11,1	0	0,0	11	61,1	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0
			GC_00001846	12	20,7	0	0,0	0	0,0	0	0,0	2	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
			GC_00001587	0	0,0	18	100,0	0	0,0	3	16,7	0	0,0	0	0,0	0	0,0	1	20,0	1	50,0	1	100,0	0	0,0	0	0,0	1	100,0	0	0,0
			GC_00001518	7	12,1	0	0,0	0	0,0	3	16,7	4	66,7	8	34,8	0	0,0	1	20,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001688	0	0,0	5	27,8	0	0,0	14	77,8	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	1	50,0
			GC_00001179	26	44,8	17	94,4	0	0,0	0	0,0	6	100,0	0	0,0	0	0,0	3	60,0	2	100,0	0	0,0	1	100,0	2	100,0	1	100,0	0	0,0
			GC_00001479	33	56,9	0	0,0	1	11,1	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001367	40	69,0	1	5,6	0	0,0	0	0,0	0	0,0	6	26,1	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00000860	57	98,3	6	33,3	8	88,9	18	100,0	0	0,0	18	78,3	3	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG0741	Soluble lytic murein transglycosylase or regulatory proteins (may contain LysM/invasin domain) (MltE)	GC_00003886	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0

REGULATION	K03086	RNA polymerase primary sigma factor	GC_00000229	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K07273	lysozyme	GC_00002024	0	0,0	2	11,1	0	0,0	0	0,0	0	0,0	0	0,0	3	100,0	1	20,0	1	50,0	1	100,0	0	0,0	0	0,0	0	0,0	2	100,0
			GC_00001718	11	19,0	2	11,1	0	0,0	3	16,7	1	16,7	4	17,4	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001206	21	36,2	6	33,3	0	0,0	0	0,0	3	50,0	5	21,7	0	0,0	1	20,0	0	0,0	0	0,0	0	0,0	2	100,0	1	100,0	2	100,0
			GC_00001226	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	5	21,7	0	0,0	5	100,0	1	50,0	1	100,0	1	100,0	2	100,0	1	100,0	0	0,0
			GC_00001040	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	17	73,9	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	K18955	WhiB family transcriptional regulator	GC_00000653	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	1	50,0	1	100,0	2	100,0
	K07816	GTP pyrophosphokinase (RelA)	GC_00000813	55	94,8	18	100,0	7	77,8	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00002087	5	8,6	0	0,0	0	0,0	3	16,7	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001631	6	10,3	2	11,1	0	0,0	5	27,8	0	0,0	7	30,4	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001509	30	51,7	0	0,0	0	0,0	0	0,0	2	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG3279	DNA-binding response regulator, LytR/AlgR family	GC_00000294	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00004371	0	0,0	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
METABOLISM	K01466	Allantoinase AIIB (Allantion utilization)	GC_00001432	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0
	K10975	Allantoin permease	GC_00001435	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	2	100,0
BIOFILM	COG0174	Glutamine synthetase (GlnA)	GC_00000366	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000381	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG0463	Glycosyltransferase involved in cell wall bisynthesis (WcaA)	GC_00001039	58	100,0	0	0,0	9	100,0	0	0,0	6	100,0	18	78,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001048	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	18	78,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00000893	57	98,3	18	100,0	9	100,0	0	0,0	1	16,7	18	78,3	0	0,0	5	100,0	2	100,0	1	100,0	0	0,0	2	100,0	1	100,0	0	0,0
			GC_00001537	0	0,0	18	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	5	100,0	2	100,0	1	100,0	0	0,0	2	100,0	1	100,0	0	0,0
			GC_00001527	0	0,0	0	0,0	0	0,0	18	100,0	0	0,0	5	21,7	3	100,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	2	100,0
	COG0438		GC_00001539	0	0,0	0	0,0	0	0,0	18	100,0	0	0,0	5	21,7	3	100,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	2	100,0

		Glycosyltransferase involved in cell wall bisynthesis (RfaB)	GC_00000680	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG1216	Glycosyltransferase, GT2 family (WcaE)	GC_00001047	57	98,3	0	0,0	9	100,0	0	0,0	6	100,0	18	78,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001546	0	0,0	18	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	5	100,0	2	100,0	1	100,0	0	0,0	2	100,0	1	100,0	0	0,0
			GC_00001535	0	0,0	0	0,0	0	0,0	18	100,0	0	0,0	5	21,7	3	100,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	2	100,0
			GC_00000335	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00000278	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG1215	Glycosyltransferase, catalytic subunit of cellulose synthase and poly-beta-1,6-N-acetylglucosamine synthase (BcsA)	GC_00001738	2	3,4	0	0,0	1	11,1	0	0,0	3	50,0	10	43,5	1	33,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
			GC_00002759	0	0,0	0	0,0	0	0,0	0	0,0	3	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00000173	58	100,0	18	100,0	8	88,9	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	K07027	glycosyltransferase 2 family protein	GC_00000310	57	98,3	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG0855	Polyphosphate kinase	GC_00000072	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
ANTIPHAGOCYTOSIS	K00033	6-phosphogluconate dehydrogenase	GC_00000118	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
			GC_00001027	58	100,0	0	0,0	8	88,9	0	0,0	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
MUCIN DEGRADATION	COG2731	Beta-galactosidase, beta subunit (EbgC)	GC_00001216	10	17,2	6	33,3	3	33,3	11	61,1	4	66,7	13	56,5	0	0,0	1	20,0	0	0,0	1	100,0	0	0,0	2	100,0	0	0,0	0	0,0
	COG3250	Beta-galactosidase/beta-glucuronidase (LacZ)	GC_00001294	47	81,0	0	0,0	0	0,0	0	0,0	3	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001292	47	81,0	0	0,0	0	0,0	0	0,0	3	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
			GC_00001145	58	100,0	0	0,0	0	0,0	0	0,0	6	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG2723	Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase (BglB)	GC_00003911	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG3345	Alpha-galactosidase (GalA)	GC_00001639	20	34,5	0	0,0	0	0,0	0	0,0	3	50,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG3669	Alpha-L-fucosidase (AfuC)	GC_00001101	58	100,0	0	0,0	0	0,0	0	0,0	6	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0

	K15923	alpha-L-fucosidase 2	GC_00002820	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
	COG0383	Alpha-mannosidase (MngB)	GC_00000882	58	100,0	0	0,0	9	100,0	12	66,7	6	100,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
	COG3934	Endo-1,4-beta-mannosidase	GC_00001565	22	37,9	0	0,0	0	0,0	0	0,0	5	83,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
	COG4409	Neuraminidase (sialidase) NanH1	GC_00000904	58	100,0	0	0,0	9	100,0	0	0,0	6	100,0	23	100,0	0	0,0	5	100,0	2	100,0	1	100,0	1	100,0	0	0,0	0	0,0	2	100,0
		Neuraminidase (sialidase) NanH2	GC_00002035	0	0,0	0	0,0	6	66,7	0	0,0	0	0,0	2	8,7	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0	0	0,0	0	0,0	0	0,0
		Neuraminidase (sialidase) NanH3	GC_00001405	10	17,2	0	0,0	9	100,0	0	0,0	0	0,0	23	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
		Neuraminidase (sialidase) exo-alpha-sialidase	GC_00002684	3	5,2	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
OTHER VIRULENCE RELATED GENES	COG3475	Phosphorylcholine metabolism protein LicD	GC_00001194	0	0,0	18	100,0	0	0,0	18	100,0	0	0,0	5	21,7	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0
	COG0544	Trigger factor	GC_00000280	58	100,0	18	100,0	9	100,0	18	100,0	6	100,0	23	100,0	3	100,0	5	100,0	2	100,0	1	100,0	1	100,0	2	100,0	1	100,0	2	100,0

Abbreviations - GV: *G. vaginalis,* GL: *G. leopoldii,* GP: *G. piotii,* GS: *G. swidsinskii,* GPICK: *G. pickettii,* GG2, GG7 - GG14: *Gardnerella* genomic species 2, 7 - 14.

Table S7: Secondary metabolite biosynthetic gene clusters (BGCs) across *Gardnerella* genomes.

	GV (*n* = 58)		**GL** (*n* = 18)		**GP** (*n* = 9)		**GS** (*n* = 18)		**GG2** (*n* = 6)		**GPICK** (*n* = 23)		**GG7** (*n* = 3)		***GG*** (*n* = 5)		**GG9** (*n* = 2)		**GG10** (*n* = 1)		**GG11** (*n* = 1)		**GG12** (*n* = 2)		**GG13** (n = 1)		**GG14** (n = 2)	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Class II lanthipeptide clusters like mutacin II (U40620)	0	0,0	0	0,0	0	0,0	15	83,3	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
Other unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP) cluster	1	1,7	0	0,0	0	0,0	18	100,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
Type III PKS	54	93,1	1	5,6	4	44,4	0	0,0	3	50,0	19	82,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0	0	0,0	0	0,0
Linear azol(in)e-containing peptides	2	3,4	0	0,0	0	0,0	1	5,6	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0

Abbreviations - GV: *G. vaginalis,* GL: *G. leopoldii,* GP: *G. piotii,* GS: *G. swidsinskii,* GPICK: *G. pickettii,* GG2, GG7 - GG14: *Gardnerella* genomic species 2, 7 – 14.

Gardnerella pickettii sp. nov. (formerly *Gardnerella* genomic species 3) and *Gardnerella greenwoodii* sp. nov. (formerly *Gardnerella* genomic species 8) isolated from female urinary microbiome

Márcia Sousa[1,2], Magdalena Ksiezarek[1,2], Svetlana Ugarcina Perovic[1,2], Tiago Antunes-Lopes[3], Filipa Grosso[1,2], Teresa Gonçalves Ribeiro[1,2,4], and Luísa Peixe[1,2,4]

[1]Associate Laboratory i4HB , Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-- 313 Porto, Portugal; [2]UCIBIO – Applied Molecular Biosciences Unit, Department of Biological Sciences, Faculty of Pharmacy, University of Porto, 4050-- 313 Porto, Portugal; [3]Centro Hospitalar de São João, Porto, Portugal; [4]Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP), Porto, Portugal.

Int. J. Syst. Evol. Microbiol. 2023;73:006140.

INTERNATIONAL JOURNAL OF
SYSTEMATIC AND EVOLUTIONARY
MICROBIOLOGY

RESEARCH ARTICLE
Sousa *et al.*, *Int. J. Syst. Evol. Microbiol.* 2023;73:006140
DOI 10.1099/ijsem.0.006140

MICROBIOLOGY SOCIETY

Gardnerella pickettii sp. nov. (formerly *Gardnerella* genomic species 3) and *Gardnerella greenwoodii* sp. nov. (formerly *Gardnerella* genomic species 8) isolated from female urinary microbiome

Márcia Sousa[1,2], Magdalena Ksiezarek[1,2], Svetlana Ugarcina Perovic[1,2], Tiago Antunes-Lopes[3], Filipa Grosso[1,2], Teresa Gonçalves Ribeiro[1,2,4,*] and Luísa Peixe[1,2,4,*]

Abstract

During an ongoing female urinary microbiome research study, strains c17Ua_112^T and c31Ua_26^T isolated from urine samples of a patient diagnosed with overactive bladder and a healthy postmenopausal woman, respectively, could not be allocated to any *Gardnerella* species with valid names. In this work, we aimed to characterize these strains. The 16S rRNA gene sequences confirmed that these strains are members of the genus *Gardnerella*. Phylogenetic analysis based on *cpn60* strongly supported two clades, one encompassing c17Ua_112^T and nine other strains from the public database, and the other including c31Ua_26^T and three other strains, which were distinct from currently recognized species of the genus *Gardnerella*. Likewise, the phylogenomic tree also showed that strains c17Ua_112^T and c31Ua_26^T formed independent and robust clusters. Average nucleotide identity (ANI) and digital DNA–DNA hybridization (dDDH) values between c17Ua_112^T and c31Ua_26^T were 79.27 and 27.4%, respectively. Strain c17Ua_112^T showed the highest ANI (94.8%) and dDDH values (59.8%) with *Gardnerella piotii* UGent 18.01^T, and strain c31Ua_26^T revealed highest ANI (84.2%) and dDDH (29.1%) values with *Gardnerella swidsinskii* GS 9838-1^T. Based on the data presented here, the two strains c17Ua_112^T and c31Ua_26^T represent two novel species of the genus *Gardnerella*, for which the names *Gardnerella pickettii* (c17Ua_112^T=DSM 113414^T=CCP 71^T) and *Gardnerella greenwoodii* (c31Ua_26^T=DSM 113415^T=CCP 72^T) are proposed.

INTRODUCTION

The genus *Gardnerella* is classified in the phylum *Actinomycetota*, class *Actinomycetes*, order *Bifidobacteriales* and family *Bifidobacteriaceae* [1]. For almost 70 years, *Gardnerella vaginalis* was the only species reported within this genus. In 2019, the genus was emended, three novel species (*Gardnerella leopoldii*, *Gardnerella piotii* and *Gardnerella swidsinskii*) were described, and nine genomic species were delineated based on genomic comparisons, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and phenotypic analyses [2].

Over the last few decades, a strong correlation has been evidenced between *G. vaginalis* and bacterial vaginosis, the most common vaginal infection in women, which has also been associated with increased risk of acquiring sexually transmitted diseases [3, 4]. Furthermore, *Gardnerella* was frequently related to infertility, preterm labour, and miscarriage [5, 6] and has been identified in women with cervical cancer [7].

Author affiliations: [1]Associate Laboratory i4HB , Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [2]UCIBIO – Applied Molecular Biosciences Unit, Department of Biological Sciences, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [3]Centro Hospitalar de São João, Porto, Portugal; [4]Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP), Porto, Portugal.
***Correspondence:** Teresa Gonçalves Ribeiro, teresampg84@gmail.com; Luísa Peixe, lpeixe@ff.up.pt
Keywords: *Gardnerella vaginalis*; *Gardnerella pickettii* sp. nov.; *Gardnerella greenwoodii* sp. nov; whole genome sequencing.
Abbreviations: ANI, average nucleotide identity; BHI, brain heart infusion; *cpn60*, gene encoding chaperonin 60; dDDH, digital DNA–DNA hybridization; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; *nanH*, gene encoding sialidase; RiPPs, ribosomal synthesized and post-translationally modified peptides; TYGS, Type (Strain) Genome Server; *vly*, gene encoding vaginolysin.
The 16S rRNA gene sequence data from strains c17Ua_112^T and c31Ua_26^T are available in the GenBank database under accession numbers OP401218 and OP402842, respectively. The GenBank accession numbers for the whole genome sequences of strains c17Ua_112^T and c31Ua_26^T are JAKNCU000000000 and JAKNCL000000000, respectively.
One supplementary figure and two supplementary tables are available with the online version of this article.

Recently, *G. vaginalis*, *G. swidsinskii*, *G. leopoldii*, *G. piotii* and *Gardnerella* genomic species 3 were isolated from women with urinary symptoms (e.g., urgency urinary incontinence, overactive bladder symptoms and stress urinary incontinence) [8]. Extra-urogenital infections due to *Gardnerella* have been sporadically reported, being isolated from blood [9, 10], pleural fluid [11, 12] bronchoscopy [9] and disc space aspirates [13]. In animals, *G. vaginalis* has been identified in the genital tracts of mares, but not associated with infection [14, 15].

In recent studies, different species of *Gardnerella* have also been identified in the urogenital microbiome of healthy/asymptomatic women [4, 16], highlighting their ambiguous role as pathogen or as a commensal inhabiting the urogenital tract. Noteworthy, the role of different *Gardnerella* species in the human urogenital microbiome is still poorly understood, due to prior insufficient taxonomic fine-tuning, with all species being lumped in as *G. vaginalis*. Moreover, the delineation of *Gardnerella* genomic species without the designated type strains, whose phenotypes and genotypes should be characterized and described, also contributes to uncertainty about their beneficial or deleterious role in the human urogenital system.

During an ongoing project on the female urinary microbiome conducted in our research laboratory [16, 17], two strains representing putative new species of the genus *Gardnerella* were found. In this study, we aim to characterize these new species that compose the urinary microbiome of healthy women and female patients diagnosed with overactive bladder.

METHODS

Sample processing and strains isolation

Within an ongoing female urinary microbiome study, a set of female voided mid-stream urine samples was collected and analysed as described previously [17, 18]. Additionally, one sample from an asymptomatic postmenopausal woman and six female patients diagnosed with overactive bladder (pre- and postmenopausal women; unpublished data) were analysed based on the same protocol [17]. Briefly, 100 µl urine samples were inoculated on Columbia agar with 5% sheep blood (Biogerm) and chromogenic agar HiCrome UTI (HiMedia). After 48 h at 37°C under aerobic, microaerophilic and anaerobic conditions, colonies were reisolated, subjected to identification by MALDI-TOF MS analysis using *in vitro* diagnostic database version 3.0 (VITEK MS system, bioMérieux), and stored in tryptic soy broth (Liofilchem) supplemented with 20 % (v/v) glycerol at −80°C.

Strains initially identified as *G. vaginalis* by MALDI-TOF MS, were subjected to *cpn60* sequencing [16]. Two strains (c17Ua_112^T isolated from a woman diagnosed with overactive bladder and c31Ua_26^T isolated from a urine sample of a healthy postmenopausal woman) identified as putative new *Gardnerella* species were further characterized.

Growth characteristics

Growth at different temperatures (4, 8, 25, 30, 37, 40, 44, 50 and 55°C) was tested in Columbia agar with 5% sheep blood under anaerobic conditions, for 24, 48, 72 and 96 h. Moreover, growth in different media [Columbia agar with 5% sheep blood (bioMérieux), Mueller–Hinton agar (bioMérieux) with 5% sheep blood (ThermoFisher), Mueller–Hinton agar with 5% human blood (from a healthy female donor), and chocolate agar (Biogerm)] were evaluated after 48 h of incubation at 37°C under aerobic, microaerophilic and anaerobic conditions, and after 7 days. Growth at different NaCl concentrations [0.5, 1.5, 2, 2.5, 3.5, 6.5, 9.5, 12.5 and 15.5% (w/v)] and pH (pH 4.0, 4.5, 5.5, 7.5 and 8.0) were examined by using brain heart infusion (BHI) broth (Liofilchem) and supplemented BHI (addition of 0.5% yeast extract, 0.1% starch, 0.1% peptone and 2% gelatin), and were evaluated after 24, 48 and 72 h of incubation at 37 °C under anaerobic conditions.

Colony morphology and biochemical characterization

Colony morphology was observed under optimum growth conditions: cells grown on Columbia agar with 5% sheep blood at 37°C under microaerophilic and anaerobic conditions for 48 h. Gram-staining was carried out using Gram-staining kit (bioMérieux). Biochemical characterization was performed using the standardized API CORYNE (bioMérieux) and API RAPID ID 32 A (bioMérieux) kits following the manufacturer's instructions. β-Galactosidase was tested using ONPG tablets (Sigma Aldrich), and oxidase activity was tested using oxidase strips (ThermoFisher), according to manufacturer's recommendations. Sialidase activity was searched using a bacterial cell pellet in the qualitative filter paper spot test with fluorogenic substrate 2′-(4-methylumbelliferyl)-α-D-*N*-acetylneuraminic acid, as previously described [19].

MALDI-TOF MS analysis

Strains c17Ua_112^T and c31Ua_26^T were grown for 72 h at 37°C on Columbia agar with 5% sheep blood, under anaerobic conditions. Colonies were directly transferred and spotted onto a reusable 96-spot steel target (Bruker Daltonik). The spots were then overlaid with 1 µl matrix solution (α-cyano-4-hydroxycinnamic acid in 50% acetonitrile with 2.5% trifluoroacetic acid) and were air dried. Mass spectra were acquired with MALDI Biotyper Compass 4.1.100 data acquisition software (Bruker Daltonics). Each isolate was spotted in at least three technical replicates. All spectra were preprocessed using the FlexAnalysis software (Bruker Daltonics) applying the 'smoothing' and 'baseline subtraction' procedures. The raw spectral

data in text file format were imported and loaded in Mmass software (version 5.5.0) [20] to average spectra replicates from each strain and compared to the list of m/z markers for *Gardnerella* species presented by Vaneechoutte *et al.* [2], and rounded to the second decimal place.

Genomic analyses

Genomic DNA of strains c17Ua_112^T and c31Ua_26^T were submitted to Eurofins Genomics (Constance, Germany) for library preparation (NEBNext Ultra II FS DNA library prep kit, New England Biolabs) and whole-genome sequencing (NovaSeq 6000; Illumina) producing 2×150 bp paired-end reads. Reads were trimmed by Trim Galore 0.6.6 [21] in Bacterial and Viral Bioinformatics Resource Center [22, 23] and quality checked using FastQC 0.11.9 [21]. *De novo* assembly was performed by Unicycler version 0.4.8 [24], and quality assessed by QUAST version 5.0.2 [25]. Annotation of the draft genome was provided by the NCBI Prokaryotic Genome Annotation Pipeline [26].

Average nucleotide identity (ANI) based on BLAST+ was calculated by JSpeciesWS [27] and digital DNA–DNA hybridization (dDDH) by the Genome-to-Genome Distance Calculator 3.0, following the recommended Formula 2 [28, 29]. A whole-proteome phylogenomic analysis based on genome BLAST distance was inferred using the amino acid sequences of the entire proteome as input and conducted using the Type (Strain) Genome Server (TYGS) available at the DSMZ website [29].

The Web server MyDbFinder 2.0 [30] was used to extract nucleotide sequences of the *vly* (encoding vaginolysin) [31], *nanH1* (formerly *sldA*), *nanH2* and *nanH3* loci, associated with sialidase activity [32], from genomes of strains c17Ua_112^T and c31Ua_26^T. Further, comparisons with previously reported sequences were performed with the selected threshold of 70% identity and a selected minimum length of 60%.

Clustered Regularly Interspaced Short Palindromic Repeats and associated proteins (Cas), which formed the CRISPR-Cas system, were searched by using CRISPRcasFinder [33]. Only CRISPR showing an evidence level of 3 or 4 (highly likely candidates) [33] were considered. Prophages were identified using PHAge Search Tool Enhanced Release (PHASTER) [34, 35]. Antimicrobial resistance genes and plasmid replicons were investigated using ResFinder 4.1 and PlasmidFinder 2.1, respectively [36, 37]. BAGEL4 was used to mine bacteriocins and ribosomal synthesized and post-translationally modified peptides (RiPPs) [38].

Complete nucleotide sequences of genes *rrs* (16S rRNA) and *cpn60* (chaperonin 60) from genomes of strains c17Ua_112^T, c31Ua_26^T, type strains of *Gardnerella* species, and closely related genomic species were extracted with MyDbFinder 2.0 [30]. Alignment and similarity scores were generated using MEGA version 7.0 (www.megasoftware.net) [39]. Phylogenetic trees were reconstructed according to the neighbour-joining method [40], and genetic distances were estimated using Kimura's two-parameter model [41]. The reliability of internal branches was assessed from bootstrapping based on 1000 resamplings [42].

RESULTS

The two strains characterized in this study were able to grow on Columbia agar with 5% sheep blood in temperatures ranging from 30 to 40 °C. Both strains were able to grow in the different media tested under microaerophilic and anaerobic conditions, but not in aerobic conditions. Poor growth was observed on Mueller–Hinton with 5% human blood. Growth occurred in the presence of 0.5% NaCl (24, 48, 72 h), and at a pH 7.5 and 8 in BHI (optimum, pH 7.5; 24, 48 and 72 h) and pH 5.5, 7.5 and 8 in supplemented BHI (optimum, 7.5%; 24, 48 and 72 h). Moreover, both strains grew at a pH 4.5 in supplemented BHI after 48 and 72 h of incubation. Colonies of c17Ua_112^T were small, circular and opaque white/greyish, and colonies of c31Ua_26^T were very small, smooth, circular and light grey (almost transparent), both with α-haemolysis on Columbia agar with 5% sheep blood after 48 h of incubation at 37 °C in microaerophilic conditions.

Cells of strain c17Ua_112^T were coccobacilli and Gram-stain-negative, and those of strain c31Ua_26^T were coccobacilli and Gram-stain-positive (Fig. S1, available in the online version of this article).

Strains c17Ua_112^T and c31Ua_26^T were negative for β-galactosidase, catalase and oxidase activity, which was confirmed by the absence of corresponding genes in genome sequences. Strain c17Ua_112^T could use glucose, ribose and maltose. However, multiple repeated testing of API Coryne and API 20A systems for strain c31Ua_26^T always revealed negative results for the carbohydrates tested. The biochemical activity of the genus *Gardnerella* has been determined in different studies, which included a variable number of strains (13–127) and was summarized by Vaneechoutte *et al.* [2]. Different features of the strains analysed in our study in comparison with those systematized for the genus *Gardnerella* [2] are presented in Table 1.

The MALDI-TOF MS profiles of strains c17Ua_112^T and c31Ua_26^T differed between them, and between those of type strains of *Gardnerella* species, as summarized in Table 2.

Phylogenetic analysis based on 16S rRNA gene sequences revealed that strain c17Ua_112^T clustered with the type strains of *G. vaginalis*, *G. piotii* and strains from *Gardnerella* genomic species 3, while strain c31Ua_26^T clustered with type strains of *G. leopoldii*, *G. swidsinskii* and strains from *Gardnerella* genomic species 8 (Fig. 1).

Table 1. Differential characteristics of strains c17Ua_112^T and c31Ua_26^T compared with those of the genus *Gardnerella*

Characteristics	c17Ua_112^T	c31Ua_26^T	Genus *Gardnerella* [2]
Gram stain	Negative	Positive	Variable
Colony morphology	Small	Very small	ND*
Colony colour	Opaque white/greyish	Light grey	White to greyish
Growth at:			
30 °C	w*	w	+
40 °C	+	w	+
API test:			
Pyrazinamidase	+	−	+
Glucose	+	−	+
Ribose	+	−	+
Leucyl-glycine arylamidase	−	−	+
L-Phenylalanine arylamidase	−	−	+
L-Tyrosine arylamidase	−	−	+
Glycine arylamidase	−	−	+
L-Histidine arylamidase	−	−	+
L-Serine arylamidase	−	−	+

*ND, No data; w, weakly positive.

Table 2. MALDI-TOF MS fingerprints of strains c17Ua_112^T and c31Ua_26^T and type strains of *Gardnerella* species

m/z value	c17Ua_112^T	c31Ua_26^T	*G. vaginalis* ATCC 14018^T [2]	*G. leopoldii* UGent 06.41^T [2]	*G. piotii* UGent 18.01^T [2]	*G. swidsinskii* GS 9838- 1^T [2]
2704.75	NP*	NP	NP	NP	NP	10932
4422.12	NP	NP	10787	NP	NP	NP
4429.68	1479	3083	NP	6510	5878	4064
4899.09	NP	2077	NP	9250	NP	10013
4928.11	2958	NP	8800	NP	11 074	NP
5162.33	NP	854	3265	4429	NP	5502
5349.29	NP	NP	NP	12 766	NP	9105
6855.63	1179	NP	8925	NP	NP	NP
6885.02	NP	NP	NP	NP	7316	NP
8842.43	230	NP	7861	NP	NP	NP
8857.67	1022	1813	NP	4952	5185	2872
9795.73	NP	1249	NP	5955	NP	6344
9853.58	1893	NP	5455	NP	7923	NP

*The numbers in the matrix represent peak intensities. NP, No Peak.

Phylogenetic analysis based on *cpn60* sequences revealed that strain c17Ua_112^T and nine other strains, previously recognized as *Gardnerella* genomic species 3 [2], were grouped together and shared high similarity (99.4% nucleotide sequence similarity), representing a well-separated lineage supported by bootstrap values of 100%, distinct from other *Gardnerella* species (Fig. 2). *G. piotii* was the closest neighbour of strain c17Ua_112^T sharing 96.9% of sequence similarity, followed by *G. vaginalis* (88.0%), *G. swidsinskii* (86.3%) and *G. leopoldii* (84.9%).

Strain c31Ua_26^T and three other strains, previously recognized as *Gardnerella* genomic species 8 [2], were grouped together and shared high similarity (98.7% sequence similarity), representing a well-separated lineage supported by bootstrap values of 100%, distinct

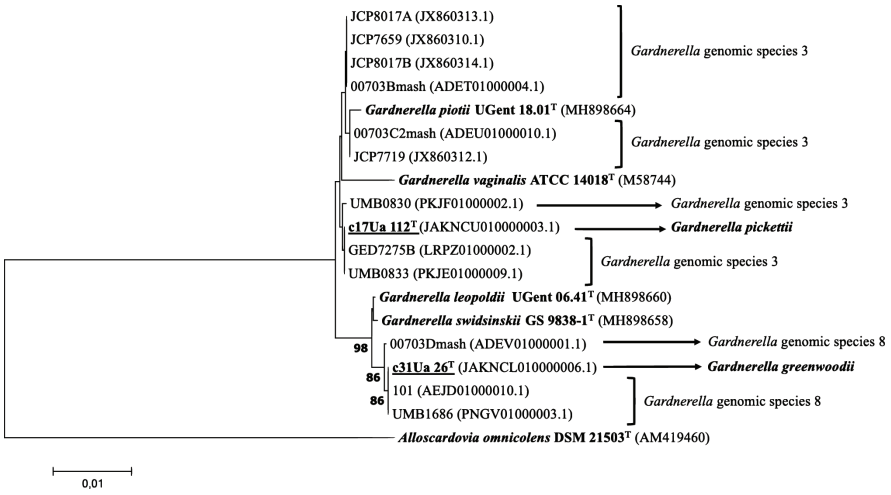


Fig. 1. Neighbour-joining tree based on 16S rRNA gene sequences showing the phylogenetic relationships of *Gardnerella pickettii* c17Ua_112^T (boldface type and underlined), *Gardnerella greenwoodii* c31Ua_26^T (boldface type and underlined), type strains of the genus *Gardnerella* (boldface type denoted by 'T' at the end of the strain name), and closely related genomic species. *Alloscardovia omnicolens* was used as the outgroup. Bootstrap percentages (based on 1000 replications) are shown at nodes. Only values above 80% are shown. Bar, 0.01 substitutions per nucleotide position. The available accession numbers of the sequences used are indicated in parentheses.

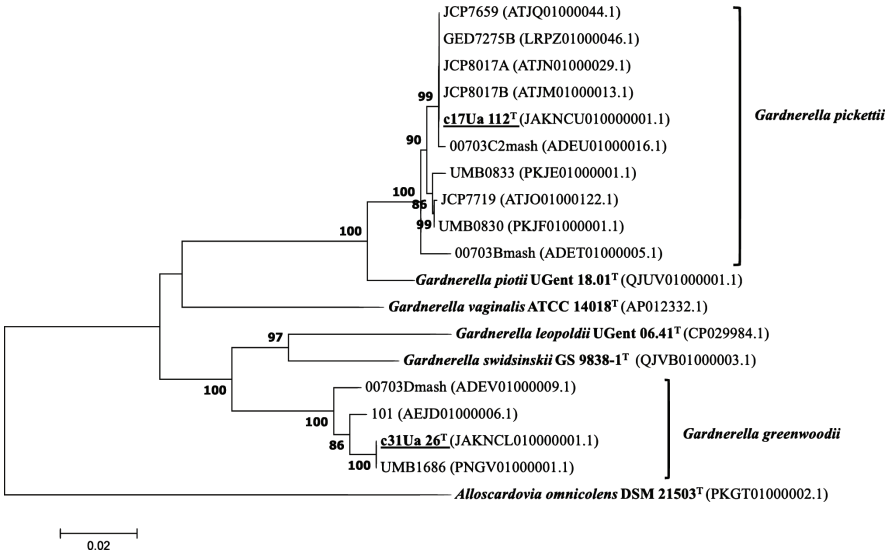


Fig. 2. Neighbour-joining tree based on complete *cpn60*-gene sequences (1626 bp) showing the phylogenetic relationships of *Gardnerella pickettii* c17Ua_112^T (boldface type and underlined), *Gardnerella greenwoodii* c31Ua_26^T (boldface type and underlined), type strains of the genus *Gardnerella* (boldface type denoted by 'T' at the end of the strain name), and closely related genomic species. *Alloscardovia omnicolens* was used as the outgroup. Bootstrap percentages (based on 1000 replications) are shown at nodes. Only values above 80% are shown. Bar, 0.02 substitutions per nucleotide position. The available accession numbers of the sequences used, are indicated in parentheses.

from other *Gardnerella* species (Fig. 2). The closest neighbour of strain c31Ua_26 ^{T} was *G. swidsinskii* (93.0% of sequence similarity), and the sequence similarity with *G. leopoldii*, *G. vaginalis* and *G. piotii* was 89.9, 89.0 and 87.3%, respectively.

The genome of the strain c17Ua_112^{T} was assembled into three contigs of total length of 1.51 Mbp and G+C content of 42.5mol%, and genome of the strain c31Ua_26^{T} was assembled into 11 contigs of total length of 1.51 Mbp and G+C content of 43.3mol%. A summary of genomic features of strains c17Ua_112^{T} and c31Ua_26^{T} is shown in Table 3.

The ANI and dDDH values between strains c17Ua_112^{T} and c31Ua_26^{T}, and between strains c17Ua_112^{T}, c31Ua_26^{T} and type strains of the closest related *Gardnerella* species were below the species cut-off levels (96% and 70%, respectively) [2]. The ANI values between c17Ua_112^{T} and strains belonging to *Gardnerella* genomic species three was ≥96.4% (Table S1), and between c17Ua_112^{T} and the closest related type strain *G. piotii* UGent 18.01^{T} was 94.8% (Table S2). The ANI values between c31Ua_26^{T} and strains belonging to *Gardnerella* genomic species 8 was ≥96.4% (Table S1), and between c31Ua_26^{T} and the closest related type strain *G. swidsinskii* GS 9838-1^{T} was 84.2% (Table S2).

The dDDH values between c17Ua_112^{T} and strains belonging to *Gardnerella* genomic species 3 was ≥70.7%, and between c17Ua_112^{T} and the closest related type strain *G. piotii* UGent 18.01^{T} was 59.8%. The dDDH values between c31Ua_26^{T} and strains belonging to *Gardnerella* genomic species 8 was ≥71.0%, and between c31Ua_26^{T} and the closest related type strain *G. swidsinskii* GS 9838-1^{T} was 29.1%.

The whole proteome tree supported the existence of six distinct *Gardnerella* species, with strain c17Ua_112^{T} clustering with strains belonging to *Gardnerella* genomic species 3, and strain c31Ua_26^{T} with strains belonging to *Gardnerella* genomic species 8 (Fig. 3).

The gene *vly,* encoding vaginolysin, was detected in strains c17Ua_112^{T} and c31Ua_26^{T}, with 91.8% of nucleotide similarity between them, and 90.2% and 91.6% of sequence similarity to *vly* from *G. vaginalis* strain 14018^{T} (accession number: EU522486), respectively. According to sequence differences in the undecapeptide region of VLY [43], strain c17Ua_112^{T} comprised a VLY Type 1A (valines at positions 1 and 6), and c31Ua_26^{T} a VLY Type 1B (valine at position one and alanine at position 6). The presence of genes encoding sialidase was variable, with *nanH1* being identified in both strains, *nanH3* in c17Ua_112^{T}, and *nanH2* not detected. However, in phenotypic tests, strain c17Ua_112^{T} was the only demonstrating sialidase activity.

Antimicrobial resistance genes were identified in the genome of both strains. The *msr*(D) in tandem with *mef*(A) efflux pump, coding for macrolide resistance, were detected in strains c17Ua_112^{T} and c31Ua_26^{T} (100% nucleotide identity with the reference gene accession number: AF274302 and 99.9% nucleotide identity with the reference gene accession number: U83667, respectively), yet premature stop codons were identified in the amino acid sequences of *mef*(A) gene. Both gene sequences were used as queries to interrogate the database of 226 *Gardnerella* genomes (accessed in March 2023), being identified in only two genomes, i.e., *G. swidsinskii* 409–05 (accession: CP001849 [2]) and *G. leopoldii* UGent 06.41^{T} (accession: CP029984). Additionally, *lsa*(C) coding for cross-resistance to lincosamides, streptogramins A and pleuromutilins (91.5–91.6% nucleotide identity with the reference gene, accession number: HM990671) were detected in both strains, and in 17 genomes of *Gardnerella* strains correctly identified at species level by dDDH (14 *G. vaginalis*, one *G. leopoldii*, one *G. swidsinskii*, one *G. piotii*).

Strain c31Ua_26^{T} harboured one CRISPR-Cas system type IE, with a total of eight spacers (sequence length: 644,135). However, the spacer sequences remain unknown and no match was found for the nucleic acid source on the database. Complete prophages, plasmids, bacteriocins and RiPPs were not detected.

DISCUSSION

Strains c17Ua_112^{T} and c31Ua_26^{T} resemble the genus *Gardnerella* in terms of cell size and morphology, colony appearance, oxygen and temperature requirements, indole-, catalase-, oxidase- and urease-negative results, and DNA G+C content [2, 44]. However, some biochemical reactions were not in agreement with that provided by Greenwood and Pickett [44] for the genus *Gardnerella*, including the

Table 3. Genome assembly statistics and annotation features of strains c17Ua_112^{T} and c31Ua_26^{T}

Genome features	c17Ua_112^{T}	c31Ua_26^{T}
Genome status	draft	draft
Total assembly size (bp)	1509345	1511834
No. of contigs	3	11
G+C content (mol%)	42.5	43.3
No. of coding sequences	1132	1511
No. of predicted RNAs	3	3
No. of predicted tRNAs	45	45

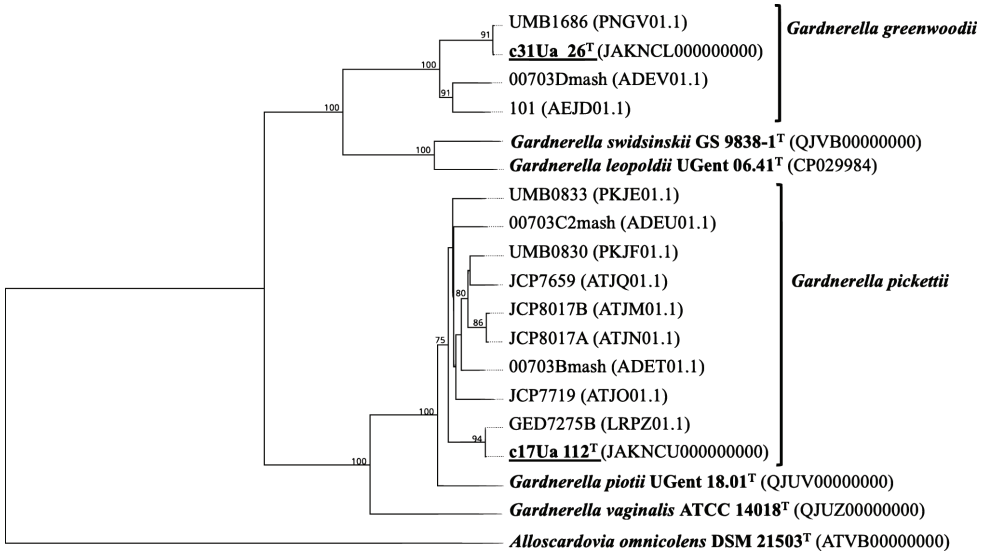


Fig. 3. Whole-proteome based GBDP distances tree showing the phylogenetic relationships between *Gardnerella greenwoodii* c31Ua_26^T (boldface type and underlined), *Gardnerella pickettii* c17Ua_112^T (boldface type and underlined), type strains of the genus *Gardnerella* (denoted by 'T' at the end of the strain name), and closely related genomic species. *Alloscardovia omnicolens* was used as the outgroup. The proteomic tree was conducted using the TYGS available at the DSMZ website [29], and created with FastME 2.1.6.1. The branch lengths are scaled via GBDP distance under the algorithm 'GreedyWithTrimming' and distance formula d5 [28]. The numbers at the branches are GBDP pseudo-bootstrap support values of >60% from 100 replications, with an average branch support of 75.5%. The tree was rooted at the midpoint. Whole genome sequence accession numbers are listed in parentheses.

inability of strain c31Ua_26^T to produce acid from glucose, maltose, and ribose, and the fact that both strains were unable to produce acid from mannose. Nevertheless, biochemical reactions can vary within *Gardnerella* strains [45], limiting the use of this method to achieve an accurate identification of *Gardnerella* species.

The differences in Gram staining reported for our strains resemble those typically associated with the genus *Gardnerella* (cells are Gram-negative to Gram-variable coccobacilli) [2]. The Gram stain reaction of *Gardnerella* has been a subject of controversy since its first description. While some authors described *Gardnerella* species as Gram-negative to Gram-variable, meaning it can appear either Gram-positive or Gram-negative under the microscope due to the varying visibility of its thin cell wall [2, 46], others suggested that it has a very thin cell wall (8–12 nm) with a characteristic Gram-negative staining pattern [47]. However, its taxonomic status as belonging to Gram-positive bacteria is clear, based on previous electron microscopy analysis of the cell wall structure [47, 48], the absence of diaminopimelic acid from the peptidoglycan layer [49] and of lipopolysaccharide from the cell wall [47]. Moreover, several other bacteria have been described as staining Gram-negative but possessing a Gram-positive cell-wall structure, e.g. *Caproicibacter fermentans* EA1^T, *Butyrivibrio fibrisolvens* D1, *Clostridium polysaccharolyticum* ATCC 33142^T, *Lachnoclostridium phytofermentans* ATCC 700394^T, *Cellulosilyticum lentocellum* NCIMB 11756^T and *Clostridium populeti* ATCC 35295^T [50].

MALDI-TOF MS protein profiles revealed that c17Ua_112^T and c31Ua_26^T presented different peaks between them and between those previously described for *Gardnerella* species, which suggests that these strains may represent two novel species. However, the unambiguously differentiation of strains c17Ua_112^T and c31Ua_26^T from all remaining *Gardnerella* species by MALDI-TOF MS can only be achieved with the inclusion of more strains per species to confirm that signature peaks are conserved and species specific.

The ANI and dDDH values between strains c17Ua_112^T, c31Ua_26^T, and type strains of *Gardnerella* species and between both strains were clearly below the recommended cut-offs for species delineation (96% for ANI [51], 70% for dDDH [28]), supporting our findings of strains c17Ua_112^T and c31Ua_26^T as representing two novel species. Likewise, the whole-proteome tree supported the identification of two novel *Gardnerella* species. Moreover, publicly available genomes of strains UMB1686, 00703Dmash and 101 deposited as *G. vaginalis* should be reclassified as *Gardnerella greenwoodii*, and strains UMB0833, 00703C2mash, UMB0830, JCP7659, JCP8017B, JCP8017A, 00703Bmash, JCP7719 and GED7275B as *Gardnerella pickettii* based on whole-genome relatedness.

Sialidase activity in the genus *Gardnerella* is mediated by two enzymes, NanH2 and NanH3, which are known to facilitate the destruction of the protective mucus layer on the vaginal epithelium [32]. We observed that the presence of NanH3 in strain c17Ua_112^T was associated with sialidase activity, while the presence of NanH1 in strain c31Ua_26^T was not, which is in agreement with recent studies contradicting the identification of NanH1 with sialidase activity [32, 52]. The degree of expression and substrate specificity of NanH3 by c17Ua_112^T was not evaluated in this study, yet it is important to clarify its role on the possible connection with overactive bladder syndrome.

The identification of the gene encoding for a type 1 cholesterol-dependent cytolysin (VLY, vaginolysin), able to cause cytotoxicity on vaginal epithelium [31] in both strains, is in line with previous findings reporting type 1 VLY in genomic species 3 (*G. pickettii*) and 8 (*G. greenwodii*) [43].

Noteworthy, the comprehensive genomic analysis of the two new *Gardnerella* species also revealed the presence of antibiotic-resistance genes. The *lsaC*, commonly found in the genomes of different *Gardnerella* strains [53], may contribute to a poor response to clindamycin treatment, as this antibiotic belongs to the lincosamide class and is one of the most used to treat genital infections [54]. Other genes such as *mef*(A) and *mrs*(D), known to contribute to the macrolide efflux resistance, have been previously identified in Gram-negative and Gram-positive bacteria, including *Gardnerella* species [55]. However, the *mef*(A) gene in tandem with *msr*(D) is barely described [55].

DESCRIPTION OF *GARDNERELLA PICKETTII* SP. NOV.

Gardnerella pickettii (pi.cke'ti.i. N.L. gen. masc. n. *pickettii* of Pickett, named after M. John Pickett, an American microbiologist who proposed the transfer of *Haemophilus vaginalis* to the new genus *Gardnerella*).

Cells are Gram-stain-negative coccobacilli and 0.79 µm long. They are facultative anaerobic and microaerophilic. Growth occurs on Columbia agar with 5% sheep blood, Mueller–Hinton agar with 5% sheep blood, Mueller–Hinton agar with 5% human blood, and chocolate agar after 48–72 h of incubation, and between 30–40°C (optimum, 37°C). Colonies are small, circular, opaque white/greyish with α-haemolysis. Positive for sialidase activity and negative for β-galactosidase.

The type strain is c17Ua_112^T (=DSM 113414^T=CCP 71^T), isolated from the urine of a woman diagnosed with overactive bladder syndrome in Portugal, 2017. The G+C content of the type strain is 42.5mol%. The 16S rRNA gene sequence data from c17Ua_112^T is available in the GenBank database under accession number OP401218. The whole genome sequence of the type strain is available from NCBI under accession number JAKNCU000000000.

Other strains that did not originate from this study are available from the NCBI database: JCP7659, GED7275B, JCP8017A, JCP8017B, 00703C2mash, UMB0833, JCP7719, UMB0830 and 00703Bmash.

DESCRIPTION OF *GARDNERELLA GREENWOODII* SP. NOV.

Gardnerella greenwoodii (green.woo'di.i. N.L. gen. masc. n. *greenwoodii* of Greenwood, named after James Greenwood, an American microbiologist who proposed the transfer of *Haemophilus vaginalis* to the new genus *Gardnerella*).

Cells are Gram-stain-positive coccobacilli and 0.94 µm long. They are facultative anaerobic and microaerophilic. Growth occurs on Columbia agar with 5% sheep blood and Mueller–Hinton agar with 5% sheep blood, Mueller–Hinton agar with 5% human blood, and chocolate agar after 48–72 h of incubation, and between 30–40°C (optimum 37°C). Colonies are very small, circular, opaque greyish with α-haemolysis. Negative for sialidase activity and β-galactosidase.

The type strain is c31Ua_26^T (=DSM 113415^T=CCP 72^T), isolated from urine of a healthy postmenopausal woman in Portugal, 2019. The G+C content of the type strain is 43.3mol%. The 16S rRNA gene sequence data from strain c31Ua_26^T is available in the GenBank database under accession number OP402842. The whole genome sequence of the type strain is available from NCBI under accession number JAKNCL000000000.

Other strains that did not originate from this study are available from the NCBI database: 00703Dmash, 101 and UMB1686.

Funding information
This work is 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 Applied Molecular Biosciences—UCIBIO and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy—i4HB. M.S., M.K. and F.G. were supported by FCT (SFRH/BD/05038/2020, SFRH/BD/132497/2017, DL57/2016/CP1346/CT0034, respectively), T.G.R. by Unidade de Ciências Biomoleculares Aplicadas (UIDP/QUI/04378/2020), with the financial support of the FCT/ MCTES through national funds, and S.U.P. by NORTH-01–0145-FEDER-000024.

Acknowledgements
We thank Carla Campos (Instituto Português de Oncologia do Porto, Portugal) for the acquisition of mass spectra on the MALDI-TOF MS device. BioMérieux Lda. (Portugal) provided material for bacterial culture and MALDI-TOF MS analysis and had no role in the study design, data collection and analysis, decision to publish, or writing of the manuscript.

Author contributions
M.S. performed the phenotypic, biochemical and genomic characterization, and wrote the original draft of the manuscript. M.K. contributed to strain isolation. S.U.P. performed the initial identification of strains by MALDI-TOF MS. T.A.L. contributed to study design and reviewing the manuscript. F.G. contributed to reviewing the manuscript. T.G.R. designed the study, contributed to reviewing the manuscript, and supervised. L.P. contributed to reviewing the manuscript, project design, administration, and funding.

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

Ethical statement
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.

References

1. Stackebrandt E, Rainey FA, Ward-rainey NL. Proposal for a new hierarchic classification system, *Actinobacteria classis* nov. *Int J Syst Bacteriol* 1997;47:479–491.
2. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, *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* 2019;69:679–687.
3. Swidsinski A, Mendling W, Loening-Baucke V, Ladhoff A, Swidsinski S, *et al*. Adherent biofilms in bacterial vaginosis. *Obstet Gynecol* 2005;106:1013–1023.
4. Castro J, Jefferson KK, Cerca N. Genetic heterogeneity and taxonomic diversity among *Gardnerella* species. *Trends Microbiol* 2020;28:202–211.
5. Hay PE, Morgan DJ, Ison CA, Bhide SA, Romney M, *et al*. A longitudinal study of bacterial vaginosis during pregnancy. *BJOG* 1994;101:1048–1053.
6. Menard JP, Mazouni C, Salem-Cherif I, Fenollar F, Raoult D, *et al*. High vaginal concentrations of *Atopobium vaginae* and *Gardnerella vaginalis* in women undergoing preterm labor. *Obstet Gynecol* 2010;115:134–140.
7. Kovachev SM. Cervical cancer and vaginal microbiota changes. *Arch Microbiol* 2020;202:323–327.
8. Putonti C, Thomas-White K, Crum E, Hilt EE, Price TK, *et al*. Genome investigation of urinary *Gardnerella* strains and their relationship to isolates of the vaginal microbiota. *mSphere* 2021;6:e00154-21.
9. Legrand JC, Alewaeters A, Leenaerts L, Gilbert P, Labbe M, *et al*. *Gardnerella vaginalis* bacteremia from pulmonary abscess in a male alcohol abuser. *J Clin Microbiol* 1989;27:1132–1134.
10. Lagacé-Wiens PRS, Ng B, Reimer A, Burdz T, Wiebe D, *et al*. *Gardnerella vaginalis* bacteremia in a previously healthy man: case report and characterization of the isolate. *J Clin Microbiol* 2008;46:804–806.
11. Calvert LD, Collins M, Bateman JRM. Multiple abscesses caused by *Gardnerella vaginalis* in an immunocompetent man. *J Infect* 2005;51:E27–9.
12. Murray L, Halpin J, Casserly B, O'Connell NH, Scanlon T. A pyohydropneumothorax with sepsis, secondary to *Gardnerella vaginalis* infection in a post-partum female. *Respir Med Case Rep* 2019;26:189–192.
13. Graham S, Howes C, Dunsmuir R, Sandoe J. Vertebral osteomyelitis and discitis due to *Gardnerella vaginalis*. *J Med Microbiol* 2009;58:1382–1384.
14. Salmon SA, Walker RD, Carleton CL, Robinson BE. Isolation of *Gardnerella vaginalis* from the reproductive tract of four mares. *J Vet Diagn Invest* 1990;2:167–170.
15. Salmon SA, Walker RD, Carleton CL, Shah S, Robinson BE. Characterization of *Gardnerella vaginalis* and *G. vaginalis*-like organisms from the reproductive tract of the mare. *J Clin Microbiol* 1991;29:1157–1161.
16. 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.
17. Ugarcina Perovic S, Ksiezarek M, Rocha J, Cappelli EA, Sousa M, *et al*. Urinary microbiome of reproductive-age asymptomatic European women. *Microbiol Spectr* 2022;10:e0130822.
18. Perovic SU, Ksiezarek M, Rocha J, Vale L, Silva C, *et al*. Time to change microbiological approach to overactive bladder. *Eur Urol Suppl* 2019;18:e103.
19. Briselden AM, Moncla BJ, Stevens CE, Hillier SL. Sialidases (neuraminidases) in bacterial vaginosis and bacterial vaginosis-associated microflora. *J Clin Microbiol* 1992;30:663–666.
20. Martin Strohalm. mMass - Open Source Mass Spectrometry Tool; (n.d.). http://www.mmass.org/
21. Babraham Institute. Bioinformatics - Trim Galore!; (n.d.). https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
22. Bioinformatics Resource Centers (BRCs). Bacterial and Viral Bioinformatics Resource Center | BV-BRC; (n.d.). https://www.bv-brc.org/
23. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, *et al*. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res* 2023;51:D678–D689.
24. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
25. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
26. 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.
27. 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.
28. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:1–14.
29. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 2019;10:2182.
30. CGE Server; (n.d.). https://cge.food.dtu.dk/services/MyDbFinder/
31. Gelber SE, Aguilar JL, Lewis KLT, Ratner AJ. Functional and phylogenetic characterization of vaginolysin, the human-specific cytolysin from *Gardnerella vaginalis*. *J Bacteriol* 2008;190:3896–3903.
32. Robinson LS, Schwebke J, Lewis WG, Lewis AL. Identification and characterization of NanH2 and NanH3, enzymes responsible for

sialidase activity in the vaginal bacterium *Gardnerella vaginalis*. *J Biol Chem* 2019;294:5230–5245.

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. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res* 2011;39:W347–52.
35. 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–W21.
36. 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.
37. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, *et al*. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58:3895–3903.
38. 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 Res* 2018;46:W278–W281.
39. Kumar S, Stecher G, Tamura K, Dudley J. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
40. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425.
41. 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.
42. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39:783–791.
43. Garcia EM, Serrano MG, Edupuganti L, Edwards DJ, Buck GA, *et al*. Sequence comparison of vaginolysin from different *Gardnerella* species. *Pathogens* 2021;10:86.
44. Greenwood JR, Pickett MJ. Transfer of *Haemophilus vaginalis* Gardner and Dukes to a new genus, *Gardnerella*: *G. vaginalis* (Gardner and Dukes) comb. nov. *Int J Syst Bacteriol* 1980;30:170–178.
45. Catlin BW. *Gardnerella vaginalis*: characteristics, clinical considerations, and controversies. *Clin Microbiol Rev* 1992;5:213–237.
46. Turovskiy Y, Sutyak Noll K, Chikindas ML. The aetiology of bacterial vaginosis. *J Appl Microbiol* 2011;110:1105–1128.
47. Sadhu K, Domingue PA, Chow AW, Nelligan J, Cheng N, *et al*. *Gardnerella vaginalis* has a Gram-positive cell-wall ultrastructure and lacks classical cell-wall lipopolysaccharide. *J Med Microbiol* 1989;29:229–235.
48. Reyn A, Birch-Andersen A, Lapage SP. An electron microscope study of thin sections of *Haemophilus vaginalis* (Gardner and Dukes) and some possibly related species. *Can J Microbiol* 1966;12:1125–1136.
49. Harper JJ, Davis GHG. Cell wall analysis of *Gardnerella vaginalis* (*Haemophilus vaginalis*). *Int J Syst Bacteriol* 1982;32:48–50.
50. Flaiz M, Baur T, Brahner S, Poehlein A, Daniel R, *et al*. *Caproicibacter fermentans* gen. nov., sp. nov., a new caproate-producing bacterium and emended description of the genus *Caproiciproducens*. *Int J Syst Evol Microbiol* 2020;70:4269–4279.
51. Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, *et al*. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol* 2018;68:2386–2392.
52. Kurukulasuriya SP, Patterson MH, Hill JE. Slipped-strand mispairing in the gene encoding sialidase NanH3 in *Gardnerella* spp. *Infect Immun* 2021;89:e00583-20.
53. Muzny CA, Sobel JD. The role of antimicrobial resistance in refractory and recurrent bacterial vaginosis and current recommendations for treatment. *Antibiotics* 2022;11:500.
54. Murphy CS, McKay G. Clindamycin. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2022. pp. 1–4.
55. Iannelli F, Santoro F, Santagati M, Docquier J-D, Lazzeri E, *et al*. Type M resistance to macrolides is due to a two-gene efflux transport system of the ATP-binding cassette (ABC) superfamily. *Front Microbiol* 2018;9:1670.

Supplementary Material

Gardnerella pickettii sp. nov. (formerly *Gardnerella* genomic species 3) and *Gardnerella greenwoodii* sp. nov. (formerly *Gardnerella* genomic species 8) isolated from female urinary microbiome

Márcia Sousa,[1,2] Magdalena Ksiezarek,[1,2] Svetlana Ugarcina Perovic,[1,2] Tiago Antunes-Lopes,[3] Filipa Grosso,[1,2] Teresa Gonçalves Ribeiro,[1,2,4*] Luísa Peixe[1,2,4*]

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

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

[3]Centro Hospitalar de São João, Porto, Portugal

[4]Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP).

Corresponding author and email address:

Teresa Gonçalves Ribeiro, teresampg84@gmail.com

Luísa Peixe, lpeixe@ff.up.pt

Table S1. Average nucleotide identity between the strains analysed in this study and 9 publicly available genomes of *Gardnerella* genomic species 3, and 3 publicly available genomes of *Gardnerella* genomic species 8.

Strain	GenBank name	GenBank Accession numbers	Host	Source	Symptoms/ Diagnoses	c17Ua_112^T	c31Ua_26^T
00703Bmash	*Gardnerella vaginalis* 00703Bmash	GCA_000263615.1	*Homo sapiens*	Vagina	Bacterial vaginosis and HSV-2[a]	96.85	
00703C2mash	*Gardnerella vaginalis* 00703C2mash	GCA_000263515.1		Vagina	Bacterial vaginosis and HSV-2[a]	96.68	
JCP8017B	*Gardnerella vaginalis* JCP8017B	GCA_000414585.1		Vagina	Bacterial vaginosis[b]	96.72	
JCP8017A	*Gardnerella vaginalis* JCP8017A	GCA_000414605.1		Vagina	Bacterial vaginosis[b]	96.74	
JCP7719	*Gardnerella vaginalis* JCP7719	GCA_000414625.1		Vagina	Bacterial vaginosis[b]	96.42	
JCP7659	*Gardnerella vaginalis* JCP7659	GCA_000414665.1		Vagina	Bacterial vaginosis[b]	96.54	
GED7275B	*Gardnerella vaginalis*	GCA_001546445.1		Vagina	Asymptomatic[c]	99.99	
UMB0833	*Gardnerella vaginalis*	GCA_002861885.1		Urine	Asymptomatic (pregnant woman)[d]	96.67	
UMB0830	*Gardnerella vaginalis*	GCA_002861905.1		Urine	Asymptomatic (pregnant woman)[d]	96.68	
00703Dmash	*Gardnerella vaginalis* 00703Dmash	GCA_000263635.1	*Homo sapiens*	Vagina	Bacterial vaginosis and HSV-2[a]		96.65
101	*Gardnerella vaginalis* 101	GCA_000165615.2		Vagina	Bacterial vaginosis[e]		96.39
UMB1686	*Gardnerella vaginalis*	GCA_002884775.1		Urine	Diabetes[f]		99.95

[a] Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, Powell E, Janto B, Eutsey R, Hiller NL, Boissy R, Dahlgren ME, Hall BG, Costerton JW, Post JC, Hu FZ, Ehrlich GD. Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. J Bacteriol. 2012 Aug;194(15):3922-37. doi: 10.1128/JB.00056-12. Epub 2012 May 18. PMID: 22609915; PMCID: PMC3416530.

[b] Lewis, W. G., Robinson, L. S., Gilbert, N. M., Perry, J. C., & Lewis, A. L. (2013). Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted actinobacterium *Gardnerella vaginalis*. Journal of Biological Chemistry, 288(17), 12067–12079. https://doi.org/10.1074/JBC.M113.453654.

[c] Bohr, L. L., Mortimer, T. D., & Pepperell, C. S. (2020). Lateral Gene Transfer Shapes Diversity of Gardnerella spp. Frontiers in Cellular and Infection Microbiology, 10. https://doi.org/10.3389/FCIMB.2020.00293.

[d] 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).

[e] Castro, J., Henriques, A., Machado, A., Henriques, M., Jefferson, K. K., & Cerca, N. (2013). Reciprocal Interference between Lactobacillus spp. and *Gardnerella vaginalis* on Initial Adherence to Epithelial Cells. Int. J. Med. Sci, 10. https://doi.org/10.7150/ijms.6304

[f] Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, Hilt EE, Price TK, Wolfe AJ, Lawley TD. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. Nat Commun. 2018 Apr 19;9(1):1557. doi: 10.1038/s41467-018-03968-5. PMID: 29674608; PMCID: PMC5908796.

Table S2. Genomic features of the *Gardnerella* strains analysed in this study.

Strain	Species	GenBank Accession numbers	G+C (%)	Genome size (Mbp)	Number of Contigs	tRNA	Proteins	CDS	ANI/ dDDH (%)			
									Gardnerella vaginalis ATCC 14018^{T}	*Gardnerella piotii* UGent 18.01^{T}	*Gardnerella leopoldii* UGent 06.41^{T}	*Gardnerella swidsinskii* GS 9838-1^{T}
c17Ua_112^{T}	*Gardnerella pickettii*	JAKNCU000000000.1	42.5	1.51	3	45	1144	1151	87.16/ 33.0	94.82/ 59.8	78.76/ 26.8	78.90/ 27.2
c31Ua_26^{T}	*Gardnerella greenwoodii*	JAKNCL000000000.1	43.3	1.51	11	45	1124	1132	78.97/ 26.0	78.63/ 26.4	84.09/ 28.3	84.22/ 29.1

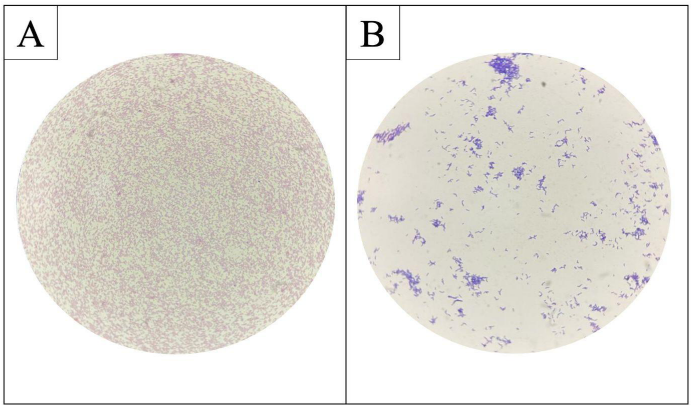


Figure S1. Gram-staining of the *Gardnerella* strains analysed in this study. A) *Gardnerella pickettii* sp. nov. (formerly *Gardnerella* genomic species 3). B) *Gardnerella greenwoodii* sp. nov. (formerly *Gardnerella* genomic species 8).

The *Gardnerella* Sialidases Repertoire: Diversity, Structure, and Function in the Urogenital Microbiome

Márcia Sousa,[1,2] Benedita Pinheiro,[3,4] Pedro Teixeira,[1,2#] Angelina Palma,[3,4] Teresa Gonçalves Ribeiro,[1,2,5] Filipa Grosso,[1,2] Luísa Peixe [1,2,5 *]

[1]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [2]UCIBIO – Applied Molecular Biosciences Unit, Department of Biological Sciences, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [3]UCIBIO, Applied Molecular Biosciences Unit, Department of Chemistry, School of Science and Technology, NOVA University Lisbon, Caparica, Portugal; [4]Associate Laboratory i4HB - Institute for Health and Bioeconomy, School of Science and Technology, NOVA University Lisbon, Caparica, Portugal; [5]Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP).

[#]current affiliation: National Reference Laboratory of Antibiotic Resistances and Healthcare Associated Infections, Department of Infectious Diseases, National Institute of Health Dr. Ricardo Jorge, 1649-016 Lisbon, Portugal

*Luísa Peixe, lpeixe@ff.up.pt

Manuscript in preparation

Abstract

Gardnerella spp., prevalent members of the female urogenital microbiome, express sialidases, enzymes implicated in bacterial vaginosis (BV) and urogenital dysbiosis. This study investigates the diversity, genomic context, and structural analysis of four *Gardnerella* sialidases (NanH1-NanH4) across 153 *Gardnerella* genomes from healthy and dysbiotic urogenital samples. Genomic analysis revealed *nanH1* and *nanH3* as the most widespread sialidase genes, while *nanH2* was less frequent and more associated with dysbiosis. We provide the first functional analysis of NanH4, previously predicted to be inactive, and the analysis of the genomic organization of all four *nanH* gene clusters. Phylogenetic analysis demonstrated distinct evolutionary relationships among the sialidases: a close relationship between NanH2 and NanH3, both proposed as primary sources of sialidase activity, while NanH4 formed a distinct cluster more closely related to NanH1. Structural models predicted conserved catalytic residues across all sialidases, with variations in interacting residues suggesting distinct substrate specificities. NanH1 and NanH4 lacked signal peptides and transmembrane domains, indicating intracellular localization, while NanH2 and NanH3 exhibited features suggesting surface tethering or secretion, potentially influencing host interactions and pathogenicity.

In vitro assays demonstrated sialidase activity in selected *Gardnerella* strains, with varying susceptibility to the inhibitors DANA and Zanamivir. Crude extracts containing NanH3 showed sialidase activity, efficiently cleaving α2-6-linked sialic acids from both bovine submaxillary gland mucin and fetuin. In contrast, extracts containing NanH1 and NanH4 displayed residual sialidase activity. Inhibition assays showed that DANA and Zanamivir effectively suppressed NanH3 activity, while the inhibition of NanH1 and NanH4 was less pronounced, potentially due to differences in enzyme structure and intracellular localization.

This study highlights the genomic and functional diversity of *Gardnerella* sialidases, offers a structural basis for understanding their catalytic mechanisms, and suggests potential therapeutic strategies for modulating sialidase activity in the urogenital tract. However, the complexity of sialidase activity and its potential regulation by host factors highlights the uncertainty of simply correlating sialidase presence with virulence. Future research should focus on characterizing the specific substrates of *Gardnerella* sialidases, examining their interactions with host immune components, and investigating their contribution to polymicrobial interactions within the vaginal microbiome to gain a comprehensive understanding of their role in *Gardnerella* virulence and overall urogenital health.

Keywords: *Gardnerella* spp., urogenital microbiome, sialidases, genetic diversity

Introduction

The *Gardnerella* genus, historically associated with bacterial vaginosis (BV), underwent a significant taxonomic revision in 2019, expanding the understanding of the genus and its diversity (1). However, the complex role of *Gardnerella* spp. in the female urogenital microbiome extends beyond their association with BV. In fact, these bacteria have been identified in women with various urogenital tract dysbiosis, including urinary tract infections and overactive bladder, but also in healthy women (2,3).

Sialidase activity in vaginal fluids remains a critical diagnostic marker for BV, but our knowledge of these enzymes, particularly in *Gardnerella* spp. remains limited. These enzymes are glycosidases and contribute to the degradation of the vaginal mucus by hydrolyzing terminal sialic acids, such as N-acetyl-neuraminic acid (Neu5Ac), from glycans, which provides a carbon source for the local microbiota (4). Sialidase virulence is associated with the hydrolysis of α2-3 or α2-6-linked sialic acids on N-linked (asparagine-linked) and O-linked (serine- or threonine-linked) (5) glycans, found in mucin glycoproteins produced by mucosal epithelia of urogenital, gastrointestinal, and respiratory tracts (6). Based on their location, substrate specificity and catalytic mechanism, sialidases can be hydrolytic, *trans*-sialidases and intramolecular *trans*-sialidase (IT-sialidase) (7).

The first sialidase gene identified in *Gardnerella* spp. was *nanH1* (previously *sldA*, sialidase A) (8). However, NanH1 is present in strains both with and without detectable sialidase activity (9). Later, analysis of NanH1 suggested an intracellular location, while NanH2 and NanH3 were proposed as the primary contributors to sialidase activity (9). Structural studies indicated that NanH2 may either be secreted or tethered to the bacterial surface, whereas NanH3 (which can occur alongside NanH2 or independently) can be secreted, intracellular, or unexpressed in certain strains (9). More recently, a putative novel sialidase, NanH4, was identified in *G. vaginalis* strains. However, it was predicted to be intracellular or inactive, as the strain in which it was identified (*G. vaginalis* ATCC 49145) did not exhibit sialidase activity (10).

Although the role of sialidases in microbial pathogenesis is well established, their use as therapeutic targets has so far been limited to viruses (11); inhibitors of influenza virus neuraminidase (sialidase), such as oseltamivir, zanamivir, laninamivir, and peramivir, have been employed in the treatment of severe influenza infections, but previous studies have reported conflicting results regarding the inhibitory effects of zanamivir on *Gardnerella* sialidase activity

(10,12). However, when used in topical form, zanamivir has been shown to be effective and safe even during pregnancy (13). Inhibition of sialidase activity helps to preserve functional IgA, thereby reducing inflammation and bacterial adhesion, while also limiting the availability of free sialic acid, which is essential for bacterial growth (10,12). These findings underscore the potential of targeting bacterial sialidases as a therapeutic strategy and highlight the need for further research into their diversity and functional roles in pathogenic processes.

In this study, we aimed to investigate the presence, diversity, and structure of the four *Gardnerella* sialidases in isolates from the urogenital tract of healthy and unhealthy women, characterizing their activity on different substrates. We analyzed the genomic environment, architectural organization and structural features of these sialidases to understand their functional implications and potential links to dysbiosis. Additionally, we evaluated sialidase activity and tested the inhibitory effects of sialidase inhibitors, providing insights into the genomic and functional diversity, substrate preferences, and the potential of sialidase inhibition as a therapeutic strategy in the urogenital tract.

Materials and methods

Genomes and bacterial collection of *Gardnerella*

A total of 153 *Gardnerella* genomes were analyzed in this study. Of these, 33 were derived from isolates in our collection, obtained during a previous study on the Female Urogenital Microbiome (FUM) conducted by our research team (3,14), with isolates carrying sialidase genes being further selected for the experimental assays performed in this study. To broaden our analysis, an additional 120 *Gardnerella* genomes were retrieved from NCBI. Additionally, our group acquired the commercially available *G. vaginalis* ATCC 49145 (accession number: GCF_003034925.1), in which the novel sialidase NanH4 was previously identified, and incorporated it into the bacterial collection used in this study.

Extraction of sialidase genes and amino acid sequences

The presence of sialidase genes in our genome dataset was assessed using the MyDbFinder 2.0 tool (Center for Genomic and Epidemiology, https://genomicepidemiology.org/). To build the reference database, we used the well-characterized *nanH1*, *nanH2*, and *nanH3* genes from *Gardnerella piotii* strain JCP8151B (NCBI accession number: GCF_000414485.1) as nucleotide

reference sequences (9). MyDbFinder 2.0 identified nucleotide sequences with a minimum identity threshold of 70% and a minimum length of 60%. Corresponding amino acid sequences for different sialidase enzymes were then extracted from the complete genome dataset using the anvi'o software.

Identification of *nanH* gene clusters and phylogenetic analysis

To identify *nanH* gene clusters, Geneious Prime® 2024.0.7 was used. Functional predictions for the enzyme-encoding genes within these clusters were performed using UniProt [https://www.uniprot.org/, (15)]. Phylogenetic trees for all sialidases (NanH1, NanH2, NanH3, and NanH4) were constructed using MEGA v7.0.26 [http://www.megasoftware.net, (16)] with the following parameters: *i)* the Neighbor-Joining algorithm was employed for tree reconstruction (17); *ii)* evolutionary distances were calculated using Kimura's two-parameter model (18); and *iii) Bifidobacterium longum* subsp. *infantis* ATCC 15697 (accession number: GCF_000020425.1, exo-alpha-sialidase NCBI-Protein ID: ACJ53406.1) was used as an outgroup. Tree topology reliability was assessed via bootstrap analysis with 1,000 replicates, considering bootstrap values ≥80% as significant (19). Multiple sequence alignments were generated using Clustal Omega (https://www.ebi.ac.uk/).

Structural analysis of the putative active center in the catalytic module

To investigate the structural organization of predicted *Gardnerella* sialidases, RAST-annotated sequences were uploaded to InterPro v102.0 (20) using an E-value threshold of 1e-5 across all available member databases. Structural predictions were generated using AlphaFold2 Colab v1.5.5 (https://colab.research.google.com) with default parameters (3 recycles, 200 maximum relax iterations) or obtained from the AlphaFold Protein Structure Database (May 2024).

Model quality was assessed based on Predicted Local Distance Difference Test (pLDDT) confidence scores and Predicted Aligned Error (PAE) scores. Regions with pLDDT scores below 70 were considered low-confidence and excluded from further analysis. Structural comparisons with previously characterized sialidases were performed using reference structures retrieved from the Protein Data Bank (PDB; https://www.rcsb.org/) with resolutions ≤2.5 Å to ensure high structural detail.

Structural analyses were conducted using Coot v0.8.9.2 CCP4 and PyMOL v2.5.2, with structural alignments performed using the align command (cycles = 5, *cutoff* = 2.0 Å).

Visualizations were rendered in ChimeraX v1.7.1 with standard lighting and surface representation (21). To validate the architectural arrangement, each structural module was analyzed using the DALI server (Z-score threshold ≥2.0) and compared against various structures in the PDB, with detailed results provided in Tables S2–S5.

Sialidase activity screening

Twenty-one isolates from the bacterial collection used in this study were selected for screening based on the presence of *nanH* genes identified in their genomes. Among them, strains carrying *nanH1* alone or in combination with *nanH3*, as well as the commercially available *G. vaginalis* ATCC 49145 (harboring *nanH1* and *nanH4*), were tested for sialidase activity using a qualitative filter paper spot test. Briefly, the isolates were cultured on Columbia agar with 5% sheep blood (Biogerm, Portugal) at 37°C under anaerobic conditions for 48 hours and one colony from each isolate was added to the fluorogenic substrate 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid (4-MU-N-acetylneuraminic acid, MUN; Merck KGaA, Germany), following previously described protocol (22).

Bacterial strains and crude extracts preparation

G. vaginalis ATCC 49145 (harboring NanH1 and NanH4) and *Gardnerella pickettii* strains c17Ua_112^{T} (carrying NanH3) and c17Ua_118 (NanH3 truncated) were selected for further analysis and substrate assays. These *G. pickettii* strains were chosen for further investigation as they both harbor NanH3 but exhibited different results in the preliminary qualitative filter paper spot test. Bacterial cells were grown in 10 mL Brain Heart Infusion (BHI) broth (Frilabo, Portugal) supplemented with 1% glucose, 1% starch, 0.5% yeast extract, and 1% gelatin at 37°C under anaerobic conditions for 48 hours.

For crude extract preparation, cultures were centrifuged (2,500 × g, 10 min, 4°C), and cells were washed with a 50 mM sodium acetate buffer (pH 5.5). Cell lysis was performed in the same buffer supplemented with 10 mg/mL lysozyme, 10 mg/mL DNase, 10 mg/mL protease inhibitors, and 5 mM $MgCl_2$. Cells were disrupted by sonication (2 min total; 5 sec ON/ 5 sec OFF cycles) while maintained on ice. After centrifugation (2,500 × *g*, 30 min, 4°C), different fractions were collected based on predicted enzyme localization: 1) for *G. vaginalis* ATCC 49145, the supernatant containing intracellular sialidases released after cell lysis was collected; and for 2) for *G. pickettii* strains, the *pellet* containing membrane-anchored sialidases was resuspended in

lysis buffer (50 mM sodium acetate buffer, pH 5.5). These preparations were immediately used for activity analysis and substrate assays.

Functional characterization: monitoring activity and substrate specificity

For activity characterization assays, crude extracts were tested in black polypropylene 96-well plates (Corning). The reaction mixture consisted of 100 mM sodium acetate buffer (pH 5.5) and 300 µM MUN in a final volume of 20 µL. Fluorescence was monitored at 37°C using a TECAN Spark microplate reader (excitation: 365 nm, emission: 440 nm) with measurements taken every 60 seconds for 2 hours. Recombinant sialidase from *Clostridium perfringens* (rSia) served as a positive control, while crude extract from *G. pickettii* c17Ua_118 (containing a truncated *nanH3* gene) was used as a negative control. For inhibition studies, DANA (2-deoxy-2,3-didehydro-N-acetylneuraminic acid) and Zanamivir (marketed as Relenza®) were tested at a 10% (v/v) concentration. These inhibitors were pre-incubated with crude extracts for 15 minutes before conducting the activity assays under the same conditions. Although DANA was never commercialized as a therapeutic drug, it was included in this study as it served as the template for the development of more effective and better tolerated sialidase inhibitors.

The enzymes' activity on sialylated substrates was assessed using an ELISA-based assay. Microlon 96-well plates (Greiner Bio-One) were coated with serial dilutions (0.1–1000 µg/mL) of bovine submaxillary gland mucin (BSM, Sigma-Aldrich) or fetuin (Sigma-Aldrich) and incubated overnight at 4ºC. Crude extracts were solubilized in a 100 mM sodium acetate buffer (pH 5.5) and incubated at 37ºC for 90 minutes with agitation (250 rpm). The rSia (0.03 µg/mL) was used as a positive control, while buffer-only samples served as negative controls.

Plates were blocked with 3% BSA containing 5 mM $CaCl_2$ in HEPES-buffered saline (50 mM HEPES, 150 mM NaCl) for 1 hour at room temperature (25ºC). The reduction in sialylation levels after treatment were detected using biotinylated *Sambucus nigra* Agglutinin (SNA, 10 µg/mL, Vector Laboratories) in 0.1% BSA with 5 mM $CaCl_2$ in HEPES-buffered saline. Detection was performed using Alkaline Phosphatase-conjugated streptavidin (10 µg/mL, Thermo Fisher), and absorbance was measured at 405 nm using a Spectramax 190 Microplate Reader with SoftMax Pro 6.5.1 software.

Results

Distribution of sialidases genes in *Gardnerella* species and their association with urogenital health

In our genomic analysis, *nanH1* was the most prevalent sialidase gene, detected in *G. vaginalis* (n=58), *G. piotii* (n=9), *G. pickettii* (n=23), *Gardnerella greenwoodii* (n=5), and *Gardnerella* genomic species (GG) 2, GG9, GG10, GG11, and GG14 (n=12). This gene was found in samples from both healthy women (HW) and those diagnosed with various forms of dysbiosis (Table S1). In contrast, *nanH2* was less common, identified only in *G. piotii*, *G. pickettii*, and GG11. Moreover, *nanH2* was present in strains isolated from women with BV (*G. piotii*, n=6) and in samples with unknown donor health conditions (*G. pickettii*, n=2; GG11, n=1). The *nanH3* gene was primarily associated with dysbiosis but was also present in samples from HW. It was detected in *G. vaginalis* (n=10), *G. piotii* (n=9), and *G. pickettii* (n=23). Interestingly, no sialidase genes were identified in *Gardnerella leopoldii*, *Gardnerella swidsinskii*, GG7, GG12, and GG13.

Phylogenetic analysis confirmed a distinct cluster containing an exo-α-sialidase in three *G. vaginalis* strains from BV women (ATCC 49145, accession number: GCF_003034925.1; FDAARGOS 296, accession number: GCF_002206225.1; FDAARGOS 568, accession number: GCF_003812765.1). A recent study identified this enzyme in twelve *G. vaginalis* strains, designating it as NanH4 (10), in accordance with the established *Gardnerella* sialidase nomenclature.

Genomic context of *Gardnerella* sialidase genes

We identified distinct genomic environments associated with the different sialidase genes (Fig. 1). Despite variations in gene organization, the overall sialidase clusters were largely conserved across *Gardnerella* species, with only minor differences in gene lengths and intergenic regions.

Upstream of *nanH1*, we consistently observed a tRNA-guanine transglycosylase (*tgt*), an enzyme essential for modifying tRNA molecules by replacing a guanine base with queuosine, a highly modified nucleoside present in bacteria (23). Additionally, a tRNA(Leu)CAG gene was identified, which encodes a transfer RNA with a CAG anticodon. This tRNA facilitates the incorporation of leucine into nascent polypeptide chains during protein synthesis, specifically recognizing the CUG codon (24). Downstream of *nanH1*, we detected an open reading frame (*ORF*; *ORF1*) potentially encoding a GntR family transcriptional regulator, known to regulate

carbohydrate transport and metabolism in various bacteria (25), and a *nagA* gene, which encodes N-acetylglucosamine-6-phosphate deacetylase, an enzyme characterized in *Bacillus subtilis* that catalyzes the hydrolysis of N-acetylglucosamine-6-phosphate, producing glucosamine-6-phosphate and acetate, a key step in amino sugar metabolism (26).

Upstream of *nanH2*, we identified a tRNA(Gly) gene, an *ORF2*, encoding a hypothetical protein of unknown function and an *ORF3*, encoding an AAA family ATPase, a multifunctional enzyme involved in cell-cycle regulation, protein proteolysis, and disaggregation, potentially influencing sialidase stability and localization (27). Downstream of *nanH2*, we found two genes both encoding hypothetical proteins with undetermined functions (*ORF4* and *ORF5*) and an *ORF6*, encoding an ATP-binding protein. Upstream of *nanH3* we identified *coaE*, encoding dephospho-CoA kinase, which catalyzes the final step in coenzyme A biosynthesis, essential for fatty acid metabolism, pyruvate oxidation, and energy generation (28), and an *ORF7*, encoding a GNAT (N-acetyltransferase) family protein, which transfers an acetyl group from acetyl-CoA to a variety of molecules, altering their stability, activity, or interactions (29). Downstream of *nanH3*, we identified an *ORF8*, encoding a transmembrane domain (TM2) protein of unknown function and *rpsA*, encoding the ribosomal protein uS2, essential for ribosome function (30).

Upstream of *nanH4*, we identified an *ORF9*, encoding an ABC transporter permease, which may facilitate the import/export of various molecules, and an *ORF10*, encoding an ABC transporter substrate-binding protein, likely functioning alongside *ORF9* to mediate molecular transport. Downstream of *nanH4*, we detected an *ORF1*, encoding a GntR family transcriptional regulator, and a *nanE*, which in *Staphylococcus aureus* encodes N-acetylmannosamine-6-phosphate 2-epimerase, an enzyme that converts N-acetylmannosamine-6-phosphate to N-acetylglucosamine-6-phosphate in the sialic acid utilization pathway, allowing bacteria to use sialic acid as a carbohydrate source (31).

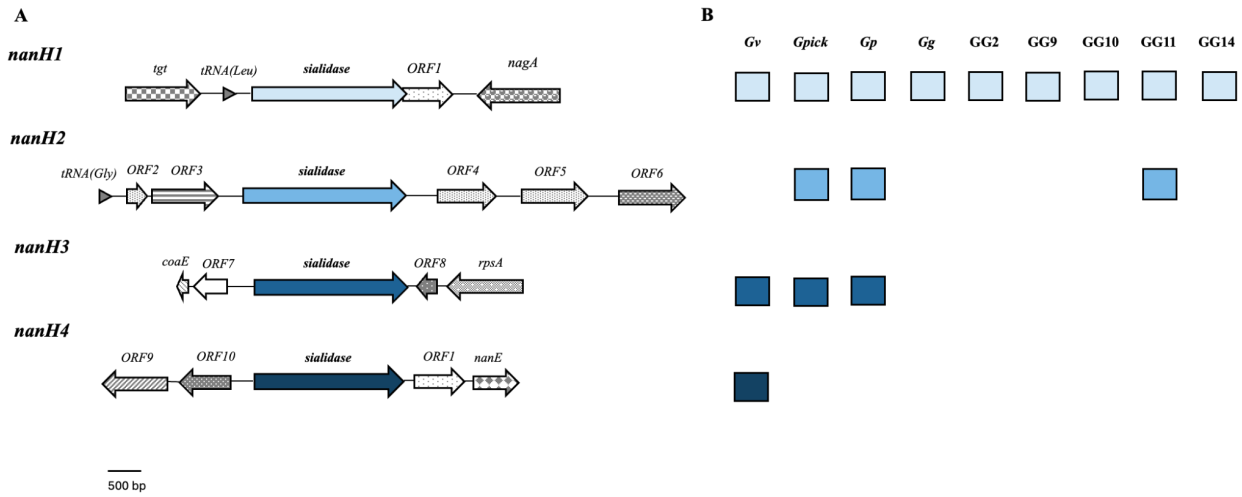


Figure 1: A) Genomic context of *Gardnerella* sialidase genes. *tgt:* tRNA-guanine transglycosylase, *tRNA(Leu):* tRNA-Leu(CAG) gene, *ORF1:* Gene encoding GntR family protein, *nagA:* N-acetylglucosamine-6-phosphate deacetylase, *tRNA(Gly):* tRNA-Gly gene, *ORF2, ORF4* and *ORF5:* encoding hypothetical proteins, *ORF3:* encoding an AAA family ATPase, *ORF6:* encoding an ATP-binding protein, *coaE:* Gene encoding a Dephospho-CoA kinase, *ORF7:* encoding a N-acetyltransferases family (GNAT), *ORF8:* encoding a transmembrane domain (TM2), *rspA:* Gene encoding a small ribosomal subunit protein uS2, *ORF9:* encoding an ABC transporter permease, *ORF10:* encoding an ABC transporter substrate-binding protein, *nanE:* Gene encoding a N-acetylmannosamine-6-phosphate 2-epimerase. **B) Distribution of sialidases across *Gardnerella* species and genomic species.** *Gv*: *G. vaginalis, Gpick*: *G. pickettii, Gp*: *G. piotii, Gg*: *G. greenwoodii,* GG2: *G.* genomic species 2, GG9: *G.* genomic species 9, GG10: *G.* genomic species 10, GG11: *G.* genomic species 11, GG14: *G.* genomic species 14.

Phylogenetic analysis confirms NanH4 as the fourth *Gardnerella* sialidase

A phylogenetic tree was constructed using 162 sialidase amino acid sequences (Fig. S1), revealing four distinct clusters, each supported by bootstrap values >70%. These results confirm NanH4 as a fourth sialidase as previously proposed. Sequence identity comparisons revealed that NanH4 shares the highest amino acid identity with NanH1 (39.29–49.01%), followed by NanH2 (26.97–27.17%) and NanH3 (23.40–24.42%). Phylogenetic trees for NanH1, NanH2, and NanH3 were strongly supported with bootstrap >80% (Fig. S2–Fig. S4). Among *Gardnerella* species, NanH1 revealed a sequence identity range of 86.96–100%, while NanH2 exhibited 96.15–100% identity. The amino acid sequences of NanH3 were similarly conserved, with

identity levels of 91.70–100% across species. All three identified NanH4 sequences exhibited 100% identity among themselves.

Structural and architectural analyses of *Gardnerella* sialidases

To better understand the enzymatic functions, substrate specificity, and interactions of *Gardnerella* sialidases, we performed tridimensional structure and architecture arrangement analyses. These insights provide a deeper understanding of their roles in biological processes, including host-pathogen interactions and microbial adaptation. While all four sialidases contained catalytic residues, aspartate boxes, and positively charged residues, they exhibited distinct structural architectures (Fig. 2, Fig. S5). Architectural analysis revealed that NanH1 possesses an N-terminal ConA-like lectin domain in addition to its enzyme domain, which includes an intradomain region resembling a trans-sialidase observed in NanH4. The ConA-like domains share structural features with galectins, though NanH1 appears to lack key residues required for carbohydrate binding. Additionally, while InterPro and BLASTp identified only a single ConA-like module, structural modeling via AlphaFold suggested the presence of two β-sandwich modules in the N-terminal region, exhibiting a lectin-like fold.

Concerning the structural features of NanH2, we observed that its N-terminal region contains a β-sandwich structure; however, this module was not classified in InterProScan or dbCAN3, and its function remains unknown.

In NanH3, the comparison of the N-terminal regions of NanH3 from the two *G. pickettii* isolates (c17Ua_112^{T} and c17Ua_118), showed that in c17Ua_118, this region is truncated. The truncation affects part of a β-sandwich structure, potentially impairing proper protein folding during expression. Additionally, an immunoglobulin-like domain (32) was identified in the C-terminal region of NanH3.

NanH4 consists of 757 amino acids with a predicted molecular weight of 83.88 kDa. The identified domains include two N-terminal ConA-like lectin domains and a sialidase domain classified as a trans-sialidase. Structural and sequence alignments confirmed that NanH4 possesses five aspartate box repeats, seven conserved active site residues, and positively charged arginine residues (Fig. S6), which are hallmarks of sialidases, further supporting its classification as the fourth *Gardnerella* sialidase based on phylogenetic evidence.

Unlike NanH2 and NanH3, NanH4 lacks a signal peptide, translocase, TM helix, and transmembrane domain, suggesting that it functions as an intracellular sialidase, such as NanH1. Additionally, its architectural similarity with NanH1 suggests that both may share a similar function. The two ConA-like modules were further analyzed using the DALI server, which classified them as part of the Carbohydrate-Binding Module CAZy Family 40 (CBM40, Table S2-S5).

To determine their functional significance within NanH4, we individually examined each CBM40 domain, comparing their structures with publicly available data from the Protein Data Bank (PDB). Concerning the first CBM40 module it contains arginine, glutamate, and aspartate residues, which may facilitate interactions with a putative sugar ligand but lacks the aromatic residues typically involved in protein-carbohydrate interactions. Structurally, it resembles the CBM40 domain of *Clostridium perfringens* NanI in complex with sialic acid (PDB ID: 2BF6). The second CBM40 module shares greater similarity with the CBM40 domain of *C. perfringens* NanJ bound to sialic acid (PDB ID: 2V73). Moreover, it contains one tyrosine and multiple glutamates, arginines, lysines, asparagines, and aspartates, strongly suggesting a functional glycan-binding site. Additionally, its overall positive charge suggests an interaction with negatively charged ligands.

These findings provide valuable insights into the structural diversity and functional specialization of *Gardnerella* sialidases, further highlighting their potential roles in microbial adaptation and host interactions.

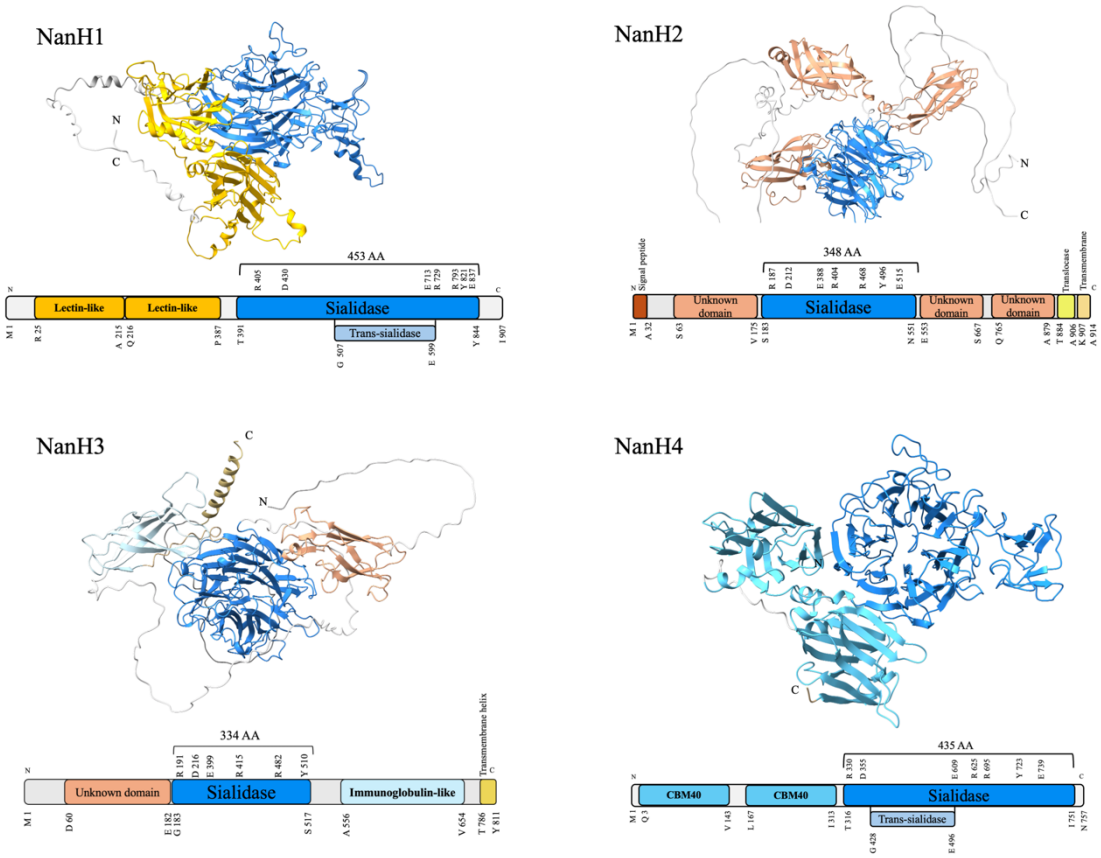


Figure 2: **Enzyme structure models and respective architectural arrangements of *Gardnerella* spp. NanH enzyme types'.** The NanH enzyme structure models were predicted by AlphaFold, rendered using ChimeraX version 1.7.1, and colored by pLDDT. The color used in the structures matches the color of the modules in the schematic representation of the protein architecture. Architectural arrangement reveals distinct modules, which were meticulously examined using coot v0.8.9.2 CCP4 and further validated through DALI server analysis. N: N-terminal; C: C-terminal.

Divergent interacting residues - a hint for distinct substrate specificities

To understand the possible catalytic mechanisms of the NanH4 and potential implications in the context of bacterial pathogenesis, we identified the catalytic and interacting residues from its putative catalytic site through the comparison of *Gardnerella* sialidases with NanI from *C. perfringens* (ID: 2BF6) (Fig. S6, Table S6), which is the closest in structural similarity to NanH4, the sialidase recently identified in *Gardnerella.*

As mentioned above, the residues considered to be involved directly in the catalysis are conserved in the sialidases from this study. In Fig. 3, the putative catalytic residues of *Gardnerella* spp. NanHs are in the same position as the catalytic residues for NanI (represented in red). Additionally, we inspected additional residues interacting with the substrate which we refer here as “interacting residues”. Three of the interacting residues of NanI (Ile267, Arg285, and Asp328), represented in gray, are also conserved in NanH1, NanH2, NanH3 and NanH4 (Fig. 3, Table S6). The primary differences in interaction with sialic acid are found in the remaining interacting residues, highlighted in gray boxes, and which are in their majority interacting with the sialic acid glycerol group. These differences illustrate the variations between NanI and *Gardnerella* NanH types, as well as the variations among the *Gardnerella* NanH types, suggesting different substrate specificity.

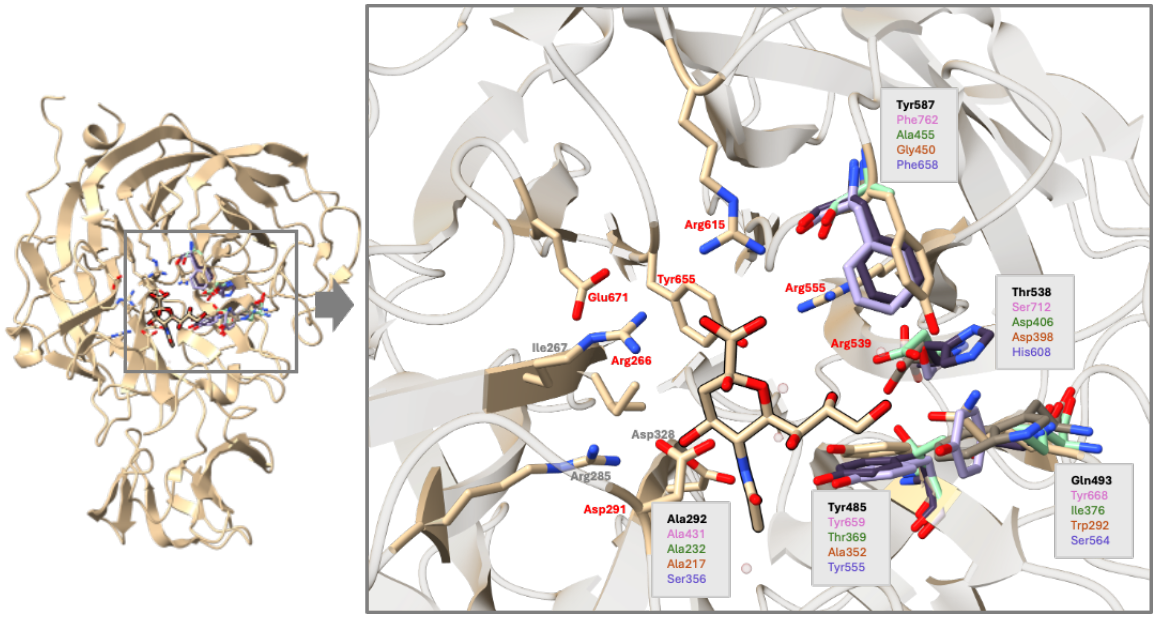


Figure 3: Secondary-Structure Matching (SSM) alignment of the enzymes catalytic domain was performed in ChimeraX version 1.7.1, highlighting the differences observed in the interacting residues (boxed in grey). The residues from *C. perfringens*, used as reference, are shown in black (bold). The enzyme NanH1 is colored pink, NanH2 is green, NanH3 is orange, and NanH4 is purple. Using *C. perfringens* as a reference, the alignment shows the positions of conserved catalytic (red) and interacting (grey) residues.

The nanH4 gene is not responsible for sialidase activity

The enzymatic activity of the crude extracts was first screened using the qualitative filter paper spot test with the neuraminic acid fluorogenic substrate 4-methylumbellifenyl-α-D-N-acetylneuraminic acid (MUN). This test revealed that isolates possessing *nanH3* (from HW and unhealthy) exhibited enzymatic activity, except for one isolate, c17Ua_118, which we observed to have a truncated *nanH3*. In addition, isolates containing only *nanH1* or *nanH1* and *nanH4* genes showed no detectable sialidase activity (Table S1).

NanH3 has strong sialidase activity and differential inhibition by DANA and Zanamivir

To further test whether the *Gardnerella* crude extracts with NanH1+NanH4 (from ATCC 49145) and with NanH3 (c17Ua_112^T) encode active sialidases, the activity was monitored over time using MUN as substrate. The MUN cleavage is represented as the change in relative fluorescence units for 120 minutes (Fig. 4a, b, c). Our results indicate that the crude extract of c17Ua_112^T containing NanH3 exhibited sialidase activity in this assay, whereas only a residual activity was monitored for NanH1+NanH4 crude extracts. The recombinant rSia used as control, showed comparatively an intermediate activity.

We also compared the effects of two sialidase inhibitors, DANA and Zanamivir, on the sialidase activity of different enzyme samples over time (Fig. 4a, b, c). Our results show a decrease in fluorescence with both inhibitors, indicating that DANA and Zanamivir inhibit crude extracts containing NanH1 and NanH4. However, their inhibitory effect is more pronounced in crude extracts with NanH3, particularly for Zanamivir, which leads to a 1.75-fold reduction in fluorescence (from approximately 35,000 to 20,000). Notably, fluorescence stabilizes around 20,000 for NanH1+NanH4 and NanH3. For rSia, both inhibitors reduced substantially enzyme activity, with resulting fluorescence levels between 5,000-10,000. The consistency between the two graphs (Fig. 4a, b) indicates that DANA and Zanamivir might be effective inhibitors of the sialidase enzymes from *Gardnerella* genus. Additionally, to evaluate NanH1+NanH4 and NanH3 activity on sialylated substrates, ELISA-based assays were performed using two types of sialylated glycoproteins: *1)* mucin derived from bovine submaxillary gland (BSM), which is characterised by heavly O-glycosylation with a predominance of the acidic disaccharide NeuAcα2-6GalNAc-Ser/Thr (sialyl-Tn antigens) (33,34) and *2)* fetuin, a heavly N-glycosylated serum protein containing bi-, tri-, and tetra antennary N-linked oligosaccharides with variable sialylation, as well as O-linked glycans (35). The substrates were examined post-enzyme treatment using ELISA with *Sambucus nigra* Agglutinin (SNA), a plant lectin specific for terminal

NeuAcα2-6-sialylated glycans (36). For these assays, binding of SNA to enzyme untreated glycoproteins and to those treated with the rSia was used as controls and for comparative analysis (Fig. 4d). Reduced SNA binding to both BSM and fetuin was observed after treatment with rSia and the enzyme crude extracts, with a more pronounced effect for NanH3 in both for both glycoproteins, while NanH1+NanH4 exhibited a less pronounced effect. These results indicate α2-6-linkage cleavage activity by the *Gardnerella* sialidases, confirming the sialidase activity of the crude extract containing NanH3 and supporting sialidase activity of NanH1+NanH4 towards α2-6 and O-linked sialic acids.

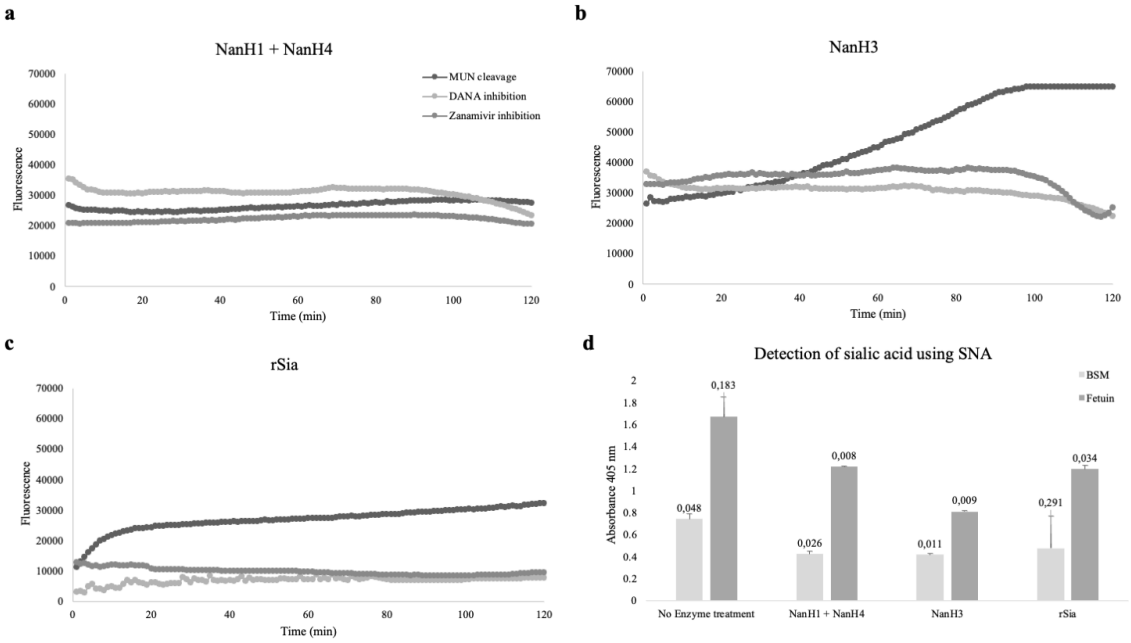


Figure 4: a, b, c) *Gardnerella* crude extracts with different sialidases are active on MUN and can be inhibited by DANA and Zanamivir. The MUN cleavage assay over time shows that different *Gardnerella* crude extracts are active on MUN substrate (fluorogenic 4-methylumbellifenyl-α-D-N-acetylneuraminic acid). Fluorescence is measured as an indicator of enzyme activity, with higher fluorescence indicating higher activity. DANA and Zanamivir demonstrated strong efficacy in reducing the sialidase activity observed in crude extracts containing NanH3. **d) Detection of α2-6-linked sialic acid using *Sambucus nigra* Agglutinin (SNA) lectin after treatment of BSM and fetuin with *Gardnerella* crude extracts containing different sialidases.** A control group without enzyme treatment was used for comparative analysis. NanH1+NanH4: crude extract with NanH1 and NanH4, NanH3: crude extract with NanH3, rSia: recombinant sialidase from *C. perfringens*, BSM: characterized by O-linked

sialoglycans enrichment, Fetuin: characterized by N-linked sialoglycans enrichment, SNA: *Sambucus nigra* Agglutinin, recognizes preferentially α2-6-linkage sialic acid.

Discussion

In this study, we found a higher prevalence of NanH1 and NanH3, while NanH2 and NanH4 were less identified. Consistent with previous studies, our investigation confirms the absence of sialidase enzymes in *G. leopoldii* and *G. swidsinskii*, and in *G.* genomic species 7, 12, and 13 (37), suggesting that these species might rely on alternative mechanisms that do not involve sialidase activity or on other bacterial enzymes, such as amylases, to obtain carbon sources in the urogenital niche (38). The intracellular location of NanH1 suggests that it may play a role in general *Gardnerella* colonization rather than serving as a specific indicator of pathogenicity. Additionally, NanH4 was detected exclusively in *G. vaginalis*, and its proximity to NanH1 suggests that it is also likely intracellular and may have a similar function of NanH1. In contrast, NanH2 showed a more limited distribution, being detected only in *G. piotii, G. pickettii,* and *G.* genomic species 11. Its presence was predominantly linked to BV samples, with a few exceptions in strains from women of unknown health status. This pattern suggests that NanH2 may be more closely associated with pathogenic strains or less favorable in certain urogenital conditions. Similarly, NanH3 appears to have a stronger association with dysbiosis, although it is also present in samples from HW. This indicates that while NanH3 may have a connection to pathogenicity, its presence alone does not exclusively indicate disease. This suggests that other genetic or environmental factors might influence its expression and impact on health.

In our analysis, NanH2 was never detected as the unique sialidase in the strain, it consistently co-occurred with NanH1 and/or with NanH3. Previous studies have shown that NanH2 can release more sialic acid than NanH3 by cleaving 9-O-acetylated linkages (9). Since complex MUC5AC, a major mucin enriched with O-linked glycans is present in the epithelia of the urogenital tract (39,40), NanH2 may be more effective than NanH3 at degrading these glycans and increasing exfoliation of the vaginal epithelium, which is associated with BV. However, it is conceivable that variations in the urogenital environments, namely interactions between microorganisms in the niche or different levels of sialylation, can lead to diverse modulations and sensitivities to the different bacterial and sialidases types. In fact, the potential role of mucin glycans as host-derived regulators of bacterial phenotype, has been unveiled, demonstrating their ability to downregulate genes associated with virulence. In fact, MUC5AC and MUC5B, the most abundant gel-forming mucins secreted in nasal and oral cavities, respiratory tract, eyes and middle ear, niches colonized by the opportunistic pathogen *Pseudomonas aeruginosa*

(PAO1), triggered the downregulation of genes related to quorum sensing, siderophore biosynthesis, toxin secretion, and was responsible for biofilm disintegration, while upregulated other pathways, such as denitrification pathway (39,40). Thus, host factors, such as mucin composition, may help explain why certain populations exhibit a certain vaginal microbial community structure and or have higher frequency or predisposition to developing urogenital pathologies compared to others.

Phylogenetic analysis revealed a close relationship between NanH2 and NanH3, both proposed as the primary sources of sialidase activity. In contrast, NanH4 formed a distinct cluster, with the highest similarity to NanH1. The distinct clustering of NanH4 and highest similarity to NanH1, suggests that while it may share some structural features with NanH1, it likely possesses unique structural elements that distinguish its function. The complete identity observed among the three NanH4 sequences further supports the notion of a highly conserved structure within this group, reinforcing its distinct phylogenetic and potentially functional identity. These phylogenetic patterns not only highlight the evolutionary divergence among sialidase types but also suggest differences in their structural domains that may support functional specialization. Moreover, the high sequence identity within the NanH1 (86.96-100%), NanH2 (96.15-100%) and NanH3 (91.70-100%) suggests a strong conservation of NanHs structure across *Gardnerella* species. In fact, upon examining the structures of sialidases and the active centers of their catalytic modules, we found that the residues directly involved in catalysis are conserved across all types (Table S6). However, variations in the interacting residues were observed, suggesting that while all sialidases have the potential for catalytic activity, their interactions with sialic acid may vary, indicating distinct substrate specificities or affinities among the sialidase types. This is further supported by the observed architectural arrangements and genomic environments, which point to unique structural features that are likely to enable specialized functions. Considering NanH1 and NanH4, the lack of a signal peptide, translocase, TM helix, or transmembrane domain in both, strongly suggests that they are intracellular enzymes. Additionally, the absence of transport-related genes in the NanH1 genomic environment implies that it may function entirely within the cell or have limited involvement in substrate import, supporting its intracellular role. The presence of GntR family transcriptional regulator, in its vicinity, hints at a more generalized metabolic role, potentially linked to carbohydrate usage. Moreover, the presence of lectin-like domains in NanH1, previously identified in *Vibrio cholerae,* suggested that this might help *Vibrio cholerae* to target sialic acid-rich environments, enhancing the catalytic efficiency of the enzyme (41). The architecture arrangement and genomic vicinity of NanH4 also supports its intracellular localization. The presence of an ABC transporter substrate-binding protein and permease upstream suggests a system for importing substrates into the cell, indicating that NanH4 might

act on substrates once they are inside the cell. Additionally, GntR transcriptional regulator and N-acetylmannosamine-6-phosphate 2-epimerase, align with intracellular metabolic pathways, particularly the utilization of sialic acid derivatives, reinforcing the idea that NanH4 operates in an intracellular environment. Moreover, the presence of two CBM40 domains may contribute to NanH4's unique ligand-binding properties and functional specialization (42). A *trans*-sialidase domain was predicted in both NanH1 and NanH4. *Trans*-sialidases have already been described in gut bacteria and in *Gardnerella* strains may provide an adaptive nutritional advantage in the urogenital tract that other microorganisms may not be able to achieve. The architecture arrangement and sialidase activity of NanH2 and NanH3 have already been demonstrated (9), however there is no previous research on its detailed architectural structure and genomic environment. The architectural structure of NanH2 suggests it could be secreted or remain tethered to the bacterial surface, and its genomic environment suggests that this enzyme might be implicated in interactions with the host or other microorganisms, particularly relevant in the context of pathogenicity. Downstream, the ATP-binding proteins could be involved in energy-dependent processes that regulate the activity or trafficking of NanH2. While the two hypothetical proteins near NanH2 remain uncharacterized, it is plausible that these proteins contribute to the proper function, regulation, or localization of NanH2 within the bacterial cell, potentially modulating its enzymatic activity and broader biological effects.

Moreover, NanH3 could be anchored to the cell or secreted, and its genomic context, which has a more pronounced association with metabolic enzymes, may suggest a specialized role in nutrient acquisition or host interaction under specific metabolic conditions. NanH3 could also be regulated differently, based on the presence of a ribosomal protein downstream. All these variations observed, suggest distinct activities toward different sialoglycoconjugates, potentially influencing host adaptation and contributing to *Gardnerella* pathogenic potential.

To correlate the genomic data with phenotypic traits, we conducted sialidase activity, substrate specificity using ELISA and inhibition assays. In the filter paper spot test, *G. vaginalis* ATCC 49415 with NanH1+NanH4 showed no detectable sialidase activity. However, in activity assays with the crude extract, residual sialidase activity was observed. Considering that it was previously predicted by Robinson *et al.,* (9), that NanH1 exhibits near-baseline activity, our findings suggest that NanH4 may possess intermediate sialidase activity, higher than NanH1 but lower than NanH2 and NanH3. To conclusively determine the enzymatic capabilities of NanH4, future studies employing purified recombinant NanH4 enzymes will be essential. The limited or absent activity of recombinant NanH1 across various substrates, as reported by Robinson *et al.,* (9) may be attributed to several factors. Firstly, the testing conditions may not have been optimized to fully unlock its potential enzymatic activity. Alternatively, NanH1 might

function through an as-yet-undiscovered mechanism or target a substrate that has not been explored in previous studies (43,44).

Overall, while these results may not offer direct insights into the substrate specificity of *Gardnerella*'s sialidases, they do underscore sialidase activity of NanH3 and NanH1+NanH4 towards α2-6 and O-linked sialic acids. In inhibition assays, DANA and Zanamivir exhibited inhibitory effects on crude extracts containing NanH1+NanH4 and NanH3, but with a notorious effect on crude extracts containing NanH3. The differences in inhibition efficacy may be attributed to the residual sialidase activity exhibited by the crude extract containing NanH1+NanH4. As a result, the inhibitory effect could be less pronounced than the inhibitory effect observed for the crude extract containing NanH3, which demonstrated high sialidase activity. Additionally, subtle structural variations among NanH1, NanH3, and NanH4 could influence their binding affinities and the effectiveness of the inhibitors. The cellular localization of these enzymes could also help explain the observed differences; NanH3, which can be membrane-associated or secreted, may be more readily released during crude extract preparation, leading to higher abundance in the extract. In contrast, NanH1 and NanH4 are primarily intracellular, making them more challenging to extract. Notably, Zanamivir has already shown effectiveness against *G. vaginalis* by significantly reducing whole cell sialidase activity, underscoring supporting our results (45). However, conflicting results have also been reported, indicating that Zanamivir is not effective specifically against *G. vaginalis* NanH3 (10), highlighting the need for more tailored approaches in inhibitor development.

Conclusion

This study provides comprehensive insights into the prevalence, diversity, and functional characteristics of *Gardnerella* sialidases, highlighting their potential roles in microbial adaptation and pathogenesis. NanH1 and NanH3 were the most detected sialidases, with phylogenetic and structural analyses revealing distinct clustering patterns and conserved catalytic residues across all sialidase types, but also highlighting structural and genomic variations that suggest functional specialization and species-specific adaptations.

Our study also emphasized the importance of genomic context in understanding sialidase activity. NanH2 and NanH3 appear to play key roles in nutrient acquisition and host interactions, potentially contributing to pathogenicity, while NanH1 and NanH4 likely function intracellularly, participating in metabolic pathways.

These findings underscore the intricate nature of *Gardnerella* sialidase activity and its potential modulation by host factors, such as mucin composition. This complexity challenges the notion of using sialidase presence alone as a reliable predictor of virulence. The interplay between bacterial sialidases and the host environment likely influences not only bacterial virulence but also community structure in ways that are not yet fully understood. Factors such as the specific sialidase isoenzymes present, their relative expression levels, and the host's unique mucosal environment may all contribute to the overall impact of sialidase activity on bacterial pathogenicity. Consequently, a more nuanced approach, considering both bacterial and host factors, may be necessary to accurately assess the role of sialidases in *Gardnerella* virulence and their broader implications for vaginal health.

Functional assays confirmed the activity of NanH3, NanH2, and NanH1+NanH4 toward specific sialoglycoconjugates, revealing notable differences in substrate specificity and inhibitor sensitivity among the sialidases. The distinct sensitivity profiles underscore the complexity of sialidase inhibition and highlight the need for further research to optimize inhibitor design to develop more targeted and effective therapeutic strategies.

Future research should focus on developing more selective and potent inhibitors of *Gardnerella* sialidases and evaluating their efficacy in *in vivo* models of BV and other urogenital disorders. By targeting these enzymes, we may be able to develop novel therapeutic strategies that restore the balance of the vaginal microbiome and improve women's health.

Ethical Approval

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.

Acknowledgements

We thank Dr. Paula Videira and Dr. Zélia Silva for providing us the commercially available sialidase inhibitors.

Supplementary Material

Fig. S1-S5 and **Table S1-S5**: Supplementary information about *Gardnerella* data used in this study.

References

1. 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 Mar 1;69(3):679–87.

2. Yoo JJ, Song JS, Kim W Bin, Yun J, Shin HB, Jang MA, *et al. Gardnerella vaginalis* in Recurrent Urinary Tract Infection Is Associated with Dysbiosis of the Bladder Microbiome. J Clin Med. 2022 May 1;11(9).

3. Perovic SU, Ksiezarek M, Rocha J, Vale L, Silva C, Dinis P, *et al.* Time To Change Microbiological Approach To Overactive Bladder. EAU19 - 34th Annual EAU Congress. 2019. 18(1):e103, ISSN 1569-9056.

4. Lewis AL, Lewis WG. Host sialoglycans and bacterial sialidases: A mucosal perspective. Cell Microbiol. 2012 Aug;14(8):1174-82.

5. Schnaar RL, Gerardy-Schahn R, Hildebrandt H. Sialic Acids In The Brain: Gangliosides And Polysialic Acid In Nervous System Development, Stability, Disease, And Regeneration. Physiol Rev 2014 94:461–518.

6. Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA. Mucins in the mucosal barrier to infection. Mucosal Immunol. 2008 May;1(3):183-97.

7. Juge N, Tailford L, Owen CD. Sialidases from gut bacteria: A mini-review. Biochem Soc Trans. 2016 Feb 15;44:166–75.

8. Von Nicolai H, Hammann R, Salehnia S, Zilliken F. A Newly Discovered Sialidase from *Gardnerella vaginalis.* Zentralbl Bakteriol Mikrobiol Hyg A. 1984 Oct;258(1):20-6.

9. Robinson LS, Schwebke J, Lewis WG, Lewis AL. Identification and characterization of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium *Gardnerella vaginalis*. J Biol Chem. 2019 Apr 5;294(14):5230–45.

10. Pelayo P, Hussain FA, Werlang CA, Wu CM, Woolston BM, Xiang CM, *et al. Prevotella* are major contributors of sialidases in the human vaginal microbiome. Proc Natl Acad Sci U S A. 2024 121(36).

11. Gubareva L, Mohan T. Antivirals Targeting the Neuraminidase. Cold Spring Harb Perspect Med. 2022 Jan 1;12(1).

12. Govinden G, Parker JL, Naylor KL, Frey AM, Anumba DOC, Stafford GP. Inhibition of sialidase activity and cellular invasion by the bacterial vaginosis pathogen *Gardnerella vaginalis*. Arch Microbiol. 2018 Sep 1;200(7):1129–33.

13. Xie HY, Yasseen AS, Xie RH, Fell DB, Sprague AE, Liu N, *et al.* Infant outcomes among pregnant women who used oseltamivir for treatment of influenza during the H1N1 epidemic. Am J Obstet Gynecol. 2013 208(4):293.e1-293.e7.

14. Ugarcina Perovic S, Ksiezarek M, Rocha J, Cappelli EA, Sousa M, Ribeiro TG, *et al.* Urinary Microbiome of Reproductive-Age Asymptomatic European Women. Microbiol Spectr. 2022 Dec 21;10(6).

15. Coudert E, Gehant S, de Castro E, Pozzato M, Baratin D, Neto T, *et al.* Annotation of biologically relevant ligands in UniProtKB using ChEBI. Bioinformatics. 2023 Jan 1;39(1).

16. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016 Jul 1;33(7):1870–4.

17. Saitou N, Nei M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees'. Molecular Biology and Evolution. 1987 4(4):406–425.

18. 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.

19. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the bootstrap. Evolution. 1985 39(4):783-791.

20. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, *et al.* InterPro in 2022. Nucleic Acids Res. 2023 Jan 6;51(D1):D418–27.

21. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, *et al.* UCSF ChimeraX: Tools for structure building and analysis. Protein Science. 2023 Nov 1;32(11).

22. Moncla BJ, Braham P. Detection of Sialidase (Neuraminidase) Activity in *Actinomyces* Species by Using 2'-(4-Methylumbelliferyl)o1-D-N-Acetylneuraminic Acid in a Filter Paper Spot Test. J Clin Microbiol. 1989 Jan;27(1):182-4.

23. Stengl B, Meyer EA, Heine A, Brenk R, Diederich F, Klebe G. Crystal Structures of tRNA-guanine Transglycosylase (TGT) in Complex with Novel and Potent Inhibitors Unravel Pronounced Induced-fit Adaptations and Suggest Dimer Formation Upon Substrate Binding. J Mol Biol. 2007 Jul 13;370(3):492–511.

24. Cinnéide E, Scaife C, Dillon ET, Wolfe KH. Evolution of the Genetic Code in the Ascoideales (CUG-Ser2) Yeast Clade: The Ancestral tRNA-Leu(CAG) Gene Is Retained in Most Saccharomycopsis Species but Is Nonessential and Not Used for Translation. Genome Biol Evol. 2024 Aug 1;16(8).

25. Li Z, Xiang Z, Zeng J, Li Y, Li J. A GntR family transcription factor in *Streptococcus mutansregulates* biofilm formation and expression of multiple sugar transporter genes. Front Microbiol. 2019 Jan 14:9:3224.

26. Vincent F, Yates D, Garman E, Davies GJ, Brannigan JA. The three-dimensional structure of the N-acetylglucosamine-6-phosphate deacetylase, NagA, from *Bacillus subtilis*: A member of the urease superfamily. Journal of Biological Chemistry. 2004 Jan 23;279(4):2809–16.

27. Kedzierska S. Structure, function and mechanisms of action of ATPases from the AAA superfamily of proteins. Postepy Biochem 2006 52(3):330–8.

28. Gong X, Chen X, Yu D, Zhang N, Zhu Z, Niu L, *et al.* Crystal structure of *Legionella pneumophila* dephospho-CoA kinase reveals a non-canonical conformation of P-loop. J Struct Biol. 2014 Dec 1;188(3):233–9.

29. Shirmast P, Ghafoori SM, Irwin RM, Abendroth J, Mayclin SJ, Lorimer DD, *et al.* Structural characterization of a GNAT family acetyltransferase from *Elizabethkingia anophelis* bound to acetyl-CoA reveals a new dimeric interface. Sci Rep. 2021 Dec 1;11(1).

30. Schnier J, Kimura M, Foulaki K, Subramanian A raman, Isono K, Wittmann-liebold B. Primary structure of *Escherichia coli* ribosomal protein S1 and of its gene rpsA. Proc Natl Acad Sci U S A. 1982 Feb;79(4):1008-11.

31. Currie MJ, Manjunath L, Horne CR, Rendle PM, Subramanian R, Friemann R, *et al.* N-acetylmannosamine-6-phosphate 2-epimerase uses a novel substrate-assisted mechanism to catalyze amino sugar epimerization. J Biol Chem. 2021 Oct 1;297(4).

32. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a (3-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure. 1995 Nov 15;3(11):1197-205.

33. Chai W, Feizi T, Yuen CT, Lawson AM. Nonreductive release of O -linked oligosaccharides from mucin glycoproteins for structure/function assignments as neoglycolipids: application in the detection of novel ligands for E-selectin. Glycobiology.1997 Jan 1;7(6):861–72.

34. Safferthal M, Bechtella L, Zappe A, Vos GM, Pagel K. Labeling of Mucin-Type O-Glycans for Quantification Using Liquid Chromatography and Fluorescence Detection. ACS Measurement Science Au. 2024 Apr 17;4(2):223–30.

35. Jia Y, Liu Y, Wang Y, Li J, Li G. Sialylation-induced stabilization of dynamic glycoprotein conformations unveiled by time-aligned parallel unfolding and glycan release mass spectrometry. Chem Sci. 2024 15, 14431-14439.

36. Dalziel M, Mcfarlane I, Axford JS. Lectin analysis of human immunoglobulin G N-glycan sialylation. Glycoconj J. 1999 Dec;16(12):801-7.

37. Kurukulasuriya SP, Patterson MH, Hill JE. Slipped-Strand Mispairing in the Gene Encoding Sialidase NanH3 in *Gardnerella* spp. Infect Immun. 2021 Feb 16;89(3):e00583-20.

38. Bhandari P, Hill JE. Transport and Utilization of Glycogen Breakdown Products by *Gardnerella* spp. from the Human Vaginal Microbiome. Microbiol Spectr. 2023 Apr 13;11(2).

39. Wheeler KM, Cárcamo-Oyarce G, Turner BS, Dellos-Nolan S, Co JY, Lehoux S, *et al.* Mucin glycans attenuate the virulence of *Pseudomonas aeruginosa* in infection. Nat Microbiol. 2019 Dec;4(12):2146-2154.

40. Gipson IK, Ho SB, Spurr-Michaud SJ, Tisdale AS, Zhan Q, Torlakovic E, *et al*. Mucin Genes Expressed by Human Female Reproductive Tract Epithelia. Biol Reprod. 1997 Apr;56(4):999-1011.

41. Moustafa I, Connaris H, Taylor M, Zaitsev V, Wilson JC, Kiefel MJ, *et al.* Sialic acid recognition by *Vibrio cholerae* neuraminidase. J Biol Chem. 2004 Sep 24;279(39):40819–26.

42. Selvi Asang G, Mamat S, Alias N, Kari A. Expression and characterization of family 40 Carbohydrate Binding Module (CBM) from *Vibrio cholerae* Non-O1 sialidase. AsPac J Mol Biol Biotechnol. 2020 28(4).

43. Agarwal K, Lewis AL. Vaginal sialoglycan foraging by *Gardnerella vaginalis*: Mucus barriers as a meal for unwelcome guests? Glycobiology. 2021 Jun 29;31(6):667-680.

44. Owen CD, Tailford LE, Monaco S, Šuligoj T, Vaux L, Lallement R. *et al.* Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont *Ruminococcus gnavus.* Nat Commun. 2017 Dec 1;8(1).

45. Govinden G, Parker JL, Naylor KL, Frey AM, Anumba DOC, Stafford GP. Inhibition of sialidase activity and cellular invasion by the bacterial vaginosis pathogen *Gardnerella vaginalis.* Arch Microbiol. 2018 Sep 1;200(7):1129–33.

Supplementary Material

Fig. S1: Neighbor-Joining tree based on all sialidases amino acid sequences showing the phylogenetic relationships of NanH1, NanH2, NanH3 and NanH4 from *Gardnerella* strains (n=162 *Gardnerella* amino acid sequences). *Bifidobacterium longum* ATCC 15697 exo-alpha-sialidase was used as an outgroup. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches.

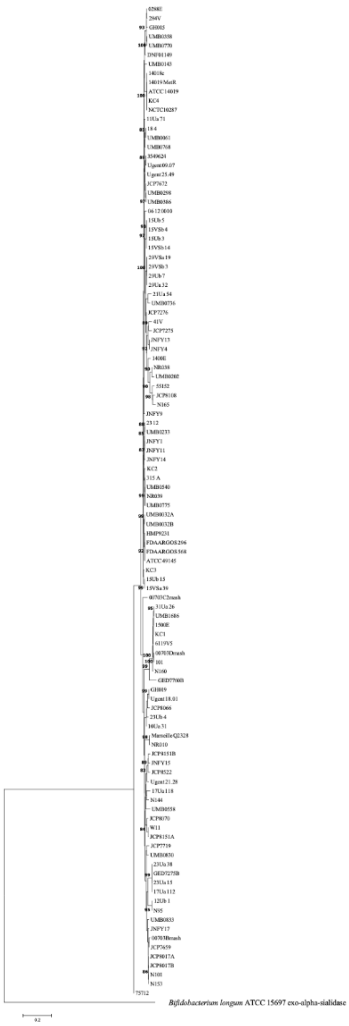


Fig. S2: Neighbor-Joining tree based on amino acid sequences showing the phylogenetic relationships of NanH1 from *Gardnerella* strains (n=108 amino acid sequences). *Bifidobacterium longum* ATCC 15697 exo-alpha-sialidase was used as an outgroup. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches.

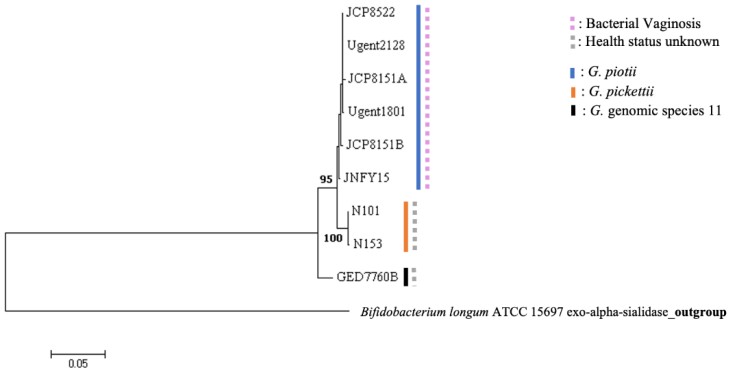


Fig. S3: Neighbor-Joining tree based on amino acid sequences showing the phylogenetic relationships of NanH2 from *Gardnerella* strains (n=9 amino acid sequences). In color is represented the correspondent species and source for each strain. *Bifidobacterium longum* ATCC 15697 exo-alpha-sialidase was used as an outgroup. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches.

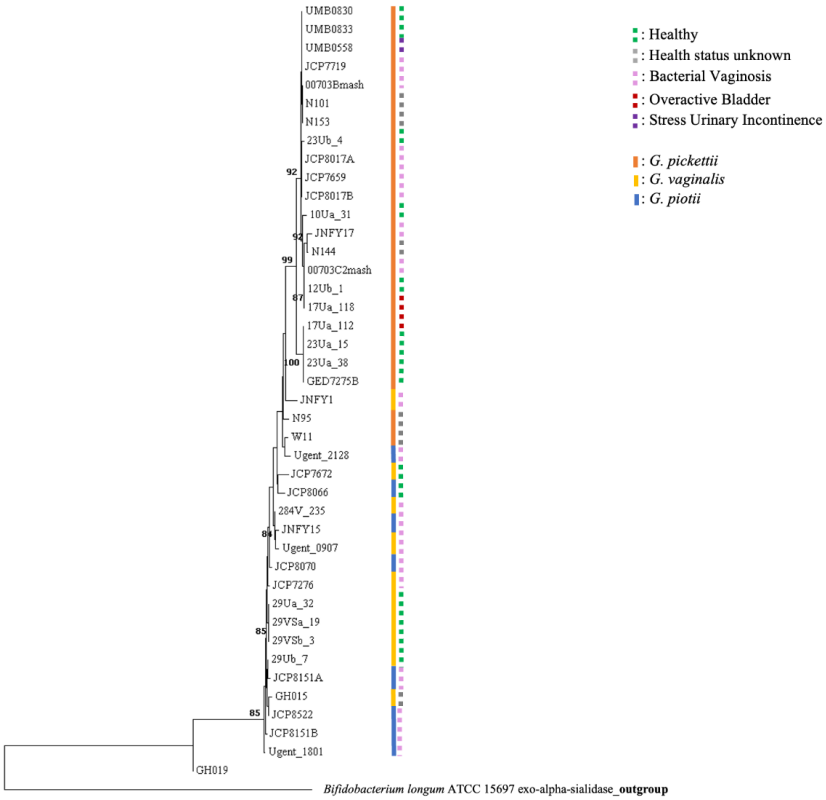


Fig. S4: Neighbor-Joining tree based on amino acid sequences showing the phylogenetic relationships of NanH3 from *Gardnerella* strains (n=42 amino acid sequences). In color is represented the correspondent species and source for each strain. *Bifidobacterium longum* ATCC 15697 exo-alpha-sialidase was used as an outgroup. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches.

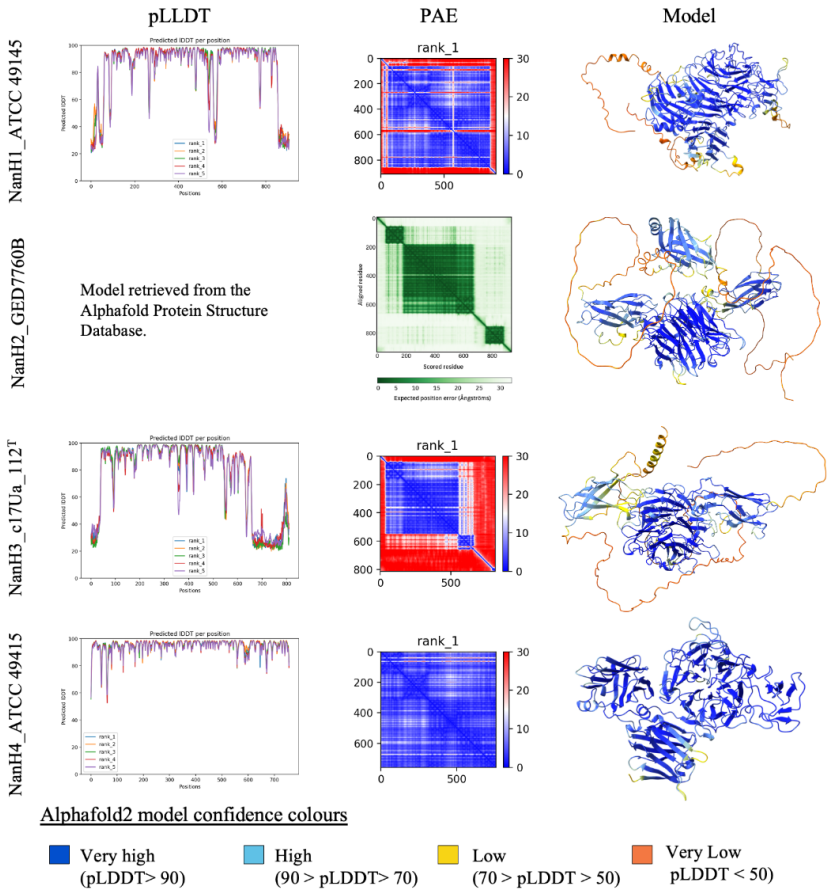


Fig. S5: High quality of the predicted Alphafold2 models of the sialidases. The predicted Local Distance Difference Test (pLDDT) estimates that the backbone and side chains of the globular domains were predicted with high accuracy for the four sialidases. Predicted Aligned Error (PAE) plots reveal that the relative placement of domains in the predicted structures of NanH1 and NanH2 is estimated to be correct, contrary to NanH2 and NanH3. The models of the predicted structures are also presented coloured with Alphafolds2's level of confidence colours. The Alphafold2 model NanH2_GED7760B was retrieved from the Alphafold Protein Structure database and, for this reason, doesn't have a pLDDT plot.

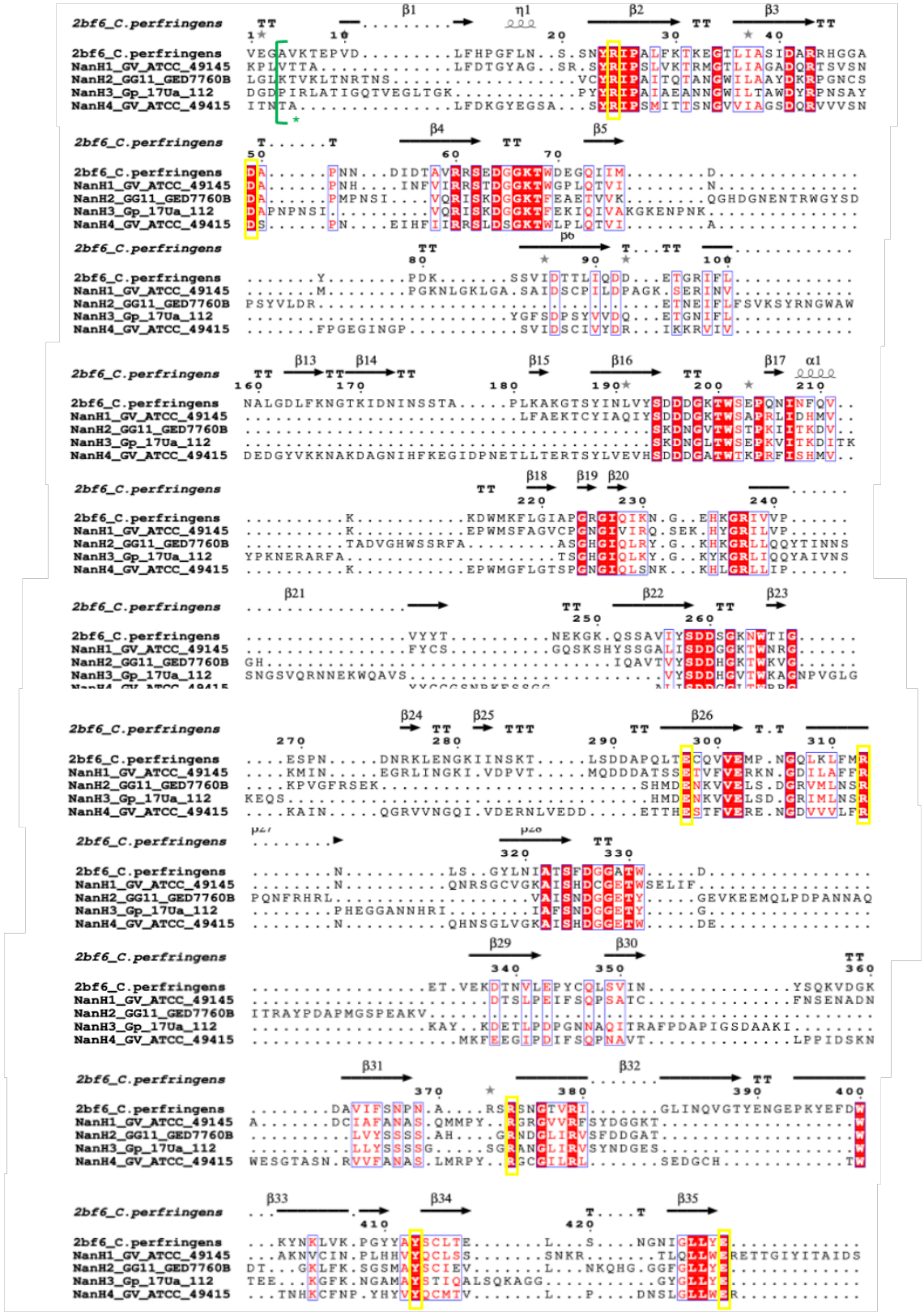


Figure S6: Sequence alignment of the catalytic domain from the sialidase superposition of structures from different *Gardnerella* spp. obtained from the superposition of Alphafold2 models with the X-ray 3D structure of the bacterial sialidase NanI from *C. perfringens* in

complex with Neu5Ac (used as query sequence; ID: 2bf6), using PyMOL v 2.5.2. The figure with catalytic residues, highlighted in yellow boxes, was generated using ESPript. Residue numbering is indicated above the superior line. *indicates the beginning of the NanH4 sialidase domain. GV: *G. vaginalis;* GG11: *G.* genomic species 11; Gp: *G. pickettii*.

Table S1: Source, donor health status, and distribution of genes encoding sialidases, sialidase activity and accession number of all isolates analyzed in this study: *Gardnerella* strains from our FUM collection (boldtype) and genomes retrieved from NCBI (available in february 2023).

Isolates	Species	Source	Donor health status	Sialidase activity	*nanH1*	*nanH2*	*nanH3*	*nanH4*	GenBank Accession numbers
c11Ua_71	*G. vaginalis*	MSU	Healthy	NT	+	-	-	-	JARFVR000000000
c15Ub_3	*G. vaginalis*	MSU	Healthy	-	+	-	-	-	JAODXZ000000000
c15Ub_5	*G. vaginalis*	MSU	Healthy	-	+	-	-	-	JANUZA000000000
c15Ub_15	*G. vaginalis*	MSU	Healthy	-	-	-	-	-	JANUYY000000000
c15VSa_39	*G. vaginalis*	VS	Healthy	-	+	-	-	-	JANUYX000000000
c15VSb_4	*G. vaginalis*	VS	Healthy	-	+	-	-	-	JANUYW000000000
c15VSb_14	*G. vaginalis*	VS	Healthy	-	+	-	-	-	JANUYU000000000
c21Ua_54	*G. vaginalis*	MSU	OAB	-	+	-	-	-	JANUYR000000000

c29Ua_32	*G. vaginalis*	MSU	Healthy	+	+	-	+	-	JANUYN000000000
c29Ub_7	*G. vaginalis*	MSU	Healthy	+	+	-	+	-	JANUYM000000000
c29VSa_19	*G. vaginalis*	VS	Healthy	+	+	-	+	-	JANUYL000000000
c29VSb_3	*G. vaginalis*	VS	Healthy	+	+	-	+	-	JANUYK000000000
ATCC 49145	*G. vaginalis*	VS	Intermediate BV	-	+	-	-	+	GCF_003034925.1
ATCC 14018^{T}	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_003397685.1
JCP7276	*G. vaginalis*	VS	Intermediate BV	+	+	-	+	-	GCF_000414685.1
284V	*G. vaginalis*	VS	BV and *Chlamydia trachomatis*	NT	+	-	+	-	GCF_000263435.1
FDAARGOS_296	*G. vaginalis*	VS	BV	NT	+	-	-	+	GCF_002206225.1

FDAARGOS_568	*G. vaginalis*	VS	BV	NT	+	-	-	+	GCF_003812765.1
75712	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_000263535.1
3549624	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_001049785.1
0288E	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_000263555.1
14018c	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_004336715.1
14019_MetR	*G. vaginalis*	NA	BV	NT	+	-	-	-	GCF_001278345.1
315-A	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_000214315.1
ATCC 14019	*G. vaginalis*	VS	BV	-	+	-	-	-	GCF_000159155.2
JNFY1	*G. vaginalis*	VS	BV	NT	+	-	+	-	GCF_023277725.1
JNFY11	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_023277645.1
JNFY13	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_023277625.1

JNFY14	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_023277605.1
JNFY4	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_023277685.1
JNFY9	*G. vaginalis*	VS	BV	NT	+	-	-	-	GCF_023277665.1
UGent 09.07	*G. vaginalis*	VS	BV	+	+	-	+	-	GCF_003397665.1
UGent 25.49	*G. vaginalis*	VS	BV	-	+	-	-	-	GCF_003397605.1
KC2	*G. vaginalis*	VS	BV (pregnant)	NT	+	-	-	-	GCF_023016185.1
KC3	*G. vaginalis*	VS	BV (pregnant)	NT	+	-	-	-	GCF_023016245.1
KC4	*G. vaginalis*	VS	BV (pregnant)	NT	+	-	-	-	GCF_023016225.1
NCTC10287	*G. vaginalis*	VS	BV (pregnant)	NT	+	-	-	-	GCF_900637625.1

UMB0061	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_002861165.1
UMB0143	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_013315005.1
UMB0202	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_013315075.1
UMB0233	*G. vaginalis*	MSU	OAB	-	+	-	-	-	GCF_002862045.1
UMB0298	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_002861975.1
UMB0358	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_013315085.1
UMB0386	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_002861965.1
UMB0540	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_013315045.1
UMB0768	*G. vaginalis*	TUC	OAB	NT	+	-	-	-	GCF_002884835.1
UMB0770	*G. vaginalis*	MSU	OAB	NT	+	-	-	-	GCF_002861945.1
UMB0775	*G. vaginalis*	MSU	SUI	NT	+	-	-	-	GCF_002861925.1

UMB0736	*G. vaginalis*	MSU	Kidney Stone	NT	+	-	-	-	GCF_013315025.1
UMB0032A	*G. vaginalis*	MSU	Healthy	NT	+	-	-	-	GCF_002862015.1
UMB0032B	*G. vaginalis*	MSU	Healthy (pregnant)	NT	+	-	-	-	GCF_002862005.1
JCP7672	*G. vaginalis*	VS	Healthy (pregnant)	+	+	-	+	-	GCF_000414645.1
UMB0768	*G. vaginalis*	TUC	NA	NT	+	-	-	-	GCF_002884835.1
NR038	*G. vaginalis*	VS	NA	-	+	-	-	-	GCF_003585655.1
NR039	*G. vaginalis*	VS	NA	-	+	-	-	-	GCF_003585755.1
06-12-0010	*G. vaginalis*	VS	NA	NT	+	-	-	-	GCF_014857145.1
18-4	*G. vaginalis*	MSU	NA	NT	+	-	-	-	GCF_001660755.1
23-12	*G. vaginalis*	MSU	NA	NT	+	-	-	-	GCF_001660735.1

DNF01149	*G. vaginalis*	VS	NA	NT	+	-	-	-	GCF_002894105.1
GH015	*G. vaginalis*	VS	NA	+	+	-	+	-	GCF_003408745.1
HMP9231	*G. vaginalis*	VS	NA	NT	+	-	-	-	GCF_000213955.1
JCP7275	*G.* genomic species 2	VS	BV	-	+	-	-	-	GCF_000414705.1
55152	*G.* genomic species 2	VS	BV	-	+	-	-	-	GCF_000263475.1
1400E	*G.* genomic species 2	VS	BV	-	+	-	-	-	GCF_000263495.1
JCP8108	*G.* genomic species 2	VS	BV	-	+	-	-	-	GCF_000414525.1
41V	*G.* genomic species 2	VS	Healthy	-	+	-	-	-	GCF_000165635.1

N165	*G.* genomic species 2	VS	NA	-	+	-	-	-	GCF_003408785.1
c10Ua_31	*G. pickettii*	MSU	Healthy	+	+	-	+	-	JANUYG010000000
c12Ub_1	*G. pickettii*	MSU	Healthy	+	+	-	+	-	JANUZC000000000
c17Ua_112^{T}	*G. pickettii*	MSU	OAB	+	+	-	+	-	GCF_029226345.1
c17Ua_118	*G. pickettii*	MSU	OAB	-	+	-	+	-	JANUYT000000000
c23Ua_15	*G. pickettii*	MSU	Healthy	+	+	-	+	-	JANUYQ000000000
c23Ua_38	*G. pickettii*	MSU	Healthy	+	+	-	+	-	JANUYP000000000
c23Ub_4	*G. pickettii*	MSU	Healthy	+	+	-	+	-	JANUYO000000000
JNFY17	*G. pickettii*	VS	BV	+	+	-	+	-	GCF_023277565.1
JCP7659	*G. pickettii*	VS	BV	+	+	-	+	-	GCF_000414665.1
JCP7719	*G. pickettii*	VS	BV	+	+	-	+	-	GCF_000414625.1

JCP8017A	*G. pickettii*	VS	BV	+	+	-	+	-	JX860314.1
JCP8017B	*G. pickettii*	VS	BV	+	+	-	+	-	GCF_000414585.1
00703Bmash	*G. pickettii*	VS	BV and HSV-2	+	+	-	+	-	GCF_000263615.1
00703C2mash	*G. pickettii*	VS	BV and HSV-2	+	+	-	+	-	GCF_000263515.1
UMB0558	*G. pickettii*	MSU	SUI	NT	+	-	+	-	GCF_013315115.1
GED7275B	*G. pickettii*	VS	Healthy	+	+	-	+	-	GCF_001546445.1
UMB0830	*G. pickettii*	MSU	Healthy (pregnant)	NT	+	-	+	-	GCF_002861905.1
UMB0833	*G. pickettii*	MSU	Healthy (pregnant)	NT	+	-	+	-	GCF_002861885.1
N101	*G. pickettii*	VS	NA	+	+	+	+	-	GCF_003369895.1

N144	*G. pickettii*	VS	NA	+	+	-	+	-	GCF_003408835.1
N153	*G. pickettii*	VS	NA	+	+	+	+	-	GCF_003369935.1
N95	*G. pickettii*	VS	NA	+	+	-	+	-	GCF_003369965.1
W11	*G. pickettii*	VS	NA	+	+	-	+	-	GCF_003369875.1
GH019	*G. piotii*	VS	BV	+	+	-	+	-	GCF_003426565.1
JCP8070	*G. piotii*	VS	BV	+	+	-	+	-	GCF_000414545.1
JCP8151A	*G. piotii*	VS	BV	+	+	+	+	-	GCF_000414505.1
JCP8151B	*G. piotii*	VS	BV	+	+	+	+	-	GCF_000414485.1
JCP8522	*G. piotii*	VS	BV	+	+	+	+	-	GCF_000414425.1
JNFY15	*G. piotii*	VS	BV	NT	+	+	+	-	GCF_023277585.1
UGent 18.01^{T}	*G. piotii*	VS	BV	+	+	+	+	-	GCF_003397585.1

UGent 21.28	*G. piotii*	VS	BV	+	+	-	+	-	GCF_003397615.1
JCP8066	*G. piotii*	VS	Healthy	+	+	-	+	-	GCF_000414565.1
c1Ub_6	*G. leopoldii*	MSU	Healthy	NT	-	-	-	-	JANUUT010000000
c1Ub_40	*G. leopoldii*	MSU	Healthy	NT	-	-	-	-	JAOPFO000000000
c18Ua_110	*G. leopoldii*	MSU	OAB	NT	-	-	-	-	JANUYS000000000
c18Ua_112	*G. leopoldii*	MSU	OAB	NT	-	-	-	-	JAPDSJ000000000
c31Ua_40	*G. leopoldii*	MSU	Healthy	NT	-	-	-	-	JANUYJ000000000
c31VSa_14	*G. leopoldii*	VS	Healthy	NT	-	-	-	-	JANUYI000000000
c31VSa_35	*G. leopoldii*	VS	Healthy	NT	-	-	-	-	JARFVP000000000
AMD	*G. leopoldii*	VS	BV	-	-	-	-	-	GCF_000176475.1
UGent 06.41^{T}	*G. leopoldii*	VS	BV	-	-	-	-	-	GCF_003293675.1

UGent 09.48	*G. leopoldii*	VS	BV	-	-	-	-	-	GCF_003397635.1
UMB0682	*G. leopoldii*	MSU	OAB	NT	-	-	-	-	GCF_002862065.1
UMB0742	*G. leopoldii*	MSU	OAB	NT	-	-	-	-	GCF_013315135.1
UMB1350	*G. leopoldii*	MSU	OAB	NT	-	-	-	-	GCF_013315125.1
6420B	*G. leopoldii*	VS	Healthy	-	-	-	-	-	GCF_000263575.1
UMB0912	*G. leopoldii*	MSU	Healthy (pregnant)	NT	-	-	-	-	GCF_002861125.1
UMB0913	*G. leopoldii*	MSU	Healthy (pregnant)	NT	-	-	-	-	GCF_002861145.1
30-4	*G. leopoldii*	MSU	NA	NT	-	-	-	-	GCF_001641215.1
UMB0662	*G. leopoldii*	TUC	NA	NT	-	-	-	-	GCF_013315255.1
UMB1489	*G. leopoldii*	TUC	NA	NT	-	-	-	-	GCF_013315195.1

c10Ua_244	*G. swidsinskii*	MSU	Healthy	NT	-	-	-	-	JARFVS000000000
c15Ua_15	*G. swidsinskii*	MSU	Healthy	NT	-	-	-	-	JAPDSI000000000
c15Ua_23	*G. swidsinskii*	MSU	Healthy	NT	-	-	-	-	JANUZB000000000
c15Ub_13	*G. swidsinskii*	MSU	Healthy	NT	-	-	-	-	JANUYZ000000000
c15VSa_36	*G. swidsinskii*	VS	Healthy	NT	-	-	-	-	JARFVQ000000000
c15VSb_13	*G. swidsinskii*	VS	Healthy	NT	-	-	-	-	JANUYV000000000
GV37	*G. swidsinskii*	Blood culture	Bacteremia	-	-	-	-	-	GCF_001953155.1
UMB1698	*G. swidsinskii*	MSU	Diabetes	NT	-	-	-	-	GCF_013315145.1
GS 10234	*G. swidsinskii*	VS	BV	-	-	-	-	-	GCF_003397745.1

GS 9838-1^{T}	*G. swidsinskii*	VS	BV	-	-	-	-	-	GCF_003397705.1
JNFY3	*G. swidsinskii*	VS	BV	NT	-	-	-	-	GCF_023277705.1
N72	*G. swidsinskii*	VS	BV	NT	-	-	-	-	GCF_003408815.1
UMB0769	*G. swidsinskii*	MSU	OAB	NT	-	-	-	-	GCF_013315215.1
409-05	*G. swidsinskii*	VS	Healthy	-	-	-	-	-	GCF_000025205.1
5-1	*G. swidsinskii*	VS	Healthy	-	-	-	-	-	GCF_000176495.1
UMB1642	*G. swidsinskii*	MSU	Healthy	NT	-	-	-	-	GCF_002884795.1

26-12	*G. swidsinskii*	MSU	NA	NT	-	-	-	-	GCF_001660745.1
DNF01162	*G. swidsinskii*	VS	NA	NT	-	-	-	-	GCF_002894125.1
UMB0170	*G. swidsinskii*	TUC	NA	NT	-	-	-	-	GCF_002884855.1
UMB0264	*G. swidsinskii*	TUC	NA	NT	-	-	-	-	GCF_002884875.1
JCP8481A	*G.* genomic species 7	VS	BV	-	-	-	-	-	GCF_000414465.1
JCP8481B	*G.* genomic species 7	VS	BV	-	-	-	-	-	GCF_000414445.1
PSS_7772B	*G.* genomic species 7	MSU	Healthy (pregnant)	-	-	-	-	-	GCF_001546485.1

c31Ua_26^T	*G. greenwoodii*	MSU	Healthy	-	+	-	-	-	GCF_029226345.1
UMB1686	*G. greenwoodii*	MSU	Diabetes	NT	+	-	-	-	GCF_002884775.1
101	*G. greenwoodii*	VS	BV	-	+	-	-	-	GCF_000165615.1
00703Dmash	*G. greenwoodii*	VS	BV and HSV-2	-	+	-	-	-	GCF_000263635.1
KC1	*G. greenwoodii*	VS	BV (pregnant)	NT	+	-	-	-	GCF_023016205.1
6119V5	*G.* genomic species 9	VS	Healthy	-	+	-	-	-	GCF_000263655.1
N160	*G.* genomic species 9	VS	NA	-	+	-	-	-	GCF_003408775.1

1500E	*G.* genomic species 10	VS	BV	-	+	-	-	-	GCF_000263595.1
GED7760B	*G.* genomic species 11	VS	NA	+	+	+	-	-	GCF_001546455.1
CMW7778B	*G.* genomic species 12	VS	Healthy (pregnant)	-	-	-	-	-	GCF_001563665.1
KA00735	*G.* genomic species 12	VS	NA	NT	-	-	-	-	GCF_002894085.1
KA00225	*G.* genomic species 13	VS	NA	-	-	-	-	-	GCF_002896555.1
NR010	*G.* genomic species 14	VS	BV	-	+	-	-	-	GCF_003408845.1
Marseille-Q2328	*G.* genomic species 14	NA	NA	NT	+	-	-	-	GCF_940796375.1

Abbreviations - BV: Bacterial Vaginosis, HSV-2: Herpes Simplex Type 2, MSU: Midstream Urine, NA: Not Available, NT: Not Tested, OAB: Overactive Bladder, SUI: Stress Urinary Incontinence, TUC: TransUrethral Catheter, VS: Vaginal Swab, +: positive, -: negative.

Table S2: Structure comparison of NanH1 modules to protein structures in PDB.

Enzyme	Module	DALI search							
NanH1	Lectin-like	**N.º**	**ID structure**	**Z-value**	**rmsd**	**lali**	**nres**	**%id**	**PDB description**
	Lectin-like 1	1	8b55	12.4	2.5	130	193	8	adhesion g-protein coupled receptor g4
		2	5ho2	12.3	2.5	136	803	8	extracellular arabinanase
		3	8byp	12.2	2.6	137	1155	13	ntnh/x
		4	7oa8	12.0	2.9	136	376	7	pilc minor pilin
		5	1h30	11.9	2.3	124	391	6	growth-arrest-specific protein

		6	1sli	11.4	3.0	145	679	11	intramolecular *trans*-sialidase
		7	4pbp	11.4	2.8	136	210	NA	c-reactive protein
		8	3azv	11.2	2.9	134	418	13	d/c mosaic neurotoxin
		9	7zl1	11.2	2.8	136	215	15	pentraxin-related protein ptx3
		10	4hl0	11.2	2.7	129	278	10	galectin
	Lectin-like 2	1	1sli	18.4	2.7	165	679	13	intramolecular *trans*-sialidase
		2	5ho2	15.4	2.6	152	803	NA	extracellular arabinanase
		3	2erf	15.0	2.6	155	209	13	thrombospondin-1

		4	8byp	14.6	2.6	150	1155	8	ntnh/x
		5	8b55	14.5	2.3	135	193	10	adhesion g-protein coupled receptor g4
		6	7zl1	14.3	2.7	145	215	21	pentraxin-related protein ptx3
		7	7ufg	14.0	2.7	148	1151	16	pappalysin-1
		8	4pbp	13.5	3.2	144	210	17	c-reactive protein
		9	2uur	13.5	2.7	152	210	12	collagen α-1(ix) chain
		10	3flp	13.0	3.1	146	217	14	sap-like pentraxin
	Sialidase	1	1eut	48.8	2.0	347	601	27	sialidase
		2	6qzh	47.9	1.9	430	744	33	human chemokine receptor 7
		3	1sli	41.3	2.3	411	679	25	intramolecular *trans*-sialidase

		4	1kit	40.6	2.6	347	757	24	sialidase
		5	7mhu	38.7	2.5	324	374	20	exo-α-sialidase
		6	2ags	36.4	2.0	329	634	19	sialidase
		7	5hx0	33.6	2.6	316	365	17	uncharacterized protein dfer_1899
		8	6myv	33.4	2.0	345	526	27	sialidase26
		9	3h6j	30.9	2.5	299	438	21	neuraminidase
		10	2w5n	29.0	2.9	307	365	14	α-l-arabinofuranosidase

Legend: Dali data description by N.º (order of prediction), ID_structure: identification of structure, Z: Alignment score normalized to compare Z scores across different alignments, rmsd: root mean square deviation, lali: The number of residues aligned, nres: total number of residues, %id: The percent identity, PDB_description: name of molecule by Protein Data Bank, a repository of experimentally determined 3D structures of biological macromolecules (https://www.rcsb.org), NA: Not available.

Table S3: Structure comparison of NanH2 modules to protein structures in PDB.

Enzyme	Module	DALI search results							
NanH2	ß-sandwich	**N.º**	**ID structure**	**Z-value**	**rmsd**	**lali**	**nres**	**%id**	**PDB description**
		1	5msz	7.9	2.7	89	195	12	thermobia domestica domestica aa15
		2	4jcl	7.6	2.3	78	687	18	cyclomaltodextrin glucanotransferase
		3	6stx	7.4	2.7	87	663	20	kelch domain-containing protein
		4	6sws	7.3	2.5	82	113	7	phosphoinositide 3-kinase adapter protein 1
		5	4fmr	7.1	2.2	79	237	13	uncharacterized hypothetical protein
		6	8qb1	7.0	2.6	76	88	12	mirolase
		7	4unm	7.0	2.9	83	609	16	secreted protein

		8	6e57	6.8	2.8	83	399	19	surface glycan binding protein b
		9	4m8r	6.7	2.6	80	406	18	hypothetical protein
		10	5hl3	6.7	2.7	82	356	15	lmo2470 protein
	Sialidase	1	1eut	8.8	3.1	97	601	12	Sialidase
		2	7nit	7.5	2.9	97	1253	14	β-galactosidase
		3	4b0m	7.4	4.3	102	212	16	f1 capsule-anchoring protein
		4	6vs7	7.0	3.4	105	409	13	adhesin
		5	2l0d	7.0	2.8	91	114	15	cell surface protein
		6	6e18	6.9	3.1	99	529	12	hapless 2
		7	6hpd	6.9	3.0	95	955	8	β-galactosidase (gh2)
		8	2j2z	6.7	3.8	95	216	15	chaperone protein papd

		9	7lr3	6.5	3.3	90	110	14	d3_2/6.14 fab light chain
		10	2e6j	6.5	3.8	92	112	10	hydin protein

Legend: Dali data description by N.º (order of prediction), ID_structure: identification of structure, Z: Alignment score normalized to compare Z scores across different alignments, rmsd: root mean square deviation, lali: The number of residues aligned, nres: total number of residues, %id: The percent identity, PDB_description: name of molecule by Protein Data Bank, a repository of experimentally determined 3D structures of biological macromolecules (https://www.rcsb.org).

Table S4: Structure comparison of NanH3 modules to protein structures in PDB.

Enzyme	**Module**	**DALI search results**							
NanH3	Unknow domain	**N.º**	**ID structure**	**Z-value**	**rmsd**	**lali**	**nres**	**%id**	**PDB description**
		1	2qsv	6.9	2.5	87	220	8	uncharacterized protein
		2	8b6l	6.4	3.5	93	153	12	protein transport protein sec61 subunit α iso
		3	6fvi	6.3	2.8	84	142	11	centrosomal protein of 192 kda

		4	1m1s	5.9	2.2	79	109	4	wr4
		5	3msp	5.8	2.7	87	126	13	major sperm protein
		6	6lp4	5.8	2.7	89	126	9	vesicle-associated membrane protein-associated pr
		7	2r39	5.7	2.9	84	109	11	fixg-related protein
		8	7p4l	5.6	4.6	85	478	14	fusexin1
		9	2e6j	5.5	2.4	83	112	10	hydin protein
		10	5cl2	5.4	5.1	88	246	8	sporulation-control protein spo0m
	Immunoglobulin-like	1	1eut	8.5	3.4	99	601	9	sialidase
		2	4b0m	7.6	3.9	101	212	4	f1 capsule-anchoring protein
		3	2j2z	7.5	3.4	96	216	7	chaperone protein papd
		4	6fqa	7.4	4.5	100	229	10	csuc

		5	7nit	7.3	2.9	97	1253	13	β-galactosidase
		6	7ptb	7.3	3.5	105	317	8	centrosomal protein of 192 kda
		7	7vqo	7.2	3.2	92	1180	7	ams1, nbr1 and male fusion protein
		8	6fvi	7.2	4.0	101	142	13	centrosomal protein of 192 kda
		9	8jmw	7.2	3.6	94	517	15	s8 family serine peptidase
		10	2l0d	7.1	3.0	95	114	11	cell surface protein
	Sialidase	1	1eut	57.7	1.1	347	601	52	sialidase
		2	6qzh	46.7	2.1	349	744	25	human chemokine receptor 7
		3	1sli	42.4	2.2	336	679	24	intramolecular *trans*-sialidase
		4	1kit	42.4	2.3	334	757	24	sialidase
		5	7mhu-a	42.1	2.2	326	374	25	exo-α-sialidase

		6	2ags	39.6	2.3	325	634	16	sialidase
		7	6myv	38.0	2.1	337	526	27	sialidase 26
		8	5hx0	35.0	2.6	316	365	20	uncharacterized protein dfer_1899
		9	7zao	32.5	3.2	329	429	15	sialidase (neuraminidase) family protein-like pro
		10	3h6j	30.9	2.6	297	438	14	neuraminidase

Legend: Dali data description by N.º (order of prediction), ID_structure: identification of structure, Z: Alignment score normalized to compare Z scores across different alignments, rmsd: root mean square deviation, lali: The number of residues aligned, nres: total number of residues, %id: The percent identity, PDB_description: name of molecule by Protein Data Bank, a repository of experimentally determined 3D structures of biological macromolecules (https://www.rcsb.org).

Table S5: Structure comparison of NanH4 modules to protein structures in PDB.

Enzyme	Module	DALI search results							
NanH4	CBM40	**N.º**	**ID structure**	**Z-value**	**rmsd**	**lali**	**nres**	**%id**	**PDB description**
	CBM40_1	1	7oa8	12.9	2.8	134	376	13	pilc minor pilin
		2	8b55	12.6	3.4	132	193	11	adhesion g-protein coupled receptor g4
		3	5ho2	12.5	2.6	137	803	14	extracellular arabinanase
		4	8byp	12.4	2.7	135	1155	15	ntnh/x
		5	1sli	11.9	3.2	143	679	14	intramolecular *trans*-sialidase
		6	7zl1	11.8	2.7	133	215	16	pentraxin-related protein ptx3
		7	3azv	11.6	2.7	132	418	17	d/c mosaic neurotoxin

		8	1h30	11.5	2.2	117	391	9	growth-arrest-specific protein
		9	4pbp	11.5	2.7	132	210	11	c-reactive protein
		10	7ufg	11.3	2.9	138	1151	10	pappalysin-1
	CBM40_2	1	1sli	16.5	2.5	146	679	16	intramolecular *trans*-sialidase
		2	7zl1	15.0	2.4	133	215	15	pentraxin-related protein ptx3
		3	2erf	14.1	2.6	143	209	9	thrombospondin-1
		4	8b55	13.8	2.2	126	193	12	adhesion g-protein coupled receptor g4
		5	5ho2	13.6	2.7	140	803	12	Extracellular arabinanase
		6	2uur	13.3	2.5	142	210	15	collagen α-1(ix) chain
		7	8byp	13.1	2.2	134	1155	11	ntnh/x
		8	4pbp	12.3	2.6	132	210	21	c-reactive protein

		9	7ufg	12.3	2.3	132	1151	14	pappalysin-1
		10	3flp	11.6	2.8	133	217	17	sap-like pentraxin
	Sialidase	1	6qzh	54.1	1.7	429	744	33	human chemokine receptor 7
		2	1eut	49.3	1.9	346	601	29	sialidase
		3	1sli	47.4	2.2	404	679	28	intramolecular *trans*-sialidase
		4	1kit	44.0	2.3	344	757	23	sialidase
		5	7mhu	42.2	2.3	322	374	21	exo-α-sialidase
		6	2ags	40.8	2.1	327	634	20	sialidase
		7	6myv	39.3	2.0	343	526	27	sialidase 26
		8	5hx0	34.7	2.6	315	365	19	Uncharacterized protein dfer_1899
		9	7zao	32.3	3.1	331	429	14	sialidase (neuraminidase) family protein-like pros

		10	3h6j	31.7	2.6	297	438	19	neuraminidase

Legend: Dali data description by N.º (order of prediction), ID_structure: identification of structure, Z: Alignment score normalized to compare Z scores across different alignments, rmsd: root mean square deviation, lali: The number of residues aligned, nres: total number of residues, %id: The percent identity, PDB_description: name of molecule by Protein Data Bank, a repository of experimentally determined 3D structures of biological macromolecules (https://www.rcsb.org).

Table S6: Comparison of residues of the catalytic and interacting sites of the different sialidases.

Residues	**NanI**	**NanH1**	**NanH2**	**NanH3**	**NanH4**
Species **Isolate**	*Clostridium perfringens* 2BF6	*Gardnerella vaginalis* ATCC 49415	*G.* genomic species 11 GED7760B	*Gardnerella pickettii* c17Ua_112^{T}	*Gardnerella vaginalis* ATCC 49415
Catalytic residues	Asp 291	Asp 430	Asp 231	Asp 216	Asp 355
	Arg 266	Arg 405	Arg 206	Arg 191	Arg 330
	Glu 539	Glu 713	Glu 407	Glu 399	Glu 609
	Arg 555	Arg 729	Arg 423	Arg 415	Arg 625
	Arg 615	Arg 793	Arg 487	Arg 482	Arg 695
	Tyr 655	Tyr 821	Tyr 515	Tyr 510	Tyr 723
	Glu 671	Glu 837	Glu 534	Glu 529	Glu 739
Interacting residues	Ile 267	Ile 406	Ile 207	Ile 192	Ile 331

	Arg 285	Arg 424	Arg 225	Arg 210	Arg 349
	Ala 292*	Ala 431*	Ala 232*	Ala 217*	Ser 356*
	Asp 328	Asp 472	Asp 272	Asp 231	Asp 395
	Tyr 485	Tyr 659	Thr 369***	Ala 352	Tyr 555
	Gln 493	Tyr 668	Ile 376	Trp 292	Ser 564
	Thr 538**	Ser 712	Asp 406	Asp 398	His 608
	Tyr 587	Phe 762	Ala 455	Gly 450	Phe 658

*The bond is not formed by the side chain, but by the C-α, regardless of which amino acid (AA) is present. **The bond is mediated by water. ***Thr 369 in the polypeptide chain has a 369, but nearby there is a phenylalanine (Phe) and a tryptophan (Trp) with space occupied by other AA side chains such as Phe 346 and Trp 296, creating a different hydrophobic aromatic environment. Interacting residues can engage with the glycerol group, potentially leading to variations in specificity and possibly impacting the overall activity.

A closer look at vaginolysin in *Gardnerella*: discovery of a novel type and characterization of distinct genetic platforms

Márcia Sousa,[1,2] Liliane Ferrador,[1,2] Teresa Gonçalves Ribeiro,[1,2,3] Filipa Grosso,[1,2] Luísa Peixe,[1,2,3*]

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

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

[3] Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP).

*Luísa Peixe, lpeixe@ff.up.pt

Manuscript in preparation

Abstract

Objectives: Vaginolysin (VLY) is a cholesterol-dependent cytolysin associated with *Gardnerella* species, playing a potential role in bacterial vaginosis (BV). However, its genetic diversity, evolutionary dynamics, and functional implications remain poorly understood. This study aims to characterize the distribution, genetic organization, and variability of *vly* across *Gardnerella* species, providing insights into its genomic context and potential role in pathogenesis.

Materials and Methods: A total of 147 *Gardnerella* genomes were analyzed to determine *vly* presence, sequence diversity, and genetic platform organization. VLY were classified based on amino acid differences in the undecapeptide and CD59-binding regions. Genetic surroundings were examined for intergenic region (IGR) variability and flanking gene conservation. Hemolysis assays were performed using selected strains to evaluate VLY functionality and cytolytic activity.

Results: *vly* was detected in 89.1% (n=131/147) of genomes, with the most prevalent variants being Type 1A (43.5%), Type 1B (26.0%), and Type 2 (28.2%). A novel variant, VLY Type 3B, was identified in *Gardnerella* genomic species 12, prompting its subclassification from the previously known Type 3. Comparative genomic analysis revealed five distinct genetic platforms, with high variability in IGR size between and within some *Gardnerella* species, flanking gene composition, and genome structure. Notably, the *gatCAB* operon was consistently found upstream of *vly* across multiple *Gardnerella* species, suggesting a possible regulatory or functional association. Functional assays demonstrated that *vly*-negative strains lacked hemolytic activity, while VLY-producing strains exhibited α- or β-hemolysis, depending on the blood sample used.

Conclusions: This study highlights the genetic diversity and structural variability of *vly* within Gardnerella, providing evidence of ongoing evolutionary adaptations. The identification of novel VLY subtypes and genetic platforms suggests distinct regulatory and functional implications across species. Further research is needed to assess the clinical relevance of VLY variation, its regulatory mechanisms, and its potential as a therapeutic target in BV and related conditions.

Introduction

Bacterial vaginosis (BV) is a reproductive health condition linked to negative outcomes and is marked by a transition from a *Lactobacillus*-dominated vaginal microbiome to a high bacterial diversity, including *Gardnerella* (1). Recent genomic advances have led to a significant reclassification of the genus *Gardnerella*, with at least 14 unique *Gardnerella* genomic species being identified, including 6 officially named species: *Gardnerella vaginalis, Gardnerella piotii, Gardnerella leopoldii, Gardnerella swidsinskii, Gardnerella pickettii,* and *Gardnerella greenwoodii* (2, 3). Notably, *Gardnerella* species have also been found in the urogenital tract of healthy women (4, 5), suggesting that their presence alone is not a definitive indicator of pathology. This underscores the importance of investigating host-microbe interactions, microbiome composition, and virulence determinants that may contribute to disease progression.

Member of *Gardnerella* genus exhibit considerable genomic and phenotypic diversity, which may contribute to their pathogenic potential, particularly through the production of virulence factors such as vaginolysin (VLY), a cholesterol-dependent cytolysin that recognizes the complement regulatory protein CD59 on human cells (e.g., vaginal epithelial cells, erythrocytes and neutrophils), leading to pore formation and cell lysis (6). In addition to direct secretion, VLY has been detected within membrane vesicles produced by *G. vaginalis*, which can be internalized by vaginal epithelial cells (7). The presence and variability of *vly* among different *Gardnerella* species remain critical to understanding their role in BV pathogenesis.

However, *vly* is not universally present in all *Gardnerella* strains. Bohr *et al*. (2020) (8) reported that *vly* is not part of the *Gardnerella* core genome, being identified in 83% of strains. The incongruence between *vly*-based phylogeny and core genome phylogeny, further suggests that it appears to be readily exchanged between *Gardnerella* species, particularly *G. vaginalis* and *G. piotii* (8).

Similar to other CD59-binding CDCs, VLY features a unique undecapeptide domain, where a proline residue at position 480 replaces the typically conserved tryptophan, a motif crucial for membrane interaction (9). The substitution of proline by tryptophan (P480W) results in a non-pore-forming protein, confirming the critical role of this region in cytotoxicity (9). Recently, five distinct VLY types were identified within *Gardnerella*, with variations in the undecapeptide and CD59-binding domains, which may influence virulence and host interactions (6). *G. vaginalis* strains carrying VLY Type 1 were found to be more prevalent and associated with an increased occurrence of vaginal symptoms (odor and discharge), while Type 2 VLY dominant profiles exhibited increased *Gardnerella* spp. abundance (6). These findings suggest that variations in VLY sequence may play a role in host-pathogen interactions and the immunopathology of BV.

Despite these findings, major gaps remain in our understanding of the genomic context, and evolutionary dynamics of *vly*. In this study, we aim to characterize the presence and genetic organization of *vly* across *Gardnerella* species, offering insights into its diversity, evolutionary patterns, and potential functional implications in virulence.

Materials and Methods

Genomes and metadata

One hundred and fourteen genomic sequences, representing good quality of assembly, and their corresponding metadata, available in May 2023 for the entry *Gardnerella*, were downloaded from the National Center for Biotechnology Information genome sequence repository (http://www.ncbi.nlm.nih.gov/genome/). The dataset included 33 whole-genome shotgun sequences from a previous study conducted by our research team (unpublished data). Species-level identification of *Gardnerella* strains has been previously performed and corresponded to: 55 *G. vaginalis,* 23 *G. pickettii*, 18 *G. leopoldii,* 19 *G. swidsinskii*, 9 *G. piotii*, 5 *G. greenwoodii*, 6 *Gardnerella* genomic species 2, 3 *Gardnerella* genomic species 7, 2 *Gardnerella* genomic species 12, 2 *Gardnerella* genomic species 14, 2 *Gardnerella* genomic species 9, 1 of each *Gardnerella* genomic species 10, *Gardnerella* genomic species 11 and *Gardnerella* genomic species 13 (unpublished data).

Gardnerella sp. were isolated from healthy women (n=45, 30.61%) or human with unknown health status (n=22, 14.97%), as well as from human with various conditions, including BV (n=56, 38.10%), overactive bladder (n=18, 12.24%), diabetes (n=2, 1.36%), stress urinary incontinence (n=2, 1.36%), bacteremia (n=1, 0.68%) and kidney stone (n=1, 0.68%) (Table S1).

Detection of Vaginolysin Genes and Identification of Types

Nucleotide and amino acid sequences of vaginolysin were extracted from all annotated genomes using Geneious Prime® 2024.0.7 to assess the distribution and genetic diversity. Vaginolysin genes and their surrounding genetic platforms were retrieved and analyzed from the genome dataset using Geneious Prime® 2024.0.7. Sequence alignments were performed using Clustal Omega (v1.2.4), and VLY variants were classified according to amino acid variations in the undecapeptide region and CD59-binding domain**,** following the criteria

outlined by Garcia *et al.* (2021) (6). The Guanine-Cytosine (GC) content was calculated using the Science Buddies online GC calculator (https://www.sciencebuddies.org).

Hemolysis assay

The ability of *Gardnerella* spp. to lyse erythrocytes was evaluated using a hemolysis assay. Eight strains with different VLY types, along with one lacking *vly*, were tested on two human blood types (O^- and B^+) to determine wether *Gardnerella* spp. secrete vaginolysin capable of lysing human erythrocytes. The tested strains included: *G. vaginalis* ATCC 49145 and c15Ub_15, both harboring Type 1A VLY, and c21Ua_54 harboring Type 1B; *G. pickettii* c17Ua_112^T harboring Type 1B, and c17Ua_118 lacking *vly*; *G. leopoldii* c1Ub_6 and c18Ua_110, both harboring Type 2 VLY; *G. swidsinskii* c15Ua_15 harboring Type 2 VLY; and *G. greenwoodii* c31Ua_26^T harboring Type 1B VLY.

Strains were cultured in Mueller-Hinton II Agar (Frilabo, France) supplemented with 5% human blood, and incubated at 37ºC for 72 h, under anaerobic conditions. The strain lacking *vly* was used as a negative control, while *G. vaginalis* ATCC 49145 obtained from a woman with BV was used as a positive control. The presence or absence of α or β-haemolysis around the colonies was observed and registered.

Results

Distribution of vaginolysin gene among *Gardnerella*

The *vly* was present in 89.1% (n=131/147) of the analyzed genomes. However, 16 strains from different *Gardnerella* species lacked *vly*, including 8 *G. piotii*, 4 *G. pickettii*, 2 *G.* genomic species 2, 1 *G. vaginalis* and 1 *G.* genomic species 11 (Table S1). Notably, only one *G. piotii* strain among the nine analyzed harbored *vly*.

In most cases, *vly* was 1551 bp in length. However, three *G. vaginalis* strains (ATCC 14019, FDAARGOS 296, and FDAARGOS 568) carried an extended *vly* variant of 1626 bp. To further investigate genetic variability, we analyzed the GC content of genomes and *vly*. The overall *Gardnerella* genome exhibited a GC content ranging from 38.0 to 45.5%, whereas *vly* displayed a GC content, ranging from 44.4 to 49.9% (Table S1).

Characterization of genetic platforms harboring or lacking *vly*

Taking into consideration that *vly* is not part of the *Gardnerella* core genome, we analyzed the regions flanking *vly* to characterize differences in genetic platform composition and organization. As a result, our analysis revealed five distinct *vly* genetic platforms (Fig. 1B–E). The gatCAB complex**,** which encodes enzymes and a membrane transporter protein involved in the N-acylglutamine utilization pathway, was consistently found upstream of *vly*, while an ATP-binding cassette domain-containing protein was always located downstream. The greatest variability was observed downstream of *vly*, including differences in the presence of additional genes**,** as well as variations in the size and identity of intergenic regions (IGRs)**.**

The most frequently identified *vly* genetic platform consisted of the *gatCAB* complex upstream of *vly*, followed by a gene encoding an ABC transporter permease and a gene encoding an ATP-binding cassette domain-containing protein positioned downstream (Fig. 1B). This genetic environment was found in multiple strains across different *Gardnerella* species, including *G. vaginalis, G. pickettii, G. piotii, G. greenwoodii*, *G. leopoldii* (3 strains), *G. swidsinskii* (13 strains), and *G.* genomic species 2, 7, 9, 10, and 14 (Table S1).

Additionally, a genetic platform with a similar gene composition to the one shown in Fig. 1B was identified in 16 strains lacking *vly* (Fig. 1A, Table S1). Further comparative analysis of genetic platforms harboring or lacking *vly* in *G. piotii* strains revealed that the first 63 bp downstream of *gatC* and the 68 bp upstream of the ABC transporter permease aligned. However, overall sequence similarity remained low, ranging from 54.7–57.9% and 58.5–62.3%**,** respectively. On the other hand, a high sequence similarity (95.8-100%) was observed in *G. piotii* strains lacking *vly.*

Distinct modifications to the *vly* genetic platform were observed in fifteen *G. leopoldii* and six *G. swidsinskii* strains. While an ABC domain-containing protein was found immediately downstream *vly* in *G. leopoldii* (Fig. 1C), a hypothetical protein was identified in *G. swidsinskii* (Fig. 1D). Additionally, two unique distinct genetic platforms were identified in *G.* genomic species 12 and 13 (Fig. 1E and 1F). In these species, the genetic platform included an ATP-binding cassette domain-containing protein adjacent to a hypothetical protein upstream of *vly*, with multiple protein-coding genes located downstream. Notably, *G.* genomic species 13 harbored a higher number of protein-coding genes in the region between vaginolysin and the ABC transporter permease.

Further analysis revealed variability in sizes and sequence similarity of ABC transporter permeases and ATP-binding cassette domain-containing proteins (960-1005 bp, 86.8-96.7% and 963-2085 bp, 76.5-90.4%, respectively) (Fig. 1*).*

Analysis of IGR sizes between *gatC* and *vly* revealed variability both across and within *Gardnerella* species. While some species maintained a consisted IGR size, others displayed a greater variability. For example, *G. swidsinskii* exhibited a highly conserved IGR of 498 bp across all strains, whereas in *G. picketii*, IGR sizes ranged from 381 to 498 bp were detected. Interestingly, *G.* genomic species 12 and 13 harbored the largest IGR sizes (711 and 710 bp, respectively)**.**

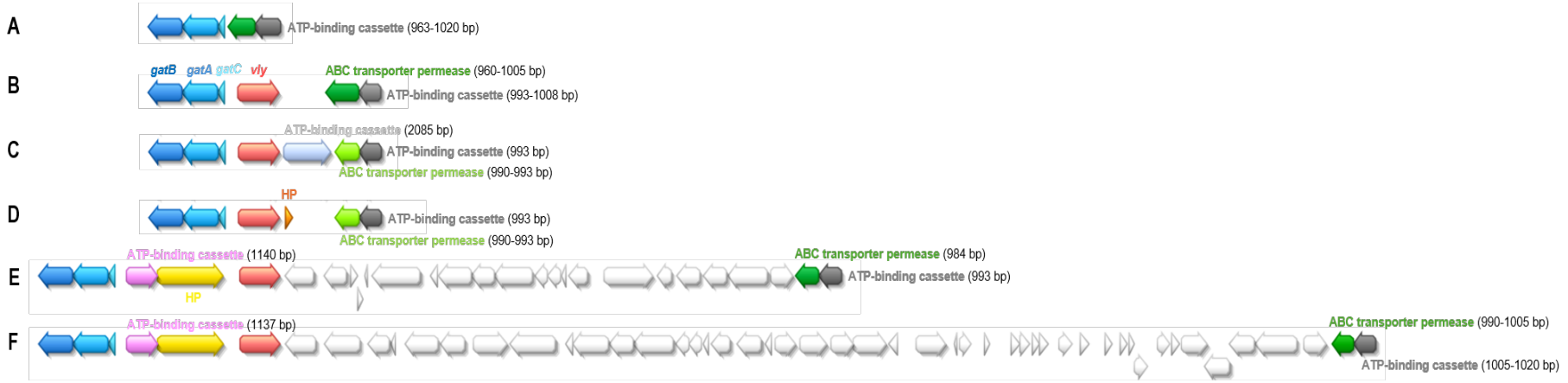


Figure 1. Schematic representation of genetic platforms harboring or lacking *vly* across *Gardnerella* spp**. A.** Genetic platform lacking *vly*. **B.** Genetic platform identified across multiple strains of *G. vaginalis, G. pickettii, G. piotii, G. swidsinskii* (13 strains), *G. leopoldii* (3 strains), *G. greenwoodii,* and *Gardnerella* genomic species 2, 7, 9, 10 and 14. **C.** Genetic platform identified in *G. leopoldii* (15 strains). **D.** Genetic platform identified in six *G. swidsinskii* strains. **E.** Genetic platform identified in *Gardnerella* genomic species 12. **F.** Genetic platform identified in *Gardnerella* genomic species 13.

VLY Types within *Gardnerella*

Garcia *et al*. (2021) (6) classified VLY into five types based on the amino acid sequence of the undecapeptide domain and CD59-binding region, both of which are involved in host cell recognition and pore formation. In our analysis of 131 VLY sequences from *Gardnerella*, Type 1A was the most prevalent (n=57/131; 43.5%), followed by Type 1B (n=34/131; 26.0%) and Type 2 (n=37/131; 28.2%). In contrast, Type 1C (n=1/131; 0.8%) and Type 3 (n=2/131; 1.5%) were rare (Table S1).

Previously, vaginal symptoms were reported as potentially associated with VLY type, with 65% of subjects experiencing at least one symptom (e.g., itching, odor, or discharge) and 35% being asymptomatic (6). In this study, we did not have access to vaginal symptom data for the strains corresponding to our analyzed genomes. While donor health status information was available, no clear association was found between VLY types and donor health conditions (Table S1).

Analysis of the VLY amino acid sequence of *G.* genomic species 12 strain KA00735 revealed characteristics of Type 3 VLY, including valine (V) and leucine (L) at positions -1 and 6 of the undecapeptide region, and serine (S) at position 6 of CD59-binding domain. However, this sequence also contained an aspartate (D) at position 4 of the CD59-binding domain instead of the asparagine (N) found in previously classified VLY Type 3 sequences. This variation suggests the presence of distinct subtypes within Type 3 VLY, prompting us to propose a subclassification into Type 3A (previous identified Type 3) and Type 3B (first identified in strain KA00735) (Fig. 2).

	471 bp	*Undecapeptide*											*CD59-binding*						489 bp	
Residue position	-1	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	
Type 1 A	V	E	K	T	G	L	V	W	E	P	W	R	T	V	Y	D	R	S	D	
Type 1 B	V	E	K	T	G	L	A	W	E	P	W	R	T	V	Y	D	R	S	D	
Type 1 C	Q	E	K	T	G	L	V	W	E	P	W	R	T	V	Y	D	R	S	D	
Type 2	E	E	K	T	G	L	V	W	E	P	W	R	T	V	Y	N	R	T	D	
Type 3 A	V	E	K	T	G	L	L	W	E	P	W	R	T	V	Y	N	R	S	D	
Type 3 B	V	E	K	T	G	L	L	W	E	P	W	R	T	V	Y	D	R	S	D	
CMW7778B	V	E	K	T	G	L	L	W	E	P	W	R	T	V	Y	N	R	S	D	GG12, HW (pregnant), vagina
KA00735	V	E	K	T	G	L	L	W	E	P	W	R	T	V	Y	D	R	S	D	GG12, NK, vagina
Marseille--Q2328	V	E	K	T	G	L	V	W	E	P	W	R	T	V	Y	D	R	S	D	GG14, NK, vagina
NR0101	V	E	K	T	G	L	A	W	E	P	W	R	T	V	Y	D	R	S	D	GG14, NK, NK

Figure 2. Classification of VLY Types based on undecapeptide and CD59-binding region

sequences. The previously known VLY Type 3 was subclassified into Type 3A (light green) and Type 3B (dark green) due to a key amino acid substitution at position 4 of the CD59-binding region [asparagine (N) to aspartate (D)]. GG12, *Gardnerella* genomic species 12; GG14, *Gardnerella* genomic species 14; NK, donor health status or sample source unknown.

Vaginolysin Activity Assays

None of the VLY sequences corresponding to the *Gardnerella* strains used in the activity assays contained the P480W alteration, a mutation previously associated with loss of pore-formed function (9). This finding leads us to hypothesize that all VLY-producing strains used in the assay harbored a functional pore-forming protein with hemolytic activity potential. The vaginolysin activity assay confirmed that *vly*-negative strain exhibited no hemolysis, as expected. In contrast, strains harboring *vly* exhibited either α- or β-hemolysis, depending on the blood sample used (Fig. 3).

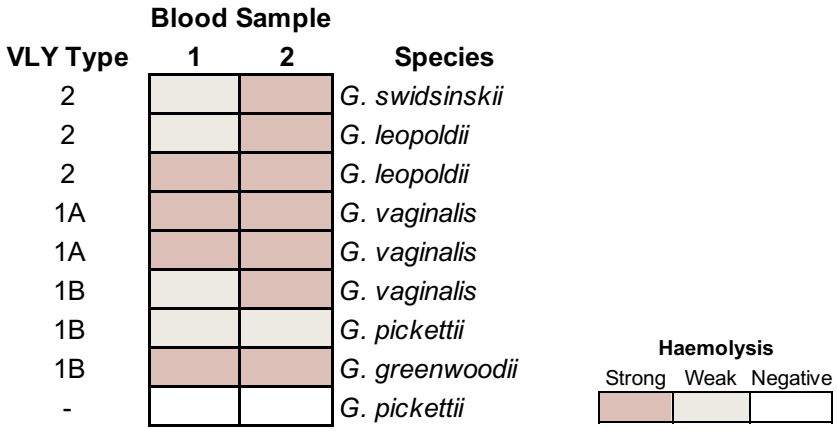


Figure 3. A hemolytic activity assay was performed to assess VLY-induced pore formation of *Gardnerella* strains in two different human blood samples.

Discussion

Our comprehensive analysis of 147 genomes provides novel insights into the distribution, genetic context, and evolution of *vly* across *Gardnerella* species.

The presence of *vly* in 89.1% of analyzed genomes supports previous findings that *vly* is not a core gene but rather an accessory component of *Gardnerella* genome (6, 8). Notably, *vly* is absent in several strains, particularly in *G. piotii*, consistent with findings by Bohr *et al*. (2020) (8). This suggests that while VLY plays a role in *Gardnerella*-associated infections, it is not universally required for survival or virulence. The absence of *vly* in certain strains could be

compensated by other virulence factors, metabolic capabilities, or adaptations that allow them to persist in the vaginal microbiome.

The five distinct *vly* genetic platforms, each characterized by differences in flanking gene composition and IGR size, expanded our understanding on its genomic context. The consistent presence of the *gatCAB* complex upstream of *vly* suggests a potential functional relationship between N-acylglutamine utilization and vaginolysin production. Furthermore, the high sequence variability downstream of *vly*, coupled with differences in IGR sizes, suggests that recombination events or gene insertions may have shaped the genetic architecture of *vly* over time (6, 8).

The classification of VLY into five types, with Type 1A being the most prevalent, aligns with previous studies (6). However, the identification of Type 3B VLY in *Gardnerella* genomic species 12 suggests further diversification within this family of toxins. This subclassification highlights the potential for ongoing mutations and adaptive changes in VLY structure**,** which may affect its interaction with host cells. The amino acid substitution in the CD59-binding region observed in Type 3B VLY suggests that modifications in this domain could alter its affinity for human cells, potentially impacting cytotoxicity and host immune evasion.

Our hemolysis assays confirmed that strains lacking *vly* did not exhibit hemolytic activity**,** supporting the role of VLY as a key determinant of *Gardnerella*-associated cytotoxicity (9, 10). However, the fact that strains producing VLY exhibited varying degrees of α- or β-hemolysis**,** depending on the blood sample used, suggests that VLY function may be influenced by host factors, including differences in CD59 expression or red blood cell membrane composition.

Conclusions

This study underscores the genetic plasticity of *vly* and its integration within diverse genomic contexts across *Gardnerella* species. The identification of a novel VLY subtype, combined with variations in genetic platforms and hemolytic activity, highlights the complexity of *Gardnerella* pathogenesis. Further research is needed to elucidate the functional significance of VLY variability and its implications for host-pathogen interactions, virulence regulation, and BV pathophysiology. Understanding these dynamics could lead to improved diagnostic markers and targeted therapeutic strategies for *Gardnerella*-associated diseases.

Funding

This work is 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. MS, FG, TGR are supported by FCT 2020.05038.BD, DL57/2016/CP1346/CT0034, UIDP/QUI/04378/2020, respectively. L.F is supported by Bolsa do Prémio Maratona da Saúde (2023/8, N.º 26).

Author Contributions

MS performed the isolates isolation, genomic DNA extraction, annotation, and species identification, and wrote the original draft of the manuscript. LF performed the Hemolysis assay FG and TGR contributed to study design, reviewing the manuscript and supervision. LP contributed to reviewing the manuscript, project design, administration, and funding.

Competing Interests

The authors declare that there are no conflicts of interest.

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.

References

1. Munch MM, Strenk SM, Srinivasan S, Fiedler TL, Proll S, Fredricks DN. *Gardnerella* Species and Their Association With Bacterial Vaginosis. J Infect Dis. 2024 Jul 25;230(1):e171-e181.
2. 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 Mar;69(3):679-687.
3. Sousa M, Ksiezarek M, Perovic SU, Antunes-Lopes T, Grosso F, Ribeiro TG, Peixe L. *Gardnerella pickettii* sp. nov. (formerly *Gardnerella* genomic species 3) and *Gardnerella greenwoodii* sp. nov. (formerly *Gardnerella* genomic species 8) isolated from female urinary microbiome. Int J Syst Evol Microbiol. 2023 Nov;73(11).
4. Ugarcina Perovic S, Ksiezarek M, Rocha J, Cappelli EA, Sousa M, Ribeiro TG, Grosso F, Peixe L. Urinary Microbiome of Reproductive-Age Asymptomatic European Women. Microbiol Spectr. 2022 Dec 21;10(6):e0130822.
5. Janulaitiene M, Paliulyte V, Grinceviciene S, Zakareviciene J, Vladisauskiene A, Marcinkute A, Pleckaityte M. Prevalence and distribution of *Gardnerella vaginalis* subgroups in women with and without bacterial vaginosis. BMC Infect Dis. 2017 Jun 5;17(1):394.
6. Garcia EM, Serrano MG, Edupuganti L, Edwards DJ, Buck GA, Jefferson KK. Sequence Comparison of Vaginolysin from Different *Gardnerella* Species. Pathogens. 2021 Jan 20;10(2):86.
7. Shishpal P, Kasarpalkar N, Singh D, Bhor VM. Characterization of *Gardnerella vaginalis* membrane vesicles reveals a role in inducing cytotoxicity in vaginal epithelial cells. Anaerobe. 2020 Feb;61:102090.
8. Bohr LL, Mortimer TD, Pepperell CS. Lateral Gene Transfer Shapes Diversity of *Gardnerella* spp. Front Cell Infect Microbiol. 2020 Jun 23;10:293.
9. Gelber SE, Aguilar JL, Lewis KL, Ratner AJ. Functional and phylogenetic characterization of Vaginolysin, the human-specific cytolysin from *Gardnerella vaginalis.* J Bacteriol. 2008 Jun;190(11):3896-903.
10. Castro J, Alves P, Sousa C, Cereija T, França Â, Jefferson KK, Cerca N. Using an *in-vitro* biofilm model to assess the virulence potential of bacterial vaginosis or non-bacterial vaginosis *Gardnerella vaginalis* isolates. Sci Rep. 2015 Jun 26;5:11640.

Supplementary material

Table S1: Genomic features of *Gardnerella* species and genomic species, detailed vaginolysin characterization.

Species	Strain	Assembly accession number	GC (%)	Host disease	Isolation Source	*vly* GC content (%)	*vly* genetic platform[a]	VLY Type[b]
G. vaginalis	c11Ua_71	JARFVR000000000	41.3	Asymptomatic	MSU	46.6	B	1A
G. vaginalis	c15Ub_15	JANUYY000000000	41.2	Asymptomatic	MSU	46.4	B	1A
G. vaginalis	c15Ub_3	JAODXZ000000000	41.3	Asymptomatic	MSU	46.4	B	1A
G. vaginalis	c15Ub_5	JANUZA000000000	41.2	Asymptomatic	MSU	46.4	B	1A
G. vaginalis	c15VSa_39	JANUYX000000000	41.2	Asymptomatic	VS	46.4	B	1A
G. vaginalis	c15VSb_14	JANUYU000000000	41.1	Asymptomatic	VS	46.4	B	1A
G. vaginalis	c15VSb_4	JANUYW000000000	41.2	Asymptomatic	VS	46.4	B	1A
G. vaginalis	c21Ua_54	JANUYR000000000	41.3	OAB	MSU	48.8	B	1B
G. vaginalis	c29Ua_32	JANUYN000000000	41.2	Asymptomatic	MSU	46.5	B	1A
G. vaginalis	c29Ub_7	JANUYM000000000	41.3	Asymptomatic	MSU	46.5	B	1A
G. vaginalis	c29VSa_19	JANUYL000000000	41.2	Asymptomatic	VS	46.5	B	1A
G. vaginalis	c29VSb_3	JANUYK000000000	41.3	Asymptomatic	VS	46.5	B	1A
G. vaginalis	75712	GCF_000263535.1	41.3	BV	VS	48.4	B	1A
G. vaginalis	0288E	GCF_000263555.1	41.2	BV	VS/ Endometrium	46.4	B	1A
G. vaginalis	06-12-0010	GCF_014857145.1	41.2	NA	VS	46.4	B	1A

G. vaginalis	14018c	GCF_004336715.1	41.3	BV	VS	46.6	B	1A
G. vaginalis	14019_MetR	GCF_001278345.1	41.3	BV	NA	46.6	B	1A
G. vaginalis	284V	GCF_000263435.1	41.2	BV	VS/ Endometrium	46.6	B	1A
G. vaginalis	315-A	GCF_000214315.1	41.4	BV	VS	49.7	B	1B
G. vaginalis	ATCC 14019	GCF_000159155.2	41.4	BV	VS	46.6	B	1A
G. vaginalis	ATCC 49145	GCF_003034925.1	41.2	Intermediate BV	VS	46.4	B	1A
G. vaginalis	DNF01149	GCF_002894105.1	41.2	NA	VS	47.6	B	1A
G. vaginalis	FDAARGOS_296	GCF_002206225.1	41.3	BV	VS	46.4	B	1A
G. vaginalis	FDAARGOS_568	GCF_003812765.1	41.3	BV	VS	46.4	B	1A
G. vaginalis	GH015	GCF_003408745.1	41.0	NA	VS	46.5	B	1A
G. vaginalis	HMP9231	GCF_000213955.1	41.2	NA	VS	46.6	B	1A
G. vaginalis	JCP7276	GCF_000414685.1	41.0	Intermediate BV	VS	46.4	B	1A
G. vaginalis	JCP7672	GCF_000414645.1	41.2	Asymptomatic	VS	46.5	B	1A
G. vaginalis	JNFY1	GCF_023277725.1	41.6	BV	VS	46.7	B	1A
G. vaginalis	JNFY11	GCF_023277645.1	41.8	BV	VS	46.6	B	1A
G. vaginalis	JNFY13	GCF_023277625.1	41.6	BV	VS	49.7	B	1B
G. vaginalis	JNFY14	GCF_023277605.1	41.7	BV	VS	49.7	B	1B
G. vaginalis	JNFY4	GCF_023277685.1	41.6	BV	VS	46.5	B	1A
G. vaginalis	JNFY9	GCF_023277665.1	41.4	BV	VS	49.7	B	1B
G. vaginalis	KC2	GCF_023016185.1	41.3	BV	VS	49.7	B	1B
G. vaginalis	KC3	GCF_023016245.1	41.2	BV	VS	46.4	B	1A

G. vaginalis	KC4	GCF_023016225.1	41.3	BV	VS	46.6	B	1A
G. vaginalis	NCTC10287	GCF_900637625.1	41.4	BV	VS	46.6	B	1A
G. vaginalis	NR038	GCF_003585655.1	42.7	NA	VS	46.6	B	1A
G. vaginalis	UGent 09.07	GCF_003397665.1	41.1	BV	VS	46.7	B	1A
G. vaginalis	UGent 25.49	GCF_003397605.1	41.2	BV	VS	48.2	B	1A
G. vaginalis	UMB0032A	GCF_002862015.1	41.2	Asymptomatic	MSU	49.7	B	1B
G. vaginalis	UMB0032B	GCF_002862005.1	41.2	Asymptomatic	MSU	49.7	B	1B
G. vaginalis	UMB0061	GCF_002861165.1	41.2	OAB	MSU	47.5	B	1B
G. vaginalis	UMB0143	GCF_013315005.1	41.3	OAB	MSU	46.3	B	1A
G. vaginalis	UMB0202	GCF_013315075.1	41.4	OAB	MSU	-	A	-
G. vaginalis	UMB0233	GCF_002862045.1	41.2	OAB	MSU	49.8	B	1B
G. vaginalis	UMB0298	GCF_002861975.1	41.2	OAB	MSU	46.4	B	1A
G. vaginalis	UMB0358	GCF_013315085.1	41.2	OAB	MSU	46.7	B	1A
G. vaginalis	UMB0386	GCF_002861965.1	41.2	OAB	MSU	46.4	B	1A
G. vaginalis	UMB0540	GCF_013315045.1	41.3	OAB	MSU	49.7	B	1B
G. vaginalis	UMB0736	GCF_013315025.1	41.1	Kidney Stone	MSU	46.6	B	1A
G. vaginalis	UMB0768	GCF_002884835.1	41.2	NA	TUC	46.6	B	1A
G. vaginalis	UMB0770	GCF_002861945.1	41.2	OAB	MSU	46.7	B	1A
G. vaginalis	UMB0775	GCF_002861925.1	41.2	SUI	MSU	46.6	B	1A
G. genomic species 2	JCP7275	GCF_000414705.1	41.0	BV	VS	-	A	-

G. genomic species 2	N165	GCF_003408785.1	41.4	NA	VS	-	A	-
G. genomic species 2	55152	GCF_000263475.1	41.3	BV	VS/Endometrium	49.3	B	1B
G. genomic species 2	1400E	GCF_000263495.1	41.2	BV	VS/Endometrium	49.3	B	1B
G. genomic species 2	41V	GCF_000165635.1	41.3	Asymptomatic	VS	49.3	B	1B
G. genomic species 2	JCP8108	GCF_000414525.1	41.1	BV	VS	49.3	B	1B
G. pickettii	c10Ua_31	JANUYG010000000	42.3	Asymptomatic	MSU	49.1	B	1A
G. pickettii	c12Ub_1	JANUZC000000000	42.4	Asymptomatic	MSU	49.6	B	1B
G. pickettii	c17Ua_112	GCF_029226345.1	42.5	OAB	MSU	46.9	B	1A
G. pickettii	c17Ua_118	JANUYT000000000	42.3	OAB	MSU	-	A	-
G. pickettii	c23Ua_15	JANUYQ000000000	42.5	Asymptomatic	MSU	46.9	B	1A
G. pickettii	c23Ua_38	JANUYP000000000	42.5	Asymptomatic	MSU	46.9	B	1A
G. pickettii	c23Ub_4	JANUYO000000000	42.3	Asymptomatic	MSU	49.6	B	1B
G. pickettii	JNFY17	GCF_023277565.1	42.7	BV	VS	49.6	B	1B
G. pickettii	00703Bmash	GCF_000263615.1	42.3	BV and HSV-2	VS	48.5	B	1A
G. pickettii	00703C2mash	GCF_000263515.1	42.3	BV and HSV-2	VS	48.0	B	1A
G. pickettii	GED7275B	GCF_001546445.1	42.5	Asymptomatic	VS	46.9	B	1A
G. pickettii	JCP7659	GCF_000414665.1	41.9	BV	VS	48.0	B	1A
G. pickettii	JCP7719	GCF_000414625.1	42.0	BV	VS	49.6	B	1B

G. pickettii	JCP8017A	GCF_000414605.1	42.1	BV	VS	49.1	B	1A
G. pickettii	JCP8017B	GCF_000414585.1	42.0	BV	VS	49.1	B	1A
G. pickettii	N101	GCF_003369895.1	42.4	NA	VS	49.6	B	1B
G. pickettii	N144	GCF_003408835.1	41.3	NA	VS	48.0	B	1A
G. pickettii	N153	GCF_003369935.1	42.4	NA	VS	49.6	B	1B
G. pickettii	N95	GCF_003369965.1	42.4	NA	VS	49.6	B	1B
G. pickettii	UMB0558	GCF_013315115.1	42.3	SUI	MSU	-	A	-
G. pickettii	UMB0830	GCF_002861905.1	42.3	Asymptomatic	MSU	49.7	B	1B
G. pickettii	UMB0833	GCF_002861885.1	42.1	Asymptomatic	MSU	-	A	-
G. pickettii	W11	GCF_003369875.1	42.3	NA	VS	-	A	-
G. piotii	GH019	GCF_003426565.1	42.7	BV	VS	-	A	-
G. piotii	JCP8066	GCF_000414565.1	42.2	Asymptomatic	VS	-	A	-
G. piotii	JCP8070	GCF_000414545.1	42.2	BV	VS	-	A	-
G. piotii	JCP8151A	GCF_000414505.1	42.0	BV	VS	-	A	-
G. piotii	JCP8151B	GCF_000414485.1	42.2	BV	VS	49.6	B	1B
G. piotii	JCP8522	GCF_000414425.1	42.2	BV	VS	-	A	-
G. piotii	JNFY15	GCF_023277585.1	42.5	BV	VS	-	A	-
G. piotii	UGent 18.01	GCF_003397585.1	42.5	BV	VS	-	A	-
G. piotii	UGent 21.28	GCF_003397615.1	42.5	BV	VS	-	A	-
G. leopoldii	c1Ub_6	JANUUT010000000	42.3	Asymptomatic	MSU	47.2	C	2
G. leopoldii	c1Ub_40	JAOPFO000000000	42.2	Asymptomatic	MSU	47.2	C	2

G. leopoldii	c18Ua_110	JANUYS000000000	42.2	OAB	MSU	47.3	C	2
G. leopoldii	c18Ua_112	JAPDSJ000000000	42.3	OAB	MSU	47.3	C	2
G. leopoldii	c31Ua_40	JANUYJ000000000	42.1	Asymptomatic	MSU	47.1	B	2
G. leopoldii	c31VSa_14	JANUYI000000000	42.1	Asymptomatic	VS	47.1	B	2
G. leopoldii	c31VSa_35	JARFVP000000000	42.1	Asymptomatic	VS	47.1	B	2
G. leopoldii	6420B	GCF_000263575.1	42.2	Asymptomatic	VS/Endometrium	47.2	C	2
G. leopoldii	AMD	GCF_000176475.1	42.1	BV	VS	47.1	C	2
G. leopoldii	UGent 06.41	GCF_003293675.1	42.1	BV	VS	47.2	C	2
G. leopoldii	UGent 09.48	GCF_003397635.1	42.2	BV	VS	47.2	C	2
G. leopoldii	UMB0662	GCF_013315255.1	42.1	NA	TUC	47.2	C	2
G. leopoldii	UMB0682	GCF_002862065.1	42.1	OAB	MSU	47.3	C	2
G. leopoldii	UMB0742	GCF_013315135.1	42.2	OAB	MSU	47.3	C	2
G. leopoldii	UMB0912	GCF_002861125.1	42.1	Asymptomatic	MSU	47.2	C	2
G. leopoldii	UMB0913	GCF_002861145.1	42.1	Asymptomatic	MSU	47.2	C	2
G. leopoldii	UMB1350	GCF_013315125.1	42.1	OAB	MSU	47.2	C	2
G. leopoldii	UMB1489	GCF_013315195.1	42.2	NA	TUC	47.2	C	2
G. swidsinskii	c10Ua_244	JARFVS000000000	42.0	Asymptomatic	MSU	47.5	B	2
G. swidsinskii	c15Ua_15	JAPDSI000000000	42.0	Asymptomatic	MSU	47.6	B	2
G. swidsinskii	c15Ua_23	JANUZB000000000	42.0	Asymptomatic	MSU	47.6	B	2
G. swidsinskii	c15Ub_13	JANUYZ000000000	42.0	Asymptomatic	MSU	47.6	B	2
G. swidsinskii	c15VSb_13	JANUYV000000000	42.0	Asymptomatic	VS	47.6	B	2

G. swidsinskii	c15VSa_36	JARFVQ000000000	42.0	Asymptomatic	VS	47.6	B	2
G. swidsinskii	409-05	GCF_000025205.1	42.0	Asymptomatic	VS	46.8	B	2
G. swidsinskii	5-1	GCF_000176495.1	42.0	Asymptomatic	VS	47.6	B	2
G. swidsinskii	DNF01162	GCF_002894125.1	41.7	NA	VS	47.6	B	2
G. swidsinskii	GS 10234	GCF_003397745.1	41.9	BV	VS	46.9	D	2
G. swidsinskii	GS 9838-1	GCF_003397705.1	41.9	BV	VS	47.1	B	2
G. swidsinskii	GV37	GCF_001953155.1	41.8	Bacteremia	B	46.8	D	2
G. swidsinskii	JNFY3	GCF_023277705.1	42.0	BV	VS	47.5	B	2
G. swidsinskii	N72	GCF_003408815.1	41.9	BV	VS	47.6	B	2
G. swidsinskii	UMB0170	GCF_002884855.1	42.3	NA	TUC	46.8	D	2
G. swidsinskii	UMB0264	GCF_002884875.1	42.3	NA	TUC	46.8	D	2
G. swidsinskii	UMB0769	GCF_013315215.1	42.0	OAB	MSU	47.5	D	2
G. swidsinskii	UMB1642	GCF_002884795.1	41.8	Asymptomatic	MSU	47.6	B	2
G. swidsinskii	UMB1698	GCF_013315145.1	41.9	Diabetes	MSU	46.8	D	2
*G.*genomic species 7	JCP8481A	GCF_000414465.1	42.9	BV	VS	48.0	B	1B
*G.*genomic species 7	JCP8481B	GCF_000414445.1	42.9	BV	VS	48.0	B	1B
*G.*genomic species 7	PSS_7772B	GCF_001546485.1	42.9	Asymptomatic	U	47.6	B	1C
G. greenwoodii	c31Ua_26	GCF_029207615.1	43.3	Asymptomatic	MSU	49.5	B	1B
G. greenwoodii	101	GCF_000165615.1	43.4	BV	VS	49.2	B	1B
G. greenwoodii	00703Dmash	GCF_000263635.1	43.4	BV and HSV-2	VS	49.3	B	1B
G. greenwoodii	KC1	GCF_023016205.1	43.3	BV	VS	49.5	B	1B

G. greenwoodii	UMB1686	GCF_002884775.1	43.3	Diabetes	MSU	49.5	B	1B
G. genomic species 9	6119V5	GCF_000263655.1	43.3	Asymptomatic	VS	49.1	B	1A
G. genomic species 9	N160	GCF_003408775.1	43.3	NA	VS	49.1	B	1A
G. genomic species 10	1500E	GCF_000263595.1	43.0	BV	VS/Endometrium	49.4	B	1B
G. genomic species 11	GED7760B	GCF_001546455.1	43.3	NA	VS	-	A	-
G. genomic species 12	CMW7778B	GCF_001563665.1	38.0	Asymptomatic	VS	44.7	E	3 A
G. genomic species 12	KA00735	GCF_002894085.1	38.2	NA	VS	44.7	E	3 B
G. genomic species 13	KA00225	GCF_002896555.1	40.8	NA	VS	45.0	F	1B
G. genomic species 14	Marseille-Q2328	GCF_940796375.1	45.3	NA	NA	49.9	B	1A
G. genomic species 14	NR010	GCF_003408845.1	45.5	BV	VS	49.9	B	1B

Abbreviations: BV, bacterial vaginosis; HSV-2, herpes simplex virus type 2; MSU, midstream urine; OAB, overactive bladder; SUI: stress urinary incontinence; TUC, transurethral catheter; VS: vaginal swab, [a]Genetic platforms identified across *Gardnerella* strains harboring or lacking *vly*, [b]Classification of VLY Type based on undecapeptide and CD59-binding region sequences.

Table S2: Intergenic region sizes in *vly* genetic platform across *Gardnerella* strains.

Species	Strain name	IGR size between *gatC* and *vly* (bp)	IRG size between *vly* and the next downstream gene (bp)
G. vaginalis	c11Ua_71	500	129
G. vaginalis	c15Ub_15	500	152
G. vaginalis	c15Ub_3	500	152
G. vaginalis	c15Ub_5	500	151
G. vaginalis	c15VSa_39	500	151
G. vaginalis	c15VSb_14	500	152
G. vaginalis	c15VSb_4	500	152
G. vaginalis	c21Ua_54	499	122
G. vaginalis	c29Ua_32	502	142
G. vaginalis	c29Ub_7	502	142
G. vaginalis	c29VSa_19	502	142
G. vaginalis	c29VSb_3	502	142
G. vaginalis	75712	501	142
G. vaginalis	0288E	500	142
G. vaginalis	06-12-0010	500	142
G. vaginalis	14018c	501	142
G. vaginalis	14019_MetR	500	142
G. vaginalis	284V	501	149
G. vaginalis	315-A	498	142

G. vaginalis	ATCC 14019	501	142
G. vaginalis	ATCC 49145	500	142
G. vaginalis	DNF01149	498	152
G. vaginalis	FDAARGOS_296	500	142
G. vaginalis	FDAARGOS_568	500	142
G. vaginalis	GH015	502	142
G. vaginalis	HMP9231	500	142
G. vaginalis	JCP7276	500	149
G. vaginalis	JCP7672	498	152
G. vaginalis	JNFY1	500	142
G. vaginalis	JNFY11	500	142
G. vaginalis	JNFY13	498	142
G. vaginalis	JNFY14	498	142
G. vaginalis	JNFY4	500	142
G. vaginalis	JNFY9	498	142
G. vaginalis	KC2	498	142
G. vaginalis	KC3	500	142
G. vaginalis	KC4	501	142
G. vaginalis	NCTC10287	501	142
G. vaginalis	NR038	500	149
G. vaginalis	UGent 09.07	500	142
G. vaginalis	UGent 25.49	498	142

G. vaginalis	UMB0032A	498	142
G. vaginalis	UMB0032B	498	142
G. vaginalis	UMB0061	500	142
G. vaginalis	UMB0143	500	142
G. vaginalis	UMB0202	144*	
G. vaginalis	UMB0233	498	142
G. vaginalis	UMB0298	500	142
G. vaginalis	UMB0358	500	142
G. vaginalis	UMB0386	500	142
G. vaginalis	UMB0540	503	142
G. vaginalis	UMB0736	500	142
G. vaginalis	UMB0768	500	129
G. vaginalis	UMB0770	500	142
G. vaginalis	UMB0775	500	129
G. genomic species 2	JCP7275	144*	
G. genomic species 2	N165	145*	
G. genomic species 2	55152	499	151
G. genomic species 2	1400E	499	151
G. genomic species 2	41V	499	192
G. genomic species 2	JCP8108	500	192
G. pickettii	c10Ua_31	497	142
G. pickettii	c12Ub_1	498	142

G. pickettii	c17Ua_112	497	154
G. pickettii	c17Ua_118	135*	
G. pickettii	c23Ua_15	497	154
G. pickettii	c23Ua_38	497	154
G. pickettii	c23Ub_4	498	142
G. pickettii	JNFY17	498	142
G. pickettii	00703Bmash	498	127
G. pickettii	00703C2mash	381	142
G. pickettii	GED7275B	497	154
G. pickettii	JCP7659	382	142
G. pickettii	JCP7719	498	142
G. pickettii	JCP8017A	498	142
G. pickettii	JCP8017B	498	142
G. pickettii	N101	497	142
G. pickettii	N144	382	143
G. pickettii	N153	498	142
G. pickettii	N95	497	142
G. pickettii	UMB0558	119*	
G. pickettii	UMB0830	498	142
G. pickettii	UMB0833	147*	
G. pickettii	W11	135*	
G. piotii	GH019	135*	

G. piotii	JCP8066	119*	
G. piotii	JCP8070	119*	
G. piotii	JCP8151A	134*	
G. piotii	JCP8151B	498	142
G. piotii	JCP8522	135*	
G. piotii	JNFY15	135*	
G. piotii	UGent 18.01	134*	
G. piotii	UGent 21.28	119*	
G. leopoldii	c1Ub_6	500	138
G. leopoldii	c1Ub_40	500	138
G. leopoldii	c18Ua_110	500	138
G. leopoldii	c18Ua_112	500	138
G. leopoldii	c31Ua_40	500	858
G. leopoldii	c31VSa_14	500	858
G. leopoldii	c31VSa_35	500	858
G. leopoldii	6420B	500	138
G. leopoldii	AMD	500	138
G. leopoldii	UGent 06.41	500	138
G. leopoldii	UGent 09.48	500	138
G. leopoldii	UMB0662	500	138
G. leopoldii	UMB0682	500	138
G. leopoldii	UMB0742	500	138

G. leopoldii	UMB0912	500	138
G. leopoldii	UMB0913	500	138
G. leopoldii	UMB1350	500	138
G. leopoldii	UMB1489	500	138
G. swidsinskii	c10Ua_244	498	848
G. swidsinskii	c15Ua_15	498	860
G. swidsinskii	c15Ua_23	498	848
G. swidsinskii	c15Ub_13	498	860
G. swidsinskii	c15VSb_13	498	860
G. swidsinskii	c15VSa_36	498	860
G. swidsinskii	409-05	498	912
G. swidsinskii	5-1	498	857
G. swidsinskii	DNF01162	498	860
G. swidsinskii	GS 10234	498	913
G. swidsinskii	GS 9838-1	498	857
G. swidsinskii	GV37	498	913
G. swidsinskii	JNFY3	498	848
G. swidsinskii	N72	498	857
G. swidsinskii	UMB0170	498	912
G. swidsinskii	UMB0264	498	912
G. swidsinskii	UMB0769	498	862
G. swidsinskii	UMB1642	498	859

G. swidsinskii	UMB1698	498	913
*G.*genomic species 7	JCP8481A	522	243
*G.*genomic species 7	JCP8481B	522	244
*G.*genomic species 7	PSS_7772B	522	148
G. greenwoodii	c31Ua_26	381	139
G. greenwoodii	101	382	144
G. greenwoodii	00703Dmash	382	147
G. greenwoodii	KC1	381	139
G. greenwoodii	UMB1686	381	139
G. genomic species 9	6119V5	381	147
G. genomic species 9	N160	381	147
G. genomic species 10	1500E	381	148
G. genomic species 11	GED7760B	135*	
G. genomic species 12	CMW7778B	711	228
G. genomic species 12	KA00735	711	329
G. genomic species 13	KA00225	710	203
G. genomic species 14	Marseille-Q2328	548	235
G. genomic species 14	NR010	548	213

*IGR size between *gatC* and the following downstream gene in the absence of *vly*.

Unveiling *Gardnerella*'s glycogen-degrading arsenal: Novel neopullulanase and metabolic machinery insights

Márcia Sousa[1,2], Teresa Gonçalves Ribeiro[1,2,3], Luísa Peixe [1,2,3 *], Filipa Grosso[1,2]

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

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

[3]Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP).

*Luísa Peixe, lpeixe@ff.up.pt

Manuscript in preparation

Abstract

Gardnerella species exhibit diverse enzymatic tools and carbohydrate transporters, enabling efficient metabolism of glycogen and complex sugars. In this study, we characterized glycoside hydrolase family 13 (GH13) enzymes and carbohydrate-binding modules (CBMs) across 33 *Gardnerella* strains, representing five species, using genome-based *in silico* analysis and, for selected strains, phenotypic characterization.

We report the first identification of neopullulanase (GH13_20 + CBM34) in *Gardnerella*, found in 51.52% of genomes, exclusively in *G. vaginalis* and *G. pickettii*, highlighting their capacity to hydrolyze α-1,4 and α-1,6-glycosidic linkages in glycogen. Glycogen debranching enzyme (GH13_11 + CBM48) was also prevalent, detected in all species except those that harbored alternative CBMs. The α-amylase domain (GH13_32), often accompanied by CBMs, was identified across most strains. *G. leopoldii* shared maltose and maltodextrin transporters with *G. swidsinskii* and *G. greenwoodii*, reflecting adaptations to diverse carbohydrate sources. Amylase activity, confirmed via iodine assay, aligned with genomic predictions.

The ability to efficiently metabolize glycogen likely contributes to *Gardnerella*'s role in shaping the vaginal microbiome, influencing resource availability for coexisting species.

Keywords: *Gardnerella,* Glycogen metabolism, Neopullulanase, carbohydrate transporters

Introduction

The urogenital microbiome frequently harbors *Gardnerella* species, which coexist at different abundances, likely due to competition for key nutrients such as glycogen, a primary energy source in the vaginal epithelium. Glycogen, a glucose homopolysaccharide, consists of linear α-1,4 glycosidic bonds with periodic α-1,6 branches approximately every 8–13 residues (1-3). Estrogen regulates glycogen accumulation in the vaginal epithelium, with levels ranging from undetectable to 32 µg/mL, depending on hormonal fluctuations and individual variation (1-3). Glycogen degradation and its conversion into oligosaccharides are crucial for microbial colonization, biofilm formation, and energy metabolism in vaginal bacteria (3,4). Several bacterial species, including *Gardnerella*, harbor glycogen-debranching enzymes, such as amylases and pullulanases, which may complement human amylases in the vagina (2). Interestingly, these enzymes have been detected even in postmenopausal women, despite their lower estrogen levels (1). Additionally, α-amylase appears to be associated with vaginal health,

showing the lowest levels in women with bacterial vaginosis (BV) and the highest in healthy individuals (2,5). The role of glycogen metabolism extends beyond the vaginal microbiome, as a deficiency in human glycogen-debranching enzymes has been linked to increased bladder cancer proliferation (6).

Pullulanases, named for their ability to hydrolyze pullulan, are crucial enzymes found in multiple bacteria, including *Gardnerella*. Although pullulan itself is absent in the vaginal environment, these enzymes also play a key role in glycogen degradation, producing smaller sugars that influence microbial community dynamics. Recent studies identified α-amylase (e.g., GH13_32 or GH13_11; EC 3.2.1.1) and α-amylase-pullulanase (e.g., GH13_14; EC 3.2.1.41) catalytic domains in *Gardnerella*, suggesting their role in glycogen metabolism in the vagina (4). While α-amylase cleaves α-1,4 glycosidic linkages, releasing maltose, maltotriose, and maltotetraose, α-amylase-pullulanase has dual activity, hydrolyzing both α-1,4 and α-1,6 bonds. The pullulanase domain specifically releases maltotriose from pullulan and maltose from glycogen, underscoring its importance in carbohydrate metabolism (2,7). Neopullulanase (EC 3.2.1.135) is a glycoside hydrolase that cleaves both α-1,4- and α-1,6-glycosidic bonds in pullulan, primarily producing panose, maltose, and glucose (8). While extensively characterized in industrially relevant bacteria such as *Parageobacillus thermoglucosidasius* (8), as far as we know, this enzyme has not been identified in *Gardnerella* species. Besides *Gardnerella*, other vaginal bacteria, including *Lactobacillus* species and *Bifidobacterium lacrimalis*, also possess pullulanase homologs, suggesting potential competition for glycogen or cross-feeding interactions within the microbiome (2,9). The presence of carbohydrate-binding modules (CBMs) in *Gardnerella*, such as CBM25, CBM41, and CBM48, further supports its ability to bind and transport sugars, facilitating glycogen metabolism (10). Moreover, the identification of six sugar transporter operons in *Gardnerella*, including MalEFG (specific to *G. leopoldii*) and MalXFGK (conserved across *Gardnerella* species), highlights its metabolic flexibility and adaptation to glycogen utilization (10). Our study aims to elucidate the role of glycogen-debranching enzymes and associated sugar transporters in *Gardnerella* by characterizing their diversity across strains isolated from vaginal swabs and urine samples of both healthy and symptomatic women. By analyzing enzymatic activity and uptake mechanisms, we seek to clarify the interplay between these enzymes, their predicted metabolic products, and the transport systems facilitating nutrient acquisition. Additionally, we aim to uncover how these metabolic pathways may contribute to *Gardnerella*'s adaptive strategies, enhancing its ability to colonize and persist within the vaginal and urinary tract environments.

Materials and methods

Bacterial collection and sampling background

This study included a previously published collection of 33 *Gardnerella* isolates, which were subjected to whole genome sequencing (WGS). The collection consists of *G. vaginalis* (n=12), *G. pickettii* (n=7), *G. leopoldii* (n=7), *G. swidsinskii* (n=6), and *G. greenwoodii* (n=1). These isolates were obtained from mid-stream urine samples and vaginal swabs of both healthy women and those with symptomatic overactive bladder (OAB) and are part of an ongoing investigation into the female urinary microbiome (11,12).

Identification of glycogen-degrading enzymes, CBMs-associated and transporter protein

To identify carbohydrate-active enzymes (CAZymes) and associated carbohydrate-binding modules (CBMs), the predicted *Gardnerella* proteomes annotated by RAST were analyzed using the dbCAN3 web server (13) (https://bcb.unl.edu/dbCAN2/). The analysis focused on Glycoside Hydrolase Family 13 (GH13) enzymes, which are known for hydrolyzing glycosidic bonds in carbohydrates. Specifically, the dbCAN3 output was filtered to retain only GH13 enzymes associated with signal peptides. This family is particularly relevant due to its diverse roles in carbohydrate metabolism. In addition, Geneious Prime® 2024.0.7 was employed to identify specific types of pullulanases within the GH13 family, and transporter proteins previously identified by Bhandari *et al.,* (10) were directly searched across all proteomes using reference amino acid sequence segments.

Detection of Amylase Activity

Amylase activity in *Gardnerella* crude extracts was assessed using the iodine test. Strains predicted to encode carbohydrate-active enzymes based on genomic analysis (*G. vaginalis* n=1, *G. leopoldii* n=3, *G. swidsinskii* n=1) were cultured in 10 mL of Brain Heart Infusion (BHI) broth (Frilabo, Portugal) supplemented with 1% glucose, 1% starch, 0.5% yeast extract, and 1% gelatin. Cultures were incubated at 37°C in 5% CO_2 for 48 hours. Following incubation, broth cultures were centrifuged, and the resulting *pellet* was washed with PBS before being sonicated at 80% amplitude duty cycle (8 minutes total; 10 sec ON/ 5 sec OFF cycles). The crude extract was then plated on starch agar and incubated for 24 hours at 37°C. The supernatant was used

as the negative control. After incubation, the plates were flooded with Lugol's iodine solution to visualize starch hydrolysis.

Results and Discussion

Diversity of glycogen-degrading enzymes: identification of neopullulanase in *Gardnerella*

Based on our analysis and current literature, we have classified the glycogen-degrading enzymes found in our *Gardnerella* collection as follows: **1)** *glycogen debranching enzymes*: GH13_11 + CBM48; **2)** *neopullulanase*: GH13_20 + CBM34; and **3)** *α-amylase domain*: GH13_32 + 2 or 4 CBM25, CBM41 or CBM48. In our collection we did not identify *α-amylase-pullulanase* domains (GH13_32 + 2 or 4 CBM25, CBM41 or CBM48 + GH13_14) (Table 1).

A key finding of this study is the identification of neopullulanases in 51.52% (16/33) of the analyzed genomes. This enzyme was detected exclusively in *G. vaginalis* (n=9) and *G. pickettii* (n=7) isolates, derived from both healthy women (HW) and women with overactive bladder (OAB). These strains possess the GH13 subfamily 20 domain (GH13_20) alongside CBM34, a configuration known to hydrolyze α-1,4 and α-1,6-glycosidic linkages in starch, releasing reducing sugars such as glucose, maltose, and maltotriose (7,14,15). These strains may also be capable of degrading glycosidic linkages from glycogen present in the vaginal environment. To our knowledge, this is the first study to identify neopullulanases in the *Gardnerella* genus, shedding new light on their metabolic capabilities.

Most isolates harbored more than one type of glycogen-degrading enzymes, with our analysis revealing 10 distinct profiles (Table 1). The most prevalent profile, observed in 13 isolates, comprised α-amylase and a glycogen-debranching enzyme with domains GH13_11 + CBM48 and GH13_9 + CBM48. This profile was detected across *G. leopoldii, G. swidsinskii*, and *G. vaginalis* species, although with variations in associated transporters. The second most common profile, found in 9 *G. vaginalis* isolates from both HW and OAB groups, as well as one *G. pickettii* isolate, included all enzyme types, α-amylase, glycogen-debranching enzyme (GH13_11 + CBM48), and neopullulanase. Again, variations in associated transporters were observed within this profile.

The α-amylase domain (GH13_32) showed considerable variation across species: the majority of isolates revealed GH13_32 with CBM25 and CBM41, but we also identified an expanded

configuration that included GH13_32 with CBM25, CBM41, and CBM48 present in all *G. swidsinskii* strains (n=6), two *G. leopoldii* strains (c18Ua_110 and c18Ua_112), one *G. vaginalis* strain (c15Ub_15) and one *G. pickettii* strain (c23Ub_4). The presence of the additional CBM48 suggests an expanded substrate range for these strains, potentially enhancing their ability to bind and degrade not only starch and β-glucans but also glycogen with greater efficiency. This likely reflects metabolic adaptations that allow these strains to thrive in diverse carbohydrate-rich environments, such as the vaginal ecosystem.

Distinct carbohydrate transporter profiles across species

Analysis of carbohydrate transporters revealed distinct patterns across *Gardnerella* species. The $MusEFGK_2I$ transporter, responsible for maltose and maltotriose uptake, was detected in *G. vaginalis* and *G. pickettii*. Conversely, the $MusE_2FGK_2I$ transporter, also involved in maltose and maltotriose uptake, was found in *G. swidsinskii* and *G. leopoldii*. The MalXFGK transporter (glucose uptake) was ubiquitous, present in all species examined (*G. vaginalis, G. pickettii, G. leopoldii, G. swidsinskii*, and *G. greenwoodii*). The MalEFG transporter (maltose and maltodextrin uptake) was identified in both *G. leopoldii* and *G. greenwoodii*, appearing to be present in all *G. leopoldii* strains except c18Ua_110 and c18Ua_112 (Table 1). Finally, RafEFGK (raffinose, panose, stachyose, melibiose, and isomaltose uptake) and TMSP transporters (trehalose, maltose, sucrose, and palatinose uptake) were exclusively identified in *G. vaginalis*. This distinct repertoire underscores the diversity of carbohydrate transport capabilities within the *Gardnerella* genus, suggesting that *G. vaginalis* may be uniquely adapted to utilize a wider range of complex carbohydrates in the vaginal environment.

Our analysis of carbohydrate transporter genes in *Gardnerella* revealed important differences compared to the findings of Bhandari *et al.,* (10). Specifically, we did not identify the hypothetical proteins associated with $MusEFGK_2I$ and $MusE_2FGK_2I$ in *G. vaginalis* and *G. swidsinskii*, respectively, within our *Gardnerella* collection. Additionally, all *G. swidsinskii* strains in our study lacked one of the two ABC transporter substrate-binding proteins (SBPs) reported by Bhandari *et al.,* (10). While amino acid sequence similarity between our and Bhandari's transporter operons was high (72.35-100%), we observed several key residue differences that could have functional implications. Future studies should prioritize functional characterization of these transporters, including substrate specificity and transport kinetics, to determine the impact of the observed amino acid variations and to resolve the taxonomic ambiguities. The amino acid sequences for the transporters discussed in this study are provided in Table S1.

Gardnerella mechanisms to efficiently metabolize glycogen

The interplay between *Gardnerella* enzymes and transporters highlights diverse carbohydrate utilization strategies, potentially shaping interspecies competition and community dynamics. *G. vaginalis* exhibits diverse enzymatic tools across strains (Table 1) yet consistently possesses the RafEFGK (raffinose uptake) and TMSP (maltose/fructose uptake) transporters. This combination suggests a “generalist” strategy for utilizing diverse, complex sugars to thrive in carbohydrate-rich environments, providing a selective advantage not found in other *Gardnerella* species. In contrast, *G. leopoldii*, *G. swidsinskii*, and *G. greenwoodii* share the GH13_11 + CBM48 (α-1,6-glycosidic linkages) and GH13_9 + CBM48 (α-1,4-glycosidic linkages) enzymes for glycogen breakdown, and the MalXFGK transporter (glucose uptake). The MalEFG (maltose/maltodextrin uptake) transporter is shared by *G. leopoldii* and *G. greenwoodii*, while *G. leopoldii* and *G. swidsinskii* share $MusE_2FGK_2I$ (also maltose/maltodextrin uptake). The shared MalXFGK transporter across all species highlights a common mechanism for glucose acquisition, providing a metabolic foundation within the urogenital microbiome, where glycogen is abundant. The ability of *Gardnerella* species to degrade and utilize shared resources like glycogen may drive “scramble” competition (16), where efficient nutrient utilization allows certain strains or species to outcompete others. Conversely, the resulting breakdown products may also be shared with other bacterial species, such as *Lactobacillus*, leading to community-level benefits and potentially altering the overall composition of the urogenital microbiome (16).

Amylase activity

To assess amylase activity in *Gardnerella* isolates, we employed Gram's iodine to detect starch degradation. Our study revealed that *Gardnerella* strains exhibited amylase activity, regardless of their specific enzyme and transporter profiles. Starch hydrolysis was evidenced by clear zones surrounding the crude extracts in the starch-iodine assay. This activity was observed in isolates from both HW and those with OAB syndrome, aligning with predictions made through genomic analysis (Figure 1).

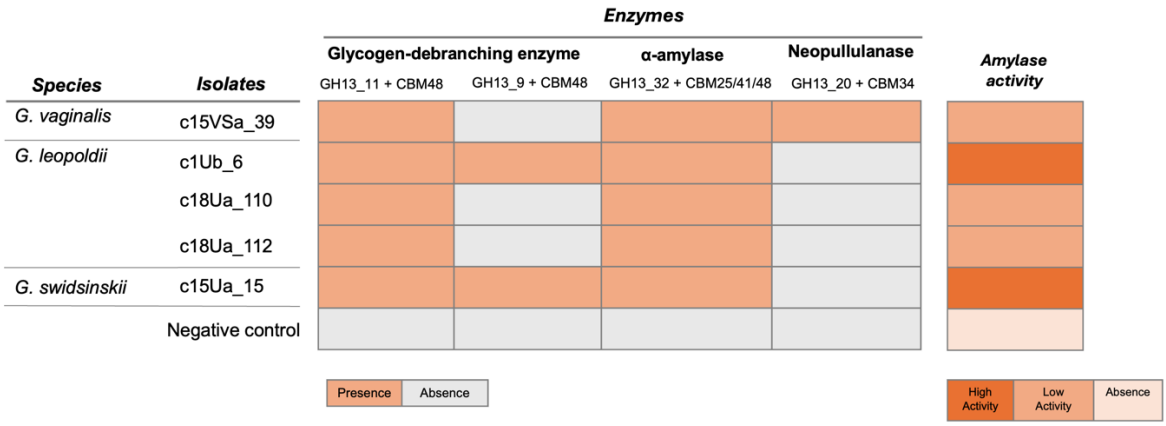


Figure 1: The presence of different enzymes was assessed through *in silico* analysis, while the corresponding amylase activity in selected *Gardnerella* isolates was assessed using the iodine test.

Table 1: Glycogen metabolism enzymes and transporter profiles present in *Gardnerella* strains.

	Enzymatic profile	Transporters	Species (n)	Health Status
1	· α-amylase · Glycogen debranching enzyme (GH13_11 + CBM48) · Neopullulanase	MusEFGK$_2$I MalXFGK RafEFGK TMSP	*G. vaginalis* (9)	HW, OAB
2	· α-amylase · Glycogen debranching enzyme (GH13_11 + CBM48 and GH13_9 + CBM48)	MalXFGK MusE$_2$FGK$_2$I	*G. leopoldii* (2) *G. swidsinskii* (6)	OAB HW
3	· α-amylase · Glycogen debranching enzyme (GH13_11 + CBM48 and GH13_9 + CBM48)	MalEFG MalXFGK MusE$_2$FGK$_2$I	*G. leopoldii* (3) *G. vaginalis* (2)	HW HW
4	· Glycogen debranching enzyme (GH13_11 + CBM48) · Neopullulanase	MusEFGK$_2$I MalXFGK	*G. pickettii* (4)	HW, OAB

5	· α-amylase · Glycogen debranching enzyme (GH13_11 + CBM48)	MusEFGK$_2$I MalXFGK RafEFGK TMSP	*G. vaginalis* (2)	HW
6	· α-amylase · Glycogen debranching enzyme (GH13_11 + CBM48) · Neopullulanase	MusEFGK$_2$I MalXFGK	*G. pickettii* (1)	HW
7	· α-amylase · Neopullulanase	MusEFGK$_2$I MalXFGK RafEFGK TMSP	*G. vaginalis* (1)	HW
8	· α-amylase · Neopullulanase	MusEFGK$_2$I MalXFGK	*G. pickettii* (1)	HW

9	· Glycogen debranching enzyme (GH13_11 + CBM48 and GH13_9 + CBM48)	MalEFG MalXFGK	*G. greenwoodii* (1)	HW
10	· Neopullulanase	MusEFGK_2I MalXFGK	*G. pickettii* (1)	HW

Abbreviations: HW - Healthy woman, OAB - Woman diagnosed with overactive bladder.

Conclusion

The correlation between glycogen-degrading enzymes and carbohydrate transporters in *Gardnerella* species highlights their metabolic versatility and adaptation to diverse carbohydrate-rich environments. While some strains exhibit a preference for degrading starch and β-glucans, others (particularly those containing the GH13_9 + CBM48 domain) may employ known or novel enzymatic mechanisms to efficiently degrade glycogen. Moreover, the discovery of distinct glycogen debranching enzymes, including the newly identified neopullulanase, within the genus, alongside various CBMs and carbohydrate transporters, highlights the capacity of *Gardnerella* to metabolize a broad spectrum of sugars. These findings suggest that *Gardnerella* strongly relies on carbohydrate metabolism for survival and persistence within the urogenital microbiome.

Future experimental studies are essential to determine the substrate specificity of CBMs and their role in sugar utilization that will provide deeper insights into the functional roles of these enzymes and transporters, advancing our understanding of *Gardnerella*'s carbohydrate metabolic pathways and their contributions to host-associated ecosystems.

Acknowledgements

We would like to thank Dr. Janet E. Hill and Dr. Pashupati Bhandari, from the Department of Veterinary Microbiology, Western College of Veterinary Medicine, University of Saskatchewan, Canada, for sharing their list of primers and sequences used in their study (10), which enabled us to perform a meaningful comparison with our results.

References

1. Dittmer KE, Pradhan P, Tompkins QC, Brittingham A, Wilson WA. Cloning and characterization of glycogen branching and debranching enzymes from the parasitic protist *Trichomonas vaginalis*. Biochimie. 2021 Jul 1;186:59–72.

2. Nunn KL, Clair GC, Adkins JN, Engbrecht K, Fillmore T, Forney LJ. Amylases in the Human Vagina. mSphere. 2020 Dec 9;5(6):e00943-20.

3. Bonafonte MA, Solano C, Sesma B, Alvarez M, Montuenga L, García-Ros D, *et al.* The relationship between glycogen synthesis, biofilm formation and virulence in *Salmonella enteritidis*. FEMS Microbiol Lett. 2000 Oct 1;191(1):31–6.

4. Bhandari P, Tingley J, Abbott DW, Hill JE. Glycogen-Degrading Activities of Catalytic Domains of a-Amylase and a-Amylase-Pullulanase Enzymes Conserved in *Gardnerella* spp. from the Vaginal Microbiome. J Bacteriol. 2023 Feb 1;205(2).

5. Nasioudis D, Beghini J, Bongiovanni AM, Giraldo PC, Linhares IM, Witkin SS. α-Amylase in Vaginal Fluid: Association with Conditions Favorable to Dominance of *Lactobacillus*. Reproductive Sciences. 2015 Nov 1;22(11):1393–8.

6. Lew CR, Guin S, Theodorescu D. Targeting glycogen metabolism in bladder cancer. Nat Rev Urol. 2015 12(7):383–91.

7. Kahar UM, Ab Latif N, Amran SI, Liew KJ, Goh KM. A Bibliometric Analysis and Review of Pullulan-Degrading Enzymes—Past and Current Trends. Catalysts 2022 12(2), 143.

8. Ibenegbu CC, Leak DJ. Simultaneous saccharification and ethanologenic fermentation (SSF) of waste bread by an amylolytic *Parageobacillus thermoglucosidasius* strain TM333. Microb Cell Fact. 2022 Dec 1;21(1).

9. Jenkins DJ, Woolston BM, Hood-Pishchany MI, Pelayo P, Konopaski AN, Quinn Peters M, *et al*. Bacterial amylases enable glycogen degradation by the vaginal microbiome. Nat Microbiol. 2023 Sep 1;8(9):1641–52.

10. Bhandari P, Hill JE. Transport and Utilization of Glycogen Breakdown Products by *Gardnerella* spp. from the Human Vaginal Microbiome. Microbiol Spectr. 2023 Apr 13;11(2).

11. Perovic SU, Ksiezarek M, Rocha J, Cappelli EA, Sousa M, Ribeiro TG, *et al.* Urinary Microbiome of Reproductive-Age Asymptomatic European Women. Microbiol Spectr. 2022 Dec 21;10(6).

12. Perovic SU, Ksiezarek M, Rocha J, Vale L, Silva C, Dinis P, *et al.* Time To Change Microbiological Approach To Overactive Bladder. EAU19 - 34th Annual EAU Congress. 2019. 18(1):e103, ISSN 1569-9056.

13. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. DbCAN3: Automated carbohydrate-Active enzyme and substrate annotation. Nucleic Acids Res. 2023 Jul 5;51(W1):W115–21.

14. Kamasaka H, Sugimoto K, Takata H, Nishimura T, Kuriki T. *Bacillus stearothermophilus* neopullulanase selective hydrolysis of amylose to maltose in the presence of amylopectin. Appl Environ Microbiol. 2002 68(4):1658–64.

15. Domań-Pytka M, Bardowski J. Pullulan degrading enzymes of bacterial origin. Crit Rev Microbiol. 2004 30(2):107-21.

16. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol. 2010 8(1):15–25.

Supplementary Material

Table S1: Amino acid sequences of the various transporters discussed in this study retrieved from the proteomes of our collection.

Species	Operon	Protein	Amino acid sequence
G. vaginalis	**MusEFGK$_2$I**	Histidine phosphatase family protein	>MLNSIDVDVNGGSFSGDSYEIQDVPATCGGAPEGRLVLLRHGQTVW SESGQHTGRTDIPLTQIGCSQAVYAGERIRQAFPKGFDADCQFVSPLV RARQTAQLAGFTYCTPLEYAAEWDYGCAEGRTRKDVSALSGVESWD VWRDGPKALPEFMSDSCEESLPSGETVKVVRCNGESLQDVSDRTQKII DSVLPKLKSGKDVIIVAHAHVLRILAMRWLGVDASQARLLRLDTAHYSA LGTYKGDRVIVRWNC
		ABC transporter substrate-binding protein	>MMNRTIKSTIALVACSAMAFGALAACGSSSSDSSDKGKVYFLNFKPE SADQWKEVAKEYTKKTGIPVKVQNAASGTYEQSLKSEISKGDEAPTLF QVNGPVGYASWKDYTADLKDTELYKQLQNKDIALKDGDKVVGVPYAM ETYGLIYNKSLLKKYTELKDAKVKSAEEINSFDKLKAVADDIQARKSELG VDGAFTSAGFDSSSDWRFKTHLANIPLYYEFKADKVTKQPAKIKGTFLD GYKKIFDLYITDSTTPATQLSSKTGEDANSEFSLGKAVFYQNGTWAWT DLQKAGMKAEDLGMLPIYIGAKGEEKQGLATGSENYWCINDKTSDANK KATKDFLKWVLTSDYGKEAMSKKMGFTTPFKSFADMKSENPLTAAAV EDAQSGKTPVSWNFTMMPSDNWKNDLGSALLEYAQGTGKWDKVKSA FVDNWAKEVAQTHEDQD
		LacI family transcriptional regulator	>MPANIQDVANEAHVSVSTVSRSFTRPNLVSAITRDKVLAIAEKLNFSIS RSAAALKSGRTLRVALLMSDSIRLWFSASIMQGLNQVLHPAGYDLSIFQ ISSSQERAEFFDMLPTRRNADAVIVCSFDVNKSEVASLKATGVPVLGIN CLYPMDCDFDATINIDDNQGARLMARHLIGLGHKNIAYIRTNRDVSLHF SVLQRYHSFIDECKINNITPTEIVAPEGLDRISSIVSSLLSNSSMPTAIACQ EDGIAIPLMFQLSRSGYSTPGDLSIIGFDDSFYARETGLTTIRQDPVSMA VDAAKMTLSLINGDYLDESRRYITIPAQLIVRSSTSQLDVA

		Sugar ABC transporter permease	>MINMSGKAIRKWWALFALPTFAAFVIGFLVPFVMGVYLSFTKFHTVTD GKFVGIRNYRKALRDPHFIHSLIFTTAFTVVTTVVINVIAFAIAYMLTKKFK GSNLFRTVFFMPNLIGGIVLGYVWQVVLNGILGYFGRAITYSATYGFWG MVLLICWQQIGYMMIIYIAGLQALPGDVIEAAAVDGATPMQTLFKITIPLM MPSITVCTFLTVTNGFKLFDQNLALTNGAPSRKSEMLALNIFNTFYGRIN GEGVGQAKAVIFFIIVAVIALIQNRLTTSKEVAA
		Sugar ABC transporter permease/ Carbohydrate ABC transporter permease	>MNEKTKHPVLWTTLFAVVSLLWIFPIALVFMNSFKSKVYIASEPFSLTP KSFIGFDNYVLGINRTNLLMSFWWTIVVTIGAVVLILLCTSMCAWWIVRV NNWAARLLYVLFLFNMIVPFQMVMFTLSNITTKLGLNTPWGLWIIYLGF GAGLAVFIFTGVVKGIPQELEESAMIDGASVPRIFFQIIVPIMRPAIISVSIL QTMWIWNDFLLPYLTLDLNKYKTVSIAVQYLKGGYGSVEMGAMMGCL VLAILPIIIFYLICQKYIIKGVMAGAVKG
	MalXFGK	Type I pullulanase (GH13_14)	>MTCTSSTQWPVYSGQLGCSLNEKSTTFTVWAPTASCVTLRLFETGN NSEPFKTRNLKPLRDGAWRTSILQRLDGVYYDFLVTFADGKVNRTSDP WGVASGVNGERSMVVSEERVSPADWHKDKSPEFVRHATVVWETHVE DFSSNPNGGFKPEHRGQYLAFTDSHTTLGGDEHFPTGIDYLKKLGVTH VQLMPIYDFATVDEVAISSARKNGESVDSLYNWGYDPHAYNVPEGAYS SNPYDGTARIRECKAMIGALHKNGIRVVMDVVYNHMFEAASNVFESVA PGYFCRRKADGSFANGSGCGNEMASDHPMFRRFMIESILHWARCYH VDGFRFDLMGLTDAETLNMIRAELDKLDGGKDILMYGEPWGADAAAV GDDCPMGDKDGRGFLDSRIGWFSDESRDTLKGNVMDHHDRGYVNG NEFWTADPVRNAFNGWRGMPCAGKSVGQIVQYVSAHDDLTLWDKLCI SMRDNASSCDFDASDPNSVADIFSANILSAGLILCSAGMPFMLGGEEF GRTKHGCDNSYNRGALMNQLDWSRIKRFQPLVEWYSALISLRAKNAE FYDAQHTRLTASDDAVLAEQIGSDLIVCVNPMNDASATIDLPAKSANPW RCVLDSTEYFAKWTGHDSGSCVSMLCDDNGSQCAVHVPPRTFVVLS RKRA >MTFTSSTQWPVYSGQLGCSLNEKSTTFTVWAPTASCVTLRLFETGN NSEPFKTRNLKPLRDGAWRTSILQRLDGIYYDFLVTFADGKVNRTSDP WGVASGVNGERSMVVSEERVSPADWHKDKSPEFVRHATVVWETHVE DFSSNPNGGFKPEHSGQYLAFTDSHTTLDSDNNFPTGIDYLKKLGVTH VQLMPIYDFATVDEVAISSARKNGESVDSLYNWGYDPHAYNVPEGAYS

			SNPYDGTARIRECKAMIGALHKNGMRVVMDVVYNHMFEAASNVFESV APGYFCRRKSDGSFANGSGCGNEMASDHPMFRRFMIESILHWARCY HVDGFRFDLMGLTDAETLNMIRAELDKLDGGKDILMYGEPWGADAAA VPDDCPMGDKDGRGFLDSRIGWFSDESRDTLKGNVMDHHDRGYVNG NEFWTADPVRNAFNGWRGMPCAGKSVGQIVQYVSAHDDLTLWDKLCI SMRDNASSCDFDASDPNSVADIFSANILSAGLILCSAGMPFMLGGEEF GRTKHGCDNSYNRGVLMNQLDWSRAERFQPLVEWYSALIKLRAKNAE FYDAQRTHLTASDNTVLAEQIGSDLIVCVNPMNDASATIDLPAKSASPW RCVLDSTEYFAKWTGHDSGSCVSMLCDDNGSQRAVHVPPRTFVVLS RKRA
		Sugar ABC transporter permease	>MSNKKDLMQGEQTVHSVRSVPFWRDQRIRRIFGDIVAHLFLIVMAVIW LAPIVWVFLESFNQNTAPYQTTFFPTKYTFNNYIQLFTEREVLDFPRMF LNTLIVSIFVCIISLFFVLVVAFCMSRLRFRGRKTFMNIALILGMFPGIMAV IAIYFILKAFGLTNGSVVLISLIIVYSAGSGMGFYVMKGYMDTIPKSLDEA AYLDGCTTWQVFWKIILPICKPMIVFQAIVSFLGPWLDFVLVKTIARTQD NYTVALGLYQMLQREYINHWFARFSAGAVCVSIPIAILFIVMQRFYEES MSGSVKG
		Sugar ABC transporter permease	>MTGTTSTKRGKRKKAPAADEYIAPSQYSLQEAAKHADVLSWVSIIIFG FGNFVRKQYVKGLVFILCEAGVLAFLATTGAEYLPKFGTLGTKKQGRVK INGFWHYVAGDNSVEILLYGVMSIAAIVVLIYLATLSLRSAYKAQAFVAR GEKPITFVKDVESLLDRHAHVGFMTLPTAGIVLFTVLPLVFMISMAFTNY DRDHTLLFHWVGLNNFFKILGNTTGTINAGIFFDVLIWTLVWAFFATFLN FFFGMFMAMIINRRTTRLKSLWRICFSLTIAVPQFVSLLVMHSMLQRDG AINRLLISSGWIHSPLPFFQDATWARVTVILINLWVGIPYTIMQITGILQNV PAELYEAAELDGASWWQRFMNVTMPYIFFVMTPYLITTFTGNVNNFNV IYLLSGGEPTPIGSSAGKTDLLITWLYKLTVDKDNYSMGAVIGIMTFIVLAI VALITYRNSGSYKNEEGFR
		Extracellular solute-binding protein/ ABC Transporter substrate-binding protein	>MTNMKKVLGAGLAIATMFGFAACGSSTNAGSSNKANPVKLTVWSSS EDQKAGGWIQAMEKAFKKDHPEVTFKNAVVAPPDAAKTVKQDAKAAA DVYLFPNDQLGDLVKANAIGELSDDAAKQVKEDNDDTIIQSVTAQDGKL YGVPYMGNTWFMYYDKSKFSDEDVKSLDKMLEKGKVAFDISNAWYLP GFYLDGNMTLFGEKSNDGKAGIKIGDKAAEITKYVAKLIQNKNFVMDGD GAGLAGIKNHTVDAYFSGSWDAGDVKKALGDNYAAAQLPKFKSEDGE

			HQMKSFAGSKAVAYNPNSKAPEMAAKFAAFLGSKESQEQMYKLHGDI PVAKSLADLVKDNPAAVAQMNTIAKTSVLQPTVPEMGAFWDPMKTFG TALANKEVNDGNAAAKIADFQKGFEEALKK
		α-amylase family glycosyl hydrolase (GH13_23)	>MQNEHAQWLRDAVFYEIYPQSFYDSNGDGIGDIPGIIEKLDYVKSLGC NAIWLNPCYDSPFKDAGYDVRDYKKVAARYGTNEDLQRLFSEAHNRG MHILLDLVPGHTSEEHEWFKQSSRAQENEFSKRYIWTDSAFSGYSMPF IGGESERDATYILNFFKCQPALNYGFAHRDRAWQSAPDSEEAAQTRAA MVDVMRFWLALGADGFRVDMADSLVKNDDNNGEGSLGKDNTIRAWQ EMLGAVKAEYPQAAFVSEWGRPRQALAAGFDMDFYLNWRWDGQPN GYARLARDVDNALDGNPERDHSYFNARGGGSICPFLDDYCPQYEATC NEGYFSFITCNHDTPRLAPRLSERERRIAFGMILTMPGVPFVYYGDEIG MKYRDIPTKEGGYARTGTRTPMQWDDSKNFGFSTAAKSKLYLPVEAR ADGRTCANSRKQAVQSGEIPTVLGQEQDGDSLLNWVRALIALRHECA ALHADATWRVLHAPVDGRSFAYERCAAESNVSDGESSTDRCIVVMNP GLNEETIALSALAGLTPEGILQPRISVGAVACDSSALKLGAQSFALFVL
	RafEFGK	Carbohydrate ABC transporter permease	>MRCGFRNVKRTAVALVTLATVLLPLGACGNGNSAGKVSISFYSYFKK NQIGNVISAFQKAHPNIKIDAQYGQNSTQYIQTLQTRLAGGTPPTIFNLT MDNRTDIMQSGQALDISGSKFFSGIDESNFKLFQHNGKTYGMPVSAW VGAMFYNKDILKQAGYTEFPKTWDDFITMGKKINDSGKVAFLEDFNTQ PSGTFEALLASKYASDKNDKQDEVIWNGSSTFSKEWTPALSEWAKAID AKVIPHKSIGLTADQVKQEFVTGNLAVMRSGPWDLADVKASGINFGVA PMPSYKNGEQWINGGPDQGFAIAAKASKAQQEAAKTFLAYLNSKDGL KTFTTNAGTLSLSDKYRADPPAELAGVIKDYFQNNKFYWVNFSKSPSA MMTEMASRQQELVQGKITPKEFTKLLDDKWNSMK
		Sugar ABC transporter permease	>MRCSPNNNSNSEQIRHKFSVSSAAFKDNLTSFWFLLPIVIIFIVLTLIPL MQSIFYSFTDFNGYDLNFHFIGLQKYQQVFSDTSLLSALIFTLMYSFIVTI VVTCLAIPLAVVLNQAMIGKSFLRSLFFFLGVPAQAVIGLIWQYIFSPLDS GVANQVLHTFGLPSIQWLSQDNWARFCVMFVAVWMQVGWHATLYIA WLQAIPADLYEQARVDGANGLQRFVHITLPQLIPGMVVSTFLLMTTALKI YDLPYTLTAGGPGHSTNTITQAIIMRGMGQSDVGLASALSTLFTIACVIVI AIQMKVSSTVARRFQ

		Extracellular solute-binding protein	>MRCGFRNVKRTAVALVTLATVLLPLGACGNGNSAGKVSISFYSYFKK NQIGNVISAFQKAHPNIKIDAQYGQNSTQYIQTLQTRLAGGTPPTIFNLT MDNRTDIMQSGQALDISGSKFFSGIDESNFKLFQHNGKTYGMPVSAW VGAMFYNKDILKQAGYTEFPKTWDDFITMGKKINDSGKVAFLEDFNTQ PSGTFEALLASKYASDKNDKQDEVIWNGSSTFSKEWTPALSEWAKAID AKVIPHKSIGLTADQVKQEFVTGNLAVMRSGPWDLADVKASGINFGVA PMPSYKNGEQWINGGPDQGFAIAAKASKAQQEAAKTFLAYLNSKDGL KTFTTNAGTLSLSDKYRADPPAELAGVIKDYFQNNKFYWVNFSKSPSA MMTEMASRQQELVQGKITPKEFTKLLDDKWNSMK
		Glycoside hydrolase family 31 protein (GH13_44)	>MNVEISRTTDVLNTAETTKTNVVSNMASNIADKFVTTIRLLEGERWWG GTVIDGIYEPFGKQVFTRDLSEWTQKLREVPEASNQSSPLLISTCGRYV WCKEPFKFTFADRTLSIEHSSSLTFTKVGDKLSDAYKYACNKYFPPSG RSIPADLIDKPQYNTWIEMPYAPTQNKVENYARQILELGMPSGIIMIDDK WSPDYGDWTFDISRFPKPRKMIDSLHEQGFRVMLWLVPFISPDSENFR YLEKENLLLRSVNGETAIRRWWNGLSAVLDLSHPKAISWLEEKLDALR QIGVDGFKFDGGDFYECHNDDISYESMTSAEFCERWAKFGLRYSYNE FRACWKMGGQPLAQRLRDKPQSWGVEGLQSLIPEIIAQGLIGHAFTCP DMIGGGEIESMQNANSIDQDFFIRYAQIAALCPMMQFSTLPHRVLDKEH MKALICAIRTREAHLPTIQKFALNAANTGEPIVRPMSYHYDHCEDIIDQFL LGDRIIVAPALHKYQKQRSVYIPDGSWESDDKIIFDGPTTIIVDTPLDRIPI FTKKM
	TMSP	ROK family transcriptional regulator	>MKDSDDSRNRLISFTPEDIRKKNRCNVLNLLYSNYILSRSDVAKLTGIS KSSASAVVESLISDGLVKEGKTKSATCPGKPAVLLHLCENSRQIIVVDIA GENTIVGAITNLVGNVIECIQTPIRLDDKLSVDDIVSLIVDLRKKTTAPLLGI GISSPGVIDDEGKILSAPHLGWKNVTLADDLRKIFNVPVRVDNNANAAA LGEKQLANGSSNLIYIQIARGVGAGTLISNKIVDGGKFTAGEIGHVVVDE DGLLCRCGKRGCLETLISIPRLQQRIADAPSEKDRIITQAGTSLGKILAIPI AMNGIEDVVVNGDSSIINDTFIANMQFAINSRIDSKLFGAVRVRFPALEQ DASLLGESVAVMRSELC
		Carbohydrate ABC transporter permease	>MKIVTLLDLLRRRVTLNKIINTLLAFVFIIVWIFPVYWMCVTAFSPRNEVL QSKPTLLPSHWSFEHMHDAFSSDSMLSYACNSLLVTVLSVVIAVIVAFL ACAALTMHAFAGRKLIMAAVIVVQMIPATAILIPQFIVFNRFGMLDTYVGL VLAYVTAVLPVAIWNLRGFFLAMPMSIFESARVEGASEWQILWRITFPL

			AAPGIIATSVFACIHAWNDYLIAYTFMKTQEKYTLPVWLASFATPSGVDV GAQMAASLIFSVPVIVFFLIIQHGLTFGSMQGAIK
		Sugar ABC transporter permease	>MRILKVSKNYSSRCASKSASNSASKIVPALMLLPACILLIVLVILPLLMLV VASLSNFNARSLFTGEFEFVGIQQYLALFADKGFYYSLGITIAFAAFLVIT SVLIGMWVAQTMFKQSAVVRVIMSIVLVAAWAMPNVAAALVFKWMFQ PGYGVVGWLLTQLKIFGNIRNLSWTSNTNLAFCCIGLIIVWQAVPYIAITL YAAQKQIAPEYAEAAAVDGASKTRIYWQITVPMLAPQLTAITILSCIWDF NVFNQIWLLTQGGPRESTATVGVFTYKKAFINFSIGQGSAISVVTTLILFF ATVYYIRRLLKNGDEL
		Extracellular solute-binding protein	>MRRIYGKCTAYAVITLAMAIMLSGCSIGASTSSISSQSSSKNGITVWVM QDDFTKETLNAINKQFEAQTGVRVHVQVQPWDGISTKLTTALSTSRPP DVVDIGNTKIADFAASGGLLDITEHEKMLSQGKRWLKGLREPAIYKDKL YAVPAFAASRVVIYNKRIWSSAGITNQPATFTQLISALNTIKQAHKNEAD FSAMYLPGQNWFAGMQFVWDCGGSLASFNGMQWHSSAISKKTLQGL KKWADFQNTFSTEASRSAEINNPDQAQILAQQKASAVLTNSAAVTKAIS INPSLSRKDFATFPMPSVNGKGIQPSVLAGSDWAIPSRSSNVALALKWI RIAASPLIQRKWIVEHDGWLPNSMDLLDEYLHSRDLDEVQKGFFETAR VSQSTPAAPGWSTIEADGSLKELFASSAMSAKSVLQAARIFHDHAQEIF SHNQVGGDL
G. pickettii	**MusEFGK$_2$I**	Histidine phosphatase family protein	>MHQAQAKNGGAKAALDMLNGLNNIDIKVNGDLCTTDATDTTDSYAIQ DVPATCSGAPEGRLVLLRHGQTAWSESGQHTGRTDIPLTQVGRNQAI SAGARIRQAFPNGFSEDCQFVSPLIRARLTAQLAGFSHCTPLEYAAEW DYGCAEGRTRKDVSALSGVESWDVWKDGPKALPDFMSDSCQEILPS GETVKVVRSNGESLQEVSNRTQKIIDSVFPKLKSGKDVIIVAHAHILRILA TRWLGVDAQQARLLRLDTAHYSALGTYKGDRVIVRWNC >MLNGLNNIDIKVNGDLCTTDATDTTDSYAIQDVPATCSGAPEGRLVLL RHGQTAWSESGQHTGRTDIPLTQVGRNQAISAGARIRQAFPNGFSED CQFVSPLIRARLTAQLAGFSHCTPLEYAAEWDYGCAEGRTRKDVSALS GVESWDVWKDGPKALPDFMSDSCQEILPSGETVKVVRSNGESLQEVS NRTQKIIDSVLPKLKSGKDVIIVAHAHILRILATRWLGVDAQQARLLRLDT

			AHYSALGTYKGDRVIVRWNC
		ABC transporter substrate-binding protein	>MMNRTIKSTIALVACSAMAFGALAACGSSSSDSSDKGKVYFLNFKPE SADQWKEVAKEYTKKTGIPVKVQNAASGTYEQSLKSEISKGNEAPTLF QVNGPVGYASWKDYTADLKDTELYKQLQNKDIALKDGDKVVGVPYAM ETYGLIYNKALLKKYTELKDAKVKSAEEINSFDKLKAVADDIQARKGDLG VDGAFTSAGFDSSSDWRFKTHLANIPLYYEFKADKVTKQPAKIKGTFLD GYKKIFDLYITDSTTPATQLSSKTGEDANSEFSLGKAVFYQNGTWAWT DLQKAGMKAEDLGMLPIYIGAKGEEKQGLATGSENYWCINEKTSDANK KATKDFLKWVLTSDYGKEAMSKKMGFTTPFKSFADMKSENPLTAAAV EDAQSGKTPVSWNFTMMPSDNWKNDLGSALLEYAQGTGKWDKVKSA FVDNWAKEVAQAHEDQD
		LacI family transcriptional regulator	>MPANIQDVADQAHVSVSTVSRSFTRPNLVSAATRNKVLAIAEKLNFSI SRSAAALKSGRTLRVALLLSDSIRLWFSASIIQGLNQVLHPAGYDLSIFQI SSSQERAEFFSMLPTRRNADAVIVCSFDVNKSEVASLKATGVPVLGINC LNPMDCDFDATINIDDDQGARLMARHLIGLGHKNIAYIRTNRDISLHFSV LQRYHSFIDECKINGITPTEIVAPEGLDRISSIVSSLLSNVNVPTAIACQED GIAIPLMFQLSRSGYSTPGDVSIIGFDDSFYARETGLTTIRQDPAEMAVD AAKITLSLINGDYLSEECKYITIPAQLIVRSSTSQLENN
		Sugar ABC transporter permease	>MINMSGKAIRKWWALFALPTFAAFVIGFLVPFVMGVYLSFTKFHTVTD GKFVGIRNYRKALRDPHFLHSLVFTTAFTVVTTIVINVIAFAIAYMLTKKF KGSNLFRTVFFMPNLIGGIVLGYVWQVVLNGILGYFGRAITYSATYGFW GMVLLICWQQIGYMMIIYIAGLQALPGDVIEAAAVDGASPTQTLFKITIPL MMPSITVCTFLTVTNGFKLFDQNLALTNGAPSRKSEMLALNIFNTFYGR LNGEGVGQAKAVIFFIIVAVIALIQNRLTTSKEVAA
		Sugar ABC transporter permease/ Carbohydrate ABC transporter permease	>MNEKTKHPALWTTLFAVVSLLWIFPIALVFMNSFKSKVYIASEPFSLTP KSFIGFDNYVLGINRTNLLMSFWWTIVVTIGAVALILLCTSMCAWWIVRV NNWAAKLLYVLFLFNMIVPFQMVMFTLSNITTKLGLNTPWGLWIIYLGF GAGLAVFIFTGVVKGIPQELEESAMIDGASVPRIFFQIIVPIMRPAIISVSIL QTMWIWNDFLLPYLTLDLNKYKTVSIAVQYLKGGYGSVEMGAMMGCL VMAIVPIIVFYLICQKYIIKGVMAGAVKG

	MalXFGK	Type I pullulanase (GH13_14)	>MTFPPSTQWPVYSGQLGFSLNERSTTFTLWAPTASCVTLRLFETGNN SEPFKTRDLKPLRDGAWRTSILQRLDGVYYDFLVTFADGKVNRTSDPW GIASGVNGERSLVIDEDRVAPAGWHKDKSPEFVRHATVVWEMHVEDF SSNANGGFKPEHRGQYLAFTDSHTTLDSDSNFPTGIDYLKKLGVTHVQ LMPIYDFATVDEVAISSARKDGKSVDSLYNWGYDPHAYNVPEGAYSSN PYDGTARIRECKAMIGALHKNGMRVVMDVVYNHMFEAASNAFESVAP GYFCRRKEDGGFANGSGCGNEMSSNHPMFRRFMIESILHWARCYHV DGFRFDLMGLTDAQTLNMIRAELDKLDGGKDILMYGEPWGADAAAVG DDCPMGDKAGRDFLDSRIGWFSDESRDTIKGNVMDHRDRGYVNSNE FWTADPVRNAFNGWRGMPCAGKSVGQIVQYVSAHDDLTLWDKLCIS MRDNASNCDFDASDPNSVADIFSANILSAGIILCSAGMPFMLGGEEFGR TKHGCDNSYNRGVAMNQLDWSRAKRFQPLVEWYSALIKLRAKNAEFY DAQHTRLNASDNSVLAEQIESDLIVCANPMKDASATVDLPELCASRKK WRCVLDSSEYFDRWIGSASGSRASMLCDDNGDPRAVHIPPRTFVVLS NKRA
		Sugar ABC transporter permease	>MSNKKDLMQSEQTVHSTRSVPFWRDQRVRRIIGDIITHIFLIIMSIIWLA PIVWVFLESFNQNTAPYQTTFFPTKYTFNNYIQLFTEREVLDFPRMFLN TLIVSIFVCIISLFFVLVVAFCMSRLRFRGRKTFMNIALILGMFPGIMAVIAI YFILKAFGLTNGSVVLISLIIVYSAGSGMGFYVMKGYMDTIPKSLDEAAY LDGCTTWQVFYKIILPICKPMIVFQAIVSFLGPWLDFVLVKTIARTQDNYT VALGLYQMLQREYINHWFARFSAGAVCVSIPIAILFIVMQRFYEESMSG SVKG
		Sugar ABC transporter permease	>MTGTTSTKRGKRKKAPSADEFVAPSQYSLQEAAKHADALSWVSIIIFG FGNFVRKQYVKGLVFLLCEAGVLAFLATTGAEYLPKFGTLGTRKQGRV KINGFWHYVAGDNSVEILLYGVMSIAAIIVLVYLATLALRSAYKAQSFVK RGEKPITFVKDVESLLDRHAHVGFMTLPTAGIVLFTVLPLAFMISMAFTN YDRDHTLLFHWVGLDNFVKILGNTTGTINAGIFFDVLIWTLVWAFFATFL NFFFGMFMAMIINRRTTRLKSLWRICFSLTIAVPQFVSLLVMHSMLQRD GAINRLLISSGWIHSPLPFFQDATWARVTVILINLWVGIPYTIMQITGILQN VPAELYEAAELDGASWWQRFMNVTMPYIFFVMTPYLITTFTGNVNNFN VIYLLSGGEPTPIGSSAGKTDLLITWLYKLTVDKDNYSMGAVIGIMTFIVL AIVALITYRNSGSYKNEEGFR

		Extracellular solute-binding protein/ ABC Transporter substrate-binding protein	>MTNMKKVLGAGLAIATMFGFAACGSSTNAGSSSKSNPVKLTVWSSS EDQKAGGWIQAMEKAFQKDHPEVTFKNAVVAPPDAAKTVKQDAKAAA DVYLFPNDQLGDLVKANAIGELSDDAAKQVKGDNDETIIQSVTAQDGKL YGVPYMGNTWFMYYDKSKFSDEDVKSLDKMLEKGKVAFDISNAWYLP GFYLDGNMTLFGEKSNDGKAGIKIGDKAAEITKYVAKLIQNKNFVMDGD GAGLSGIKNHTVDAYFSGSWDAGDVKKALGDNYAAAQLPKFKSEDGE HQMKSFAGSKAVAYNPNSKAPEMAAKFAAFLGSKESQEQMYKLHGDI PVAKALSDLVKDNPAAVAQMNTIAKTSVLQPTVPEMGAFWDPMKTFG TALANKEVNDGNAAAKIADFQKGFEEALKK
		α-amylase family glycosyl hydrolase (GH13_23)	>MQNEHAQWLRDAVFYEIYPQSFYDSNGDGIGDIPGIIEKLDYVKSLGC NAIWLNPCYDSPFKDAGYDVRDYKKVAARYGTNEDLQRLFSEAHNRG MHILLDLVPGHTSEEHEWFKQSSRAQENEFSKRYIWTDSAFSGYSMPF IGGESERDATYILNFFKCQPALNYGFAHRDRAWQSAPDSAEAAQTRAA MVDVMRFWLALGADGFRVDMADSLVKNDDNNGPGSLGKDNTIRAWQ EMLGAVKAEYPQAAFVSEWGRPRQALTAGFDMDFYLNWRWDGNPN GYARLARDVDNALDSSPEHDHSYFNARGGGSICPFLDDYCPQYEATC NEGYFSFITCNHDTPRLAPRLSERERRIAFGMILTMPGVPFVYYGDEIG MKYRDIPTKEGGYARTGTRTPMQWDDSKNFGFSTAAKSKLYLPVEAR ADGRTCANSRKQAVQSGEIPTVLGQEQHGDSLLNWVRALIALRHECA ALHADATWRVLHAPVDGRSFAYERCAAVSDGESSTERIIVVMNPGLNE ETIALSALAGLTPEAILQPRLSIGAVACDSSALKLGGQSFALFAL >MQNEHAQWLRDAVFYEIYPQSFYDSNGDGIGDIPGIIEKLDYVKSLGC NAIWLNPCYDSPFKDAGYDVRDYKKVAARYGTNEDLQRLFSEAHNRG MHILLDLVPGHTSEEHEWFKQSSRAQENEFSKRYIWTDSAFSGYSMPF IGGESERDATYILNFFKCQPALNYGFAQRDRAWQSAPDSAEAAQTRAA MVDVMRFWLSLGADGFRVDMADSLVKNDDNNGPGSLGKDNTIRAWQ EMLGAVKAEYPQAAFVSEWGRPRQALTAGFDMDFYLNWRWDGNPN GYARLARDVDNALDSNPEHDHSYFNARGGGSICPFLDDYCPQYEATC NEGYFSFITCNHDTPRLAPRLSERERRIAFVMILTMPGVPFVYYGDEIG MKYRDIPTKEGGYARTGTRTPMQWDDSKNFGFSTAAKSKLYLPVEAR ADGRTCANSRKQAVQSGEIPTVLRQEQDGDSLLNWVRALISLRHECAA LHADATWRVLHAPVDGRSFAYERCAAVSDESDGSNGESSTERIIVVMN PGLNEETIALSALAGLTPEAILQPRLSIGAVACDSSALKLGGQSFALFVL

G. leopoldii	**MalEFG**	ABC transporter substrate-binding protein	>MLKLKKFGAISVMLVAALSLTACGGGTSSTSNAQGKTTVSVWHYWD GANADAFDKKVKAFNASHKNIEIKTTNVPNADFLTKLRASATSNTLPDM AIGDLIWMPQMAQIGKSADLKKLLKPEVIDNINPALKNFGSIDGRLVSVP MSANNLAYMYNRTIFKKAGLDPDKPPKTWAELKQMAKQIKDKTGLPGY DLPTEAGDSGEALTWNFQVNLWQAGGEFLKKDNSAAAFNTPAGKKAA NFWMDLIKSGVSPYAGWGKFEKGKGGSAQEGSWMVGIWAADPPFDF AAAPLPTPEGGKPSTNLGGEQAMLFNNDDAKVKAAAEFVTWFLQDKQ VLDWCESTGFLPATKSAANNAEYRSWVEKKEPRLLPYIEQMATAHTRP NTKLYPQISLAFAKEMEKAFSGEVNVGTALKNAEKAVNEVISQGK
		Sugar ABC transporter permease	>MRNPIVQLKKNKSAANAEAKASNGKLIRDSRKTVGKWEQILTAWGLLL PGLVILVIFTFYPIISAAYTSLTNWNGFSVNPDFIGLKNYVKLSQDPEFW NSLTVTVIYAFGVAVLSVVSGLLFAVLLDAPIRARGLYRTIYFLPVVTSSV AVAIVWKYMLDPSGLVNSVLSSLGIAGPDWLHNKWLALLATTLLTVWK NLGFNIILYLTALQSLSKSTYEAASLDGATSWQQLRYITVPLLRPMTFFV VVQALISSFQSFDLIYVFTEGGPRGGTDVLGMFMYRQAFRLGNFGYGT AIAIASLVLVLIVTLVQWKVSNAENK
		Carbohydrate ABC transporter permease	>MQKISKVFTRLRSLHLARHIVLIIGAASVLIPFMWMFTTSLQTKAETYAV QSVIPTSWHWENYLHAWQSAPFANYYINSLIMSAIIVVGHLAFDAMAAF AFARLEFPGKKILFVLLLSAFMVPAFVTVIPMYGIIAQLRWIDTMAALTVP RLADVFGIILLRQHFTTIPRELDEAARIDGCSSLGIFIKVILPLSKPALATLG VFSFLFAWNDFLWPLLVTNAEEMRTIQIGLSGFVGRYGTSWNYLMAGT LTSTLPSIIIFFFFQKALERGIASTGLKD
		Glycoside hydrolase family 13 protein (GH13_20)	>MQPEWVKSSVFYQIFPERFANANASINPNPVESWDKSPTRDNFFGG DLQGISEHLSYLADMGVNALYLTPIFQADTNHRYDATDYFKIDSRLGND ADFDEFIKRAHELQIRVVLDGVFNHCGIGHWAFQDVVRNGEQSQYVN WFSVEGFPVVSHPEPNYKTCSGCYYLPKWNVYNPEVRRHHYNVAKY WLNRGIDGWRLDVPYFVNKTFWKGFHKVVKSFGDDKYLVAEEWRNP SQWLTPELMDGTMNYTVRDLVLGFCADETLDAYDFADGMNALYNDIP EEHRSCMLNLLGSHDTERLLTRTKAHEYAWDTSAEDAENLAYALLFAA DGAPMVYYGDENGMEGENDPGCRAPMIWNENAWNLSVRKTVCKLM RLRRENAALQYGRQHVAALGKHTVCITRNLPSGRACLCLVHRGSGAYI NCDDLPVRMDRSLFGVPSIIGKQYRMGEKSIILYEGEAVGKL

		LacI family transcriptional regulator	>MNVRRVTIKDVAEKANVSKSAVSFAYNDPSKLSEATVEHIMQTAEEM GYVRSATARALRTNSSNALGLLLPQGIDTILQNPYYSLLIQGIGQVCQSE GYTLMLVPPLRGSMLKAIPTAAVSGFIICGLEAERAEVVEILNKHVPVVIV DPAQPINVPTVEVDDEHDFEDLVDRLIDLGHRNIGIAAIETKSHTDYPG WQGIVGRRMASVVRALKKHNLSPEDKEITVIETPCTYQGGTDVFNKM WNDTNLEYRPTAIVSFSDIIALGVLGAAQQQGLRVPEDLSITGYDDLPLS TICTPRLTTVHQPITSKGRLAAESLIDQIRFDEEALPVSHKKLGTALIVRE SVGAAPLKK
	MalXFGK	Type I pullulanase (GH13_14)	>MTTPSSATWPVYEGPLGFTLNAKSTTFTVWAPTAQNVTLRLFETGNN SEPIQTHELKPLHDGAWRINFRKRLDGIYYDYLVTFQDGTVNRTSDPW GVASGVNGERSLVIDDERVSPANWHKDKSPKFVRHAMVVWETHVAD FSSNPNGGFKPEHRGQYLAFTDEHTSLQNPNSSECNFPTGLDYLKQL GITHVQLEPIYDFATVDESTIDNARKNGASAEELDSLYNWGYDPHAYNV PEGAYSSNPYDGTARIRECKAMIAALHANGQRVIMDVVYNHMYKSAEN AFECMAPGYFCRRDSDGSFANGSGCGNEMASEHPMFRRFIIESLLHW ARCYHIDGFRFDLMGLTDTETLNQARTALDKLPGGKDIVMYGEPWGAS DPAVDEKVPMGDKDGRDTLNERIGWFSDESRDSLKGNVMNHRDRGY VNGNEFWCSDAARQTLNAWRDTPSAGKEVGQIVQYTSAHDDLTLWD KLCISMRENASDFEFDAQDVQQVADILNANMFAAGIVLCSAGTPFMLS GEEFGRTKHGCDNSYNRGVLMNQLDWARTERFNPLVQWYRALIKLRA EYSSLYDGKRKQLETSNHTVLATRIGNNLIVCANPMYHEDATIQLPAIHS AKSQEWHCVLNSSDYFAKWTGKQPDTKVNITTEKSTATLHLPIRTFVVL SNK
		Sugar ABC transporter permease	>MSDKTTTKNEEAVVKSTRSVSFLKDQHVRRVIGDTFSHLFLLVLSVIW LIPILWVLLESFNKNTGSYQKTFFPTEYSLNNYIKLFTDTDDINFPRMFLN TLIVSIFVCIISLVFVLAVSFCMSRLRFKGRKTFMNLALILGMFPGIMAVIAI YFILKALGLTKGSIVLIALIIVYSAGSGAGFYIMKGYMDTIPKSLDEAAYLD GCTKWQVFWKIIVPICKPMIVFQAITGFLGPWLDFVMVKTVVRDNKENY TVAYGLMEMLTKKYLPDWYAAFAAGAVCISIPIVILFIIMQRFYEESMSG SVKG
		Sugar ABC transporter permease	>MTGTTSTKRERRQKKTSTAERFAPSPYSLKEAAKHADILSWISVVIFG FGNFVRKQYIKGVIFLACEAALFGFLFTIGLDYLPKINTLGTVEQGRVKV NGFWRYSAGDNSVEILLYGVMSILAIVLLVYLATLALRSAYKAQSVMAD

			GKKPLNFIDDAKTLLDEKAHVGFMTLPTAGIVLFTILPLFFMISMAFTSYD ENHQQKFSWIGLDVFSQFLTGKSDLIKADVFFDVLIWTLVWAFLATFLN FFLGMFIAMIINRKSTRLKGLWRTCFSLTIAVPQFVSLLIMNNMLSQNGV INIVFKQNGWIHQELPFLTDVTWARVTVILINLWVGIPFTIMQITGILQNVP AELYEAAELDGAGWWQRFMNVTMPYIIFVMTPYLITTFTGNINNFNVIYL LTKGGPASSKVNSAGGTDLLITWLYKLTVDQQNYNAGAVIGILTFMVLA VVALITYRNSGSYKNEEGFR
		Extracellular solute-binding protein/ ABC Transporter substrate-binding protein	>MTNIKKMLGAGLAVATLFGFAACGSSTNNSDKKAADSAPVQLTVWGS ADDQKDGGWIKTVEAEFKKENPNVTFKNAVVEPPDAAKAVKQDAKAA ADVFLFPNDQLGDLVKAGAIGELSDDATKQVKEQNEDTLIKSVTAQDG KLYGVPYTGNTWFMYYNKSKFSADDVKSLDKMLEKGKVSFDLANAWY LPGFYLDGNMTLFGESGNDAKAGMKLSDKAAEVTKYIATKLVGNEKHF VLDGNGRGLAGLKDGSVDAIFSGSWDAAKVKDALKDNYAAAALPTFKT ESGEHQMKSFAGSKAVAYNTNTKSPEMAAKFAAFLGSKKSQQIAFEKE GAIPSDKSLADSVKAEPAAVAQMETIAKTSVLQPAIPEMGKFWDPVQA FGTSISTGKITGDNAAAQTAEFAKGLEK
		α-amylase family glycosyl hydrolase (GH13_23)	>MALYDHAQWLKDAVFYEIYPQSFCDSNGDGIGDIQGIIEKLDYVKSLG CNAIWLNPCYDSPFKDAGYDVRDYKKVAARYGTNEDLKQLFSEAHNR GIRILLDLVPGHTSEEHEWFKESSKAEENEFSKRYIWTDSAFSGYSMPF IGGETERDATYILNFFKCQPALNYGFAHRDRAWQSAPDSKEAAETRAA MVDIMRFWLSLGADGFRVDMADSLVKNDDNNGEGSLGKDNTIRAWQ EMLGTVKEEYPQAAFVSEWGRPRQALTAGFDMDFYLNWRWDGNPN GYARLARDVDNALDGNPEHDHSYFNAHGGGSICPFLDDYCPQYESTH SQGYFSFITCNHDTPRLAPRLGDRERRVAFGMILTMPGVPFVYYGDEI GMKYRDIPTKEGGYARTGTRTPMQWDDAKNFGFSMAAKSKLYLPVEA RADGRTCANSRKQAVESGEIPTVSAQINCEDSFLSWVRSLIALRHSRK SLQADASWRVLYAPVDGRGFAYERCAKGDVGESAVERSIVVMNPGVN SETVSLEALANLTQESAYNPLLKIGEISVNGDCLTLGAQSFAVFGM
	MusE$_2$FGK$_2$I	Histidine phosphatase family protein	>MQNNGTKESSQKECSKQGCLVLLRHGQTAWSVSGQHTGRTDLPLT EVGVEQAKEAGARLQCVFTKGFDKDCVLVSPLRRAQQTAQFAGFDSY TICDSALEWDYGAAEGRTRADISKLSGVEAWDVWKHGPKVLPSHMVS DAKEVLPSGEAVDVHRTSGESLQEVSDRTQKIIDMAMPKLKLGKDVLIV

			AHAHVLRILTARWLGVSPDFARLLRMDTAHYSVLGAYKGDNVIVHWNC
		ABC transporter substrate-binding protein	>MMNRTIKSTIALVACSAMAFGALAACGSSSSDSSAKGKVYFLNFKPEA ADQWKEVAKAYTKKTGVEVKVQTAASGTYEQSLKSEISKGDEAPTLFQ INGPVGYASWKDYTADLKDTELYSQLQNKDIALKDGDKVVGIPYAMET YGLIYNKALLKKYTELKDAKVKSAEEINSFDKLKAVADDIQKHKSDLGVE GAFTSAGFDSSSDWRFKTHLANLPLYYEFKEDHVTKQPAKVKGSFLD GYKKIFDLYITDSTTPATKIGSKTGEDANAEFSQGKAVFYQNGTWAWG DLSKAGMKAEDLGMLPIYIGAKGEEKQGLATGSENYWCINEKTSDANK KATKDFLKWLLTSDAGKDALSKKMGFTTPFKSFADIKSDNPLTQAAVE DAKSGKTPVSWNFTVMPSDNWKNDLGSALLEYAQGTGKWDKVKTAF VDGWAKEYSQAHEDDD
		LacI family transcriptional regulator	>MSANIQDVANKAHVSVSTVSRSFTRPNLVSASTRNKVLAIAEQLNFSI SKSAAALKSGKTLRIALLISDQIRLWFSASIIQGLNQVFHTAGYDLSIFQIS SSKERSEFFTMLPTRRNADAVVVCSFDVNTNEIAQLKSTGVPIVGINCL YPQKCDFDATINIDDDQGARLMARHLIGLGHHNIAYVRTTRDVSLHFSV LQRYHSFIDECQNSGITPTEIVAPANSDRISAIVSMLLGSSIMPTAIACQE DGIAIPLMFQLARSGYSTPKDVSIIGFDDSFYAHETGLTTIRQDPVDIAST AANITLALINAEKVEDQYRIVPAQLIVRSSTAALLKE
		Sugar ABC transporter permease	>MISMSGKAIRKWWWLFALPTFAAFVVGFLVPFIMGVYLSFRKFTTVTD GKFVGFRNYYRALKDPHLVYSFLFTTLFAVITTIIINVIAFAIAYMLTKKFK GTTIFRAVFFMPNLIGGIVLSYIWQLLLNGILGYFNRTITFSEVFGFWGLV IVVCWQQIGYMMIIYIAGMQALPGDVMEAAAVDGASPRQTLFKITIPLM MPSITVCTFLTITNGFKLFDQNFALTNGGPSRKTEMLALNIYRTFTTPRC EGVGQAQAVMFFIVVAVIAMVQNMLTRSKEVAA
		Carbohydrate ABC transporter permease	>MNEKTKHPALWTVLFTLISLLWIFPIALVVLNSFKSKVDIASNPFTFSSK SFVGMSNYVLGSNRTDFPMSFLWTIVVTVGSVILILLCTSMCAWWIVRV NNWAAKLLYVLFLFNMIVPFQMVMYTLAYITNTLKLNTPWGLWIIYLGF GAGLAVFIFTGVIKGIPQELEESAMIDGASVPRIFFQIIVPIMRPAIISVSIL QAMWIWNDFLLPFLTLDLNKYKTISIAVQYLKGGYGSVEMGAMMGCLV MAIVPIIIFYLICQKYIIKGVMSGAVKG

G. swidsinskii	**MusE_2FGK_2I**	Histidine phosphatase family protein	>MQNNGTKESSQKECSKQGCLVLLRHGQTAWSVSGQHTGRTDLPLT EVGVEQAKEAGARLQCVFTKGFDEDCVLVSPLRRAQQTAQFAGFASY TICDSALEWDYGAAEGRTRADISKLSGVEAWDVWKHGPKVLPSHMVS DAKEVLPSGETVDVHRTSGESLQEVSDRTQKIIDMAMPKLRLGKDVLV VAHAHVLRILTARWLGVNPDFARLLRMDTAHYSVLGTYKGDNVIVHWN C
		ABC transporter substrate-binding protein	>MMSRSVKSIIALVVCAVLAACAFTAFGFSQSFGANADRSANGKVIFFN LKPEATAMWEEVAKEYTKQTGVTVSIVTPTGGQYEKRLKTEVEKYKDN PDSMPTLFQINGPVGYANWKDYTADLTNSPLYTQLSNKNLAFMDNGK PVAVPYVTENYGLIYNKALLARYADLKDAKLQPEENGKSFEDQITNFAA LKAVADDLQKRKDELGIDGAFASAGFDLSSDWRFKTHLANMPLYYQFV KNNVNGQPQSINYDYVDNFRKIFDLYISDSTTPVSQLSAKTGEDAVYEF ALGKAVFYQNGTWAWGDLSKAGIKPDDVGMLPIYIGAKGEEKQGMAT GSENYWCINKKASRRNQEATNAFLNWLLTTNYGKDALANKMGFAAPF KGFPKAENPLVQAAVDYTAKPGKLPVKWVFVTMPSNGWKDDVGSDLL TYAQATKDGMTSIGANKAWATLKSAFVDGWAKEYKIAHGQS
		ABC transporter substrate-binding protein	>MMNRTIKSTIALVACSAMAFGALAACGSSSSDSSAKGKVYFLNFKPEA ADQWKEVAKAYTKKTGVEVKVQTAASGTYEQSLKSEISKGNEAPTLFQ INGPVGYASWKDYTADLKDTELYSQLQNKDIALKDGDKVVGIPYAMET YGLIYNKALLKKYTELKDAKVKSAEEINSFDKLKAVADDIQKHKSDLGVE GAFTSAGFDSSSDWRFKTHLANLPLYYEFKEDHVTKQPAKVKGSFLD GYKKIFDLYITDSTTPATKIGSKTGEDANAEFSQGKAVFYQNGTWAWG DLSKAGMKAEDLGMLPIYIGAKGEEKQGLATGSENYWCINEKTSDANK KATKDFLKWLLTSDAGKDALSKKMGFTTPFKSFADIKSDNPLTQAAVE DAKSGKTPVSWNFTVMPSDNWKNDLGSALLEYAQGTGKWDKVKTAF VDGWAKEYSQAHEDD
		LacI family transcriptional regulator	>MSANIQDVANKAHVSVSTVSRSFTRPNLVSASTRNKVLAIAEQLNFSI SKSAAALKSGKTLRIALLISDQIRLWFSASIIQGLNQVFHTAGYDLSIFQIS SSKERSEFFTMLPTRRNADAVVVCSFDVNTNEIAQLKSTGVPIVGINCL YPQKCDFDATINIDDDQGARLMARHLIGLGHRNIAYVRTTRDVSLHFSV LQRYHSFIDECQNSGITPTEIVAPANSDRISAIVSMLLGSSIMPTAIACQE DGIAIPLMFQLARSGYSTPKDVSIIGFDDSFYAHETGLTTIRQDPVDIAST

			AANITLALINAEKVEDQYRIVPAQLIVRSSTAALLKE
		Sugar ABC transporter permease	>MISMSGKAIRKWWWLFALPTFAAFIVGFLVPFIMGVYLSFRKFTTVTD GKFVGFRNYYRALKDPHLVYSFLFTTLFAVITTIIINVIAFAIAYMLTKKFK GTTIFRAVFFMPNLIGGIVLSYIWQLLLNGILGYFNRTITFSEVFGFWGLV IVVCWQQIGYMMIIYIAGMQALPGDVMEAAAVDGASPRQTLFKITIPLM MPSITVCTFLTITNGFKLFDQNFALTNGGPSRKTEMLALNIYRTFTTPRC EGVGQAQAVMFFIVVAVIAMVQNILTRSKEVAA
		Carbohydrate ABC transporter permease	>MNEKTKHPALWTVLFTLISLLWIFPIALVVLNSFKSKVDIASNPFTFSSK SFVGMSNYVLGSNRTDFPMSFLWTIVVTVGSVILILLCTSMCAWWIVRV NNWAAKLLYVLFLFNMIVPFQMVMYTLAYITNALKLNTPWGLWIIYLGF GAGLAVFIFTGVIKGIPQELEESAMIDGASVPRIFFQIIVPIMRPAIISVSIL QAMWIWNDFLLPFLTLDLNKYKTISIAVQYLKGGYGSVEMGAMMGCLV MAIVPIIIFYLICQKYIIKGVMSGAVKG
	MalXFGK	Type I pullulanase (GH13_14)	>MTTPSSATWPVYEGPLGFTLNAKSTTFTVWAPTAQNVTLRLFETGNN SEPIQTHELKPLHDGAWRINFRKRLDGIYYDYLVTFQDGTVNRTSDPW GVASGVNGERSLVIDDERVSPANWHKDKSPKFVRHAMVVWETHVAD FSSNPNGGFKPEHRGQYLAFTDEHTSLQNPDSSECNFPTGLDYLKQL GITHVQLEPIYDFATVDESAIDNARKNGASAEELDSLYNWGYDPHAYN VPEGAYSSNPYDGTARIRECKAMIAALHANGQRVIMDVVYNHMYKSAE NAFECMAPGYFCRRDSDGSFANGSGCGNEMASEHPMFRRFIIESLLH WARCYHIDGFRFDLMGLTDTETLNQARAALDELPGGKDIVMYGEPWS ASDPAVDEKVPMGDKDGRSTLNERIGWFSDESRDSLKGNVMNHLDR GYVNGNELWCSDAARQTLNAWRDTPSAGKEVGQIVQYTSAHDDLTF WDKLCISMRENASDFEFDAQDAQQVADILNANMFAAGIVLCSAGTPFM LSGEEFGRTKHGCDNSYNRGVLMNQLDWARAERFNPLVQWYRALIKL RAEYSSLYDGKRKQLETSNHTVLATRIGNNLIVCANPMYHEDATIQLPAI HSAKAQEWHCVLNSSDYFAKWTGKQPDTAVNITTEKSAATLHLPIRTF VVLSNK
		Sugar ABC transporter permease	>MSDKTTTKNEEAVVKSTRSVSFLKDQHVRRVIGDTFSHLFLLVLSVIW LIPILWVLLESFNKNTGSYQKTFFPTEYSLNNYIKLFTDTDDINFPRMFLN TLIVSIFVCIISLVFVLAVSFCMSRLRFKGRKTFMNLALILGMFPGIMAVIAI

			YFILKALGLTKGSIVLIALIIVYSAGSGAGFYIMKGYMDTIPKSLDEAAYLD GCTKWQVFWKIIVPICKPMIVFQAITGFLGPWLDFVMVKTVVRDNKENY TVAYGLMEMLTKKYLPDWYAAFAAGAVCISIPIVILFIIMQRFYEESMSG SVKG
		Sugar ABC transporter permease	>MTGTTSTKRERRQKKTSTAERFAPSPYSLKEAAKHADILSWISVVIFG FGNFVRKQYIKGVIFLACEATLFVFLFAIGLDYLPKINTLGTVEQGRVKV NGFWRYSAGDNSVEILLYGVMSILAIVLLVYLATLALRSAYKAQSVMAD GKKPLNFIDDAKTLLDEKAHVGFMTLPTAGIVLFTILPLFFMISMAFTSYD ENHQQKFSWIGLDVFSQFLTGKSDLIKADVFFDVLIWTLVWAFLATFLN FFLGMFIAMIINRKSTRLKGLWRTCFSLTIAVPQFVSLLIMNNMLSQNGV INIVFKQNGWIHQELPFLTDVTWARVTVILINLWVGIPFTIMQITGILQNVP AELYEAAELDGAGWWQRFMNVTMPYIIFVMTPYLITTFTGNINNFNVIYL LTKGGPASSKVNSAGGTDLLITWLYKLTVDQQNYNAGAVIGILTFIVLAV VALITYRNSGSYKNEEGFR
		Extracellular solute-binding protein/ ABC Transporter substrate-binding protein	>MTNIKKMLGAGLAVATLFGFAACGSSTNNSAKKAADSAPVQLTVWGS ADDQKDGGWIKTVEAEFKKENPNVTFKNAVVEPPDAAKAVKQDAKAA ADVFLFPNDQLGDLVKAGAIGELSDDATKQVKEQNEDTLIKSVTAQDG KLYGVPYTGNTWFMYYNKSKFSADDVKSLDKMLEKGKVSFDLANAWY LPGFYLDGNMTLFGENGNDAKAGMKLSDKAAEVTKYIATKLVGNEKHF VLDGNGRGLAGLKDGSVDAMFSGSWDAAKVKDALKDNYAAAALPTFK TESGEHQMKSFAGSKAVAYNTNTKSPEMAAKFAAFLGSKKSQQIAFEK EGAIPSDKSLADSVKAEPAAVAQMETIAKTSVLQPAIPEMGKFWDPVQ AFGTSISTGKITGDNAAAQTAEFAKGLEK
		α-amylase family glycosyl hydrolase (GH13_23)	>MALYDHAQWLKDAVFYEIYPQSFCDSNGDGIGDIQGIIKKLDYVKSLG CNAIWLNPCYDSPFKDAGYDVRDYKKVAARYGTNEDLKQLFSEAHNQ GIRILLDLVPGHTSEEHEWFRESSKAEENEFSKRYIWTDSAFSGYSMPF IGGETERDATYILNFFKCQPALNYGFAHRDRAWQSAPDSKEAAETRAA MVDIMRFWLSLGADGFRVDMADSLVKNDDNNGEGSLGKDNTIRAWQ EMLGTVKEEYPQAAFVSEWGRPRQALTAGFDMDFYLNWRWDGNPN GYARLLRDVDNALDNNSERDRSYFNAHGGGSICPFLDDYCPQYESTC NQGYFSFITCNHDTPRLAPRLGDRERRVAFGMILTMPGVPFVYYGDEI GMKYRDIPTKEGGYARTGTRTPMQWDDAKNFGFSMAAKSKLYLPVEA RADGRTCANSRKQAVESGEIPTVSAQINCEDSFLSWVRSLIALRHSRK

			SLQADASWRVLYAPVDGRGFAYERCAKGDAGESAVERSIVVMNPGVN SETVSLEALANLTQESAYNPLLKIGEISVDGDCLTLGAQSFAVFGM
G. greenwoodii	**MalEFG**	ABC transporter substrate-binding protein	>MLKLKKFGAISMMLAAALSLTACGGGTSSTSNTQGKTTVSVWHYWD GANADAFDKKVKAFNASHKNIEIKTTNVPNADFLTKLRASATSNTLPDM AIGDLIWMPQMAQIGKSADLKKLLKPEVIDNINPALKNFGSIDGHLVSVP MSANNLAYMYNRTVFKKAGLDPDKPPKTWDELKQMAKQIKDKTGLPG YDLPTEAGDSGEALTWNFQVNLWQAGGEFLKKDNSAAAFNTPAGKKA ANFWMDLIKSGVSPYAGWGKFEKGQGGSAQEGSWMVGIWAADPPF DFAAAPLPTPEGGKPSTNLGGEQAMLFNNDDAKVKAAAEFVTWFLQD KQVLDWCESTGFLPATKSAASNAEYRSWVEKKEPRLLPYIEQMATAHT RPNTKLYPQISLAFAKEMEKAFSGEVNVDTALKNAEKAVNDVIAQGK
		Sugar ABC transporter permease	>MRNPIAQLKNKQPATIDNVKVGSSKLMRNNHKTVGKWEQILTAWGLL LPGLVILVIFTFYPIISAAYTSLTNWNGFSANPDFIGLKNYVKLSQDPEFW NSLTVTVIYAFGVAVLSVVSGLIFAVLLDAPIRARGLYRTIYFLPVVTSSV AVAIVWKYMLDPSGLVNSVLSSLGISGPDWLHDKWLALFATIFLTVWKN LGFNIILYLTALQSLSKSTYEAASLDGATSWQQLRYITVPLLRPMTFFVV VQALISSFQSFDLIYVFTEGGPRGGTDVLGMFMYRQAFRLGNFGYGTA IAIASLVLVLIVTLVQWKVSNAENK
		Carbohydrate ABC transporter permease	>MQKISKVFTRLRSLHLARHIVLIVGAASVLIPFMWMFTTSLQTKAETYA VQSVIPTSWHWENYLHAWQSAPFANYYINSLIMSAIIVVGHLAFDAMAA FAFARLEFPGKKILFVLLLSAFMVPAFVTVIPMYGIIAQLRWIDTMAALTV PRLADVFGIILLRQHFTTIPRELDEAARIDGCSSLGIFIKVILPLSKPALATL GVFSFLFAWNDFLWPLLVTNAEEMRTIQIGLSGFVGRYGTSWNYLMA GTLTSTLPSIIIFFFFQKALERGIASTGLKD
		Glycoside hydrolase family 13 protein (GH13_20)	>MQPEWVKSSVFYQIFPERFANANASINPNPVESWDKSPTRDNFFGG DLQGISKHLPYLADMGINALYLTPIFQADTNHRYDATDYFKIDSRLGNDS DFDEFIKRAHELQIRVVLDGVFNHCGIGHWAFQDVVRNGEQSQYVNW FSVEGFPVVSHPEPNYKTCSGCYYLPKWNVYNPEVRRHHYNVAKYWL NRGIDGWRLDVPYFVNKTFWKGFHNIVKSFGDDKYLVAEEWRNPSQ WLTPELMDGTMNYTVRDLVLGFCADETLDAYDFADGMNALYNDIPEE HRSCMLNLLGSHDTERLLTRTKAHEYAWDTPAEDAENLAYALLFAADG

			APMVYYGDENGMEGENDPGCRAPMIWDENAWNLSVRKTVCKLMRLR RENAALQYGRQHVAALGKHTVCITRNLPSGRACLCLVHRGSGAYINCD DLPVRMDRALFGVPSIIGNQYRMGEKSIIMYEGEAVGKL
		LacI family transcriptional regulator	>MNVRRVTIKDVAKKANVSKSAVSFAYNDPSKLSEATVEHIMQTAEEM GYVRSATARALRTNSSNALGLLLPQGIDTILQNPYYSLLIQGIGQVCQSE GYTLMLVPPLRGSMLKAIPTAAVSGFIICGLEAERAEVVEILNKHVPVVIV DPAQPINVPTVEVDDEHDFEDLVDRLIDLGHRNIGIAAIETKSHTDYPG WQGIVGRRMASVVRALKKHNLSPEDKEITVIETPCTYQGGADVFNKM WNDTNLECRPTAIVSFSDIIALGVLGAAQQQGLRVPEDLSITGYDDLPL STICTPRLTTVHQPITSKGRLAAESLIDQIRFDEEALPVSHKKLGTALIVR ESVGAAPLKK
	MalXFGK	Type I pullulanase (GH13_14)	>MTFPSSATWPVYEGPLGFTLNSKSTTFTMWAPTAKTVTLRLFETGNN SEPTQTHELKPLHDGAWRINFRKRLDGIYYDYLIAFPDGTVNRTSDPW GIASGVNGERSLVVDEERISPANWHKDKSPKFVRHAMVVWETHVADF SSNPNGGFKPENRGQYLAFTEEHTTLQKPDSSDCNFPTGLDYLKKLG VTHVQLEPIYDFATVDEAAIDNALKNGASAEELDSLYNWGYDPHAYNV PEGAYSSNPYDGTARIRECKAMIASLHANGQRVIMDVVYNHMYRSAEN AFECVAPGYFCRRNADGSFSNGSGCGNEMASEHPMFRRFIIESLLHW ARCYHIDGFRFDLMGLTDTETLNQARAALDELPGGKDIVMYGEPWGA SDPAADETISLGNKEGRHTLNERIGWFSDESRDSLKGNVMNHRDRGY VNGNEFWCCDAARQALDAWRDTPSAGKEVGQIVQYTSAHDDLTLWD KLCISMRDNAGDFEFNAQDKQQVADILNANMFAAGIVLCSAGIPFMLSG EEFGRTKHGCDNSYNRGVLLNQLDWARAEQFKPLVNWYKDLIALRAE YSALYDGKRKQLTTSNRTVLATRIGNNIIVCANPMYHEDATVELPTIHST KSKGWNCILNSNDYYSEWTGSKPESDVFIANEENNTTLHLPSRTFVVL SAK
		Sugar ABC transporter permease	>MSNKNTAMQQEPVVHSTRSVSLFKDQRIRRIIGDIFSHLFLAVLSIIWLI PIVWVFLESFNKNESPFQKTFFPTEYSFNNYIRLFTEREALDFPRMFLN TLVVSIFVCIISLLFVLAVSFCMSRLRFRGRKTFMNIALILGMFPGIMAVIA IYFILKALGLTNGSVVLIALIIVYSAGSGAGFYVMKGYMDTIPKSLDEAAY LDGCTTWQVFYKIILPICKPMIVFQAITSFLGPWLDFVMVKTIANSQDNY TVALGLWKMLDREFISHWFARFCAGAVCVSIPIAILFIIMQRFYEESMSG

			SVKG
		Sugar ABC transporter permease	>MTGTTSTKRERRQKKTSNIERFAPSPYSLKEAAEHADILSWISVIIFGF GTFVRKQFIKGILFLLCEAVLFGFLFTTGAEYLPKILTLGTKKQGKIKVNG FYHYTAGDNSVEILLYSVMAIAAIIVLIYLATLSLRSAYKAQAITASGQKPV NFIKDASSLLDSKAHIGLMTLPTAGIVLFTILPLIFMISMAFTSYDRKHPVL FTWTGLNNFFAIGNAKSSINAATFFDVLLWTLIWAFFATFLNFFFGMFIA MIINRKSTRLKGLWRTCFSLTIAVPQFVSLLVMNNMLQRNGAINSILLEN HWISQPLPFFQDVTWARVTVILINLWVGIPFTIMQITGILQNVPAELYEAA ELDGASWWQRFMNVTMPYIIFVMTPYLITTFTGNINNFNVIYLLSGGEP TPLDSSAGKTDLLITWLYKLTVDRGDYNIGAVIGIMTFIVLAVVALITYRN SGSYKNEEGFR
		Extracellular solute-binding protein/ ABC Transporter substrate-binding protein	>MTNIKKMLGAGLAVATLFGFAACGSSTNSDNKAAESGPVQLTVWSAS EDQKAGGWVDTVEKEFQKANPNVKFKNTVVEPPKAAESVKKDAKAAA DVYLFPNDQLGDLVKAGAIGQLSDDAVKQVKEQNEDTLVNSVTAQDG NMYGVPYTGNTWFMYYNKAKFSADDVKSLDKMLEKGKVAFDLSNAW YLPGFYLDGNMTLFGEKGTDAKAGMKLGDKAAEVTKYIAKLFANKNFV LDGNSAGLNGLKSGKVDAYFSGSWDAADVKKALGDNYAAAALPTFKT ESGEHQIKSFAGSKAVAYNPNSKHPEMAAKFAAFLGSKKSQQIAYEKQ GVIPSDKSLADSVKSDPAAVAQMETIAKTSVLQPAIPEMGSFWEPMGN FGKAIANHEVTEANADAKTADFAKGLAGAVK
		α-amylase family glycosyl hydrolase (GH13_23)	>MGLYNHAQWLKDAVFYEIYPQSFYDANGDGIGDIRGIIEKLDYVKSLG CNAIWLNPCYDSPFKDAGYDVRDYKKVAARYGTNEDLQQLFNEAHTR GMHILLDLVPGHTSEEHEWFRESSKAQENEFSKRYIWTDSAFSGYSM PFIGGESERDATYILNFFKCQPALNYGFAHRDRAWQSSPDSKEAAETR AAMVDVMRFWLSLGADGFRVDMADSLVKNDDNNGEGSLGKDNTIRA WQEMLGAVKEEYPEAAFVSEWGRPRQALAAGFDLDFYLSWRWDGN PNGYARLARDVDNALDSNPEHDHSYFSARGGGSICPFLDDYCPQYES TCNQGYFSFITCNHDTPRMAPRLDERERRVAFGMLLTMPGVPFVYYG DEIGMKYRDIPTKEGGYARTGTRTPMQWNDTKNLGFSTAAKSKLYLPV EARADGRTCANSRKQAVESGEITTVSAQINCEDSFLSWVRSLIALRHN RASLQADASWRVLYAPVDGRGFAYERCAKNNEGESASERSIVVMNPG VNAETVPFEALANLTQEAVENPLLKIGEVLSDGNSLKLGAQSFAIFNLPL

Connecting the Dots: phenotypic and genotypic antimicrobial resistance in *Gardnerella*

Márcia Sousa,[1,2] Filipa Grosso,[1,2] Teresa Gonçalves Ribeiro,[1,2,3] Luísa Peixe [1,2,3]

[1]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [2]UCIBIO – Applied Molecular Biosciences Unit, Department of Biological Sciences, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [3]Culture Collection of Porto – Faculty of Pharmacy, University of Porto (CCP).

Manuscript in preparation

Abstract

Antimicrobial resistance (AMR) in *Gardnerella* spp. remains poorly characterized, in part due to the absence of epidemiological *cut-off* values (ECOFFs) and clinical breakpoints. This study aimed to evaluate minimum inhibitory concentrations (MICs) and antimicrobial resistance genes (ARGs) in *Gardnerella* isolates from healthy women and overactive bladder donors, bridging the gap between genotypic and phenotypic resistance and informing future ECOFF development.

Thirty-three *Gardnerella* isolates, previously recovered from urine and vaginal swabs and belonging to *Gardnerella vaginalis* (n=12), *Gardnerella pickettii* (n=7), *Gardnerella leopoldii* (n=7), *Gardnerella swidsinskii* (n=6), and *Gardnerella greenwoodii* (n=1) species were analyzed. ARGs were identified in genomes using ResFinder. MICs for antibiotics were determined using agar dilution (CLSI M11-A6), and interpretation followed EUCAST guidance for species without established breakpoints.

ARGs were detected in 72.7% (24/33) of isolates, with *lsaC* (72.7%) being the most prevalent, often alongside *mefA* and *msrD* (12.1%). Despite *lsaC* prevalence, only four isolates exhibited clindamycin MICs ≥1 mg/L, indicating a low genotype-phenotype correlation. All *mefA/msrD*-harboring strains showed low erythromycin MICs. Two isolates (*G. pickettii* and *G. leopoldii*) displayed high metronidazole MICs (64 and 8 mg/L, respectively), yet no known metronidazole resistance genes were detected. A unique 16 amino acid deletion in MurF was found in *G. pickettii* isolate c10Ua_31, potentially linked to its elevated vancomycin MIC (32 mg/L).

Despite high ARG prevalence, MIC values remained low for most antibiotics. These findings highlight the importance of MIC determination in *Gardnerella* isolates to establish baseline resistance levels, providing essential reference data for future ECOFF development and AMR surveillance. The observed discrepancies between genotype and phenotype and the identification of potential novel resistance mechanisms warrant further investigation.

Introduction

Gardnerella is a genus of facultative anaerobic bacteria, historically represented by the single species *Gardnerella vaginalis*. However, recent taxonomic revisions conducted by Vaneechoutte and colleagues identified three additional species, including *Gardnerella piotii, Gardnerella leopoldii* and *Gardnerella swidsinskii* (1). Sousa *et al*. further expanded the genus with the description of *Gardnerella pickettii* and *Gardnerella greenwoodii* (2). Members of the *Gardnerella* genus can be commensal inhabitants of the urine or vagina in healthy women (3, 4). However, they are also frequently implicated as key microorganisms in bacterial vaginosis (BV), a dysbiosis associated with increased risk of adverse health outcomes, including preterm birth, pelvic inflammatory disease, and susceptibility to sexually transmitted infections (5–7).

Although antimicrobial resistance (AMR) among anaerobic bacteria increased worldwide over the past decade (8–10), few studies comprehensively evaluated AMR within the *Gardnerella*, particularly considering the recent species delineation (11-18). AMR profiles of *G. vaginalis* have been reported with some variability (11-18). For instance, a study conducted by Kharsany *et al*., (1993) revealed that *Gardnerella* was susceptible to penicillin, ampicillin, erythromycin, clindamycin, chloramphenicol, and trimethoprim, while showing poor susceptibility to vancomycin, cephalosporins, ciprofloxacin, and imipenem (16). Tetracycline and minocycline exhibited intermediate susceptibility, whereas aztreonam, amikacin, and sulfamethoxazole showed marked resistance (16). A recent study reported significant variability in AMR patterns, particularly for metronidazole, clindamycin, and tetracycline, with some strains exhibiting MIC values ≥128 µg/mL for metronidazole (18).

The most commonly detected antimicrobial resistance genes (ARGs) in *Gardnerella* include *tet* genes (*tetL, tetM* and *tetQ*) associated with tetracycline resistance (16,18), *lsaC* associated with resistance to clindamycin, lincomycin, dalfopristin, and tiamulin resistance (18, 20), and *ermX* and *mefA* linked to macrolide resistance (18, 21) Metronidazole resistance was associated with *nimJ* (20), *yadH* and *mcrA* (18), as well as alterations in OmpR and NadE proteins (22). Despite the presence of these genes, a low correlation between genotypic and phenotypic resistance has been observed, suggesting that additional regulatory factors, uncharacterized resistance mechanisms, or variations in gene expression may be involved (18, 21).

A significant challenge in studying *Gardnerella* AMR is the lack of standardized antimicrobial susceptibility testing (AST) methods and established epidemiological *cut-off* values (ECOFFs) within the European Committee on Antimicrobial Susceptibility Testing (EUCAST) database. Without established ECOFFs, the interpretation of MIC values remains challenging, particularly in non-clinical isolates, where AMR selection pressures may differ from those in clinical settings. Defining MIC distributions in non-clinical *Gardnerella* isolates is therefore essential for

establishing baseline resistance levels, identifying potential AMR gene reservoirs, and providing crucial reference data for future EUCAST breakpoint development.

This study aimed to address the gap between genotypic and phenotypic resistance in *Gardnerella* spp. by evaluating MIC values and ARG prevalence in isolates from healthy women and women diagnosed with overactive bladder. By characterizing MIC distributions in different *Gardnerella* species from different origins, this study contributes critical data for AMR surveillance and helps establish baseline resistance levels in *Gardnerella* spp., providing valuable insights for future antimicrobial susceptibility guidelines and therapeutic strategies.

Material and methods

Bacterial Collection and *in silico* detection of antimicrobial resistance genes (ARGs)

Gardnerella strains were isolated from urine and vaginal swab samples collected from eight healthy women (HW) and three women diagnosed with overactive bladder (OAB), as part of an ongoing study on the female urogenital microbiome (22, 23). A total of 33 isolates previously identified by *cpn60* as *Gardnerella vaginalis* (n=12), *Gardnerella pickettii* (n=7), *Gardnerella leopoldii* (n=7), *Gardnerella swidsinskii* (n=6), and *Gardnerella greenwoodii* (n=1) were stored in tryptic soy broth (Liofilchem) supplemented with 20% (v/v) glycerol at −80°C until further analysis. The isolates were revived from frozen stocks, subcultured on Columbia agar with 5% sheep blood (Biogerm, Portugal), and incubated at 37°C for 48h under anaerobic condition to assure purity before testing. All strains were deposited and are available at the Culture Collection of Porto (CCP; https://ccp.ff.up.pt/). Antimicrobial resistance genes (ARGs) and chromosomal point mutations (minimum identity of 90% and a minimum gene coverage of 60%) were identified by ResFinder 4.6.0 (http://genepi.food.dtu.dk/resfinder).

Antimicrobial susceptibility testing

The minimum inhibitory concentrations (MICs) of *Gardnerella* isolates were determined using the reference agar dilution method (Wadsworth method) in accordance with CLSI guidelines M11-A6 (24) on Fastidious Anaerobe Agar (FAA; Neogen, USA), supplemented with 5% of lysed horse blood (ThermoScientific, USA).

Susceptibility was tested against 7 antimicrobial agents: benzylpenicillin (ThermoScientific, USA), piperacillin-tazobactam (piperacillin: Sigma, Germany; tazobactam: ThermoScientific, USA), meropenem (ThermoScientific, USA), vancomycin (Fisher BioReagents, UK), clindamycin (Arcos Organic, UK), metronidazole (ThermoScientific, USA), and erythromycin (ThermoScientific, USA). Before the start of each experiment, solutions of all but one antimicrobial agent (metronidazole in dimethyl sulfoxide) were prepared in sterile ultrapure water at a concentration

of 5.120 mg/mL, filtered by a filter membrane to ensure sterility, and stored at –20°C. The final antimicrobial concentrations were obtained via serial two-fold dilutions (512 to 0.06 mg/mL).
An inoculum adjusted to a 0.5 McFarland standard was spot inoculated using a Cathra replicator. Plates were incubated at 37°C under anaerobic conditions for 48 h. Antibiotic-free plates were used as growth controls and to check for potential contamination. Growth on the antibiotic plates was compared with the growth on antibiotic-free plates. The lowest concentration of the tested antimicrobial agent that inhibited the visible bacterial growth was considered the MIC (mg/mL). All tests were performed in duplicate. The MIC was interpreted according to the EUCAST guidance on "When there are no breakpoints" (25). *Clostridium perfringens* CCUG 1795^{T} and *Streptococcus pneumoniae* CCUG 33638 were used as the reference strains for quality control of susceptibility testing in each test batch.

Results

The detection of ARGs and their correlation with MICs is summarized in Table 1. ARGs were identified in 72.7% (n=24/33) of isolates, with *lsaC* being the most prevalent gene (72.7%; n=24/33), detected across multiple isolates of all species. In addition to *lsaC,* four strains (12.1%) co-carried *msrD* and *mefA* (2 *G. pickettii*, 2 *G. greenwoodii*), both of which are associated with macrolide resistance.
Most strains (72.7%, n=24/33) exhibited low MICs for the tested antibiotics. MIC values for clindamycin ranged from 0.06 mg/L to 8 mg/L, with four strains displaying MICs ≥ 1 mg/L (2 *G. vaginalis*, 2 *G. swidsinskii*). Metronidazole MICs ranged from 0.06 to 64 mg/L; it is of note that two strains displayed the highest MICs (*G. pickettii* strain c10Ua_31, MIC=64 mg/L; *G. leopoldii* strain c18Ua_112, MIC=8 mg/L). MICs for benzylpenicillin were 1 mg/L in three strains (2 *G. vaginalis*, 1 *G. pickettii*). The highest vancomycin MIC was observed in *G. pickettii* isolate c10Ua_31 at 32 mg/mL, significantly exceeding MICs in other strains (0.06-1 mg/L). All but one strain (*G. vaginalis* c15Ub_3, MIC=2 mg/L) showed low MIC values for meropenem (0.06-0.5 mg/L). Moreover, all isolates showed low MICs for piperacillin-tazobactam and erythromycin. Notably, *G. pickettii* c10Ua_31 exhibited elevated MICs to multiple antibiotics.
A low correlation between genotype and phenotype was observed (Table 1). Despite the high prevalence of *lsaC*, only four isolates exhibited MICs-values above which therapy with clindamycin should be discouraged. Similarly, all four strains co-harboring *mefA* and *msrD* displayed low MICs to erythromycin. No ARGs or chromosomal point mutations were detected in strains displaying MIC-values above which therapy with benzylpenicillin, vancomycin or metronidazole should be discouraged.

Table 1. MICs and ARGs in different *Gardnerella* species from different origins.

Species	Health condition	Strain	Accession number	ARG (% identity)	MIC (mg/L)						
					CLI	PEN	VAN	PIP-TAZ	MEM	MTZ	ERY
G. vaginalis	HW	c11Ua_71	JARFVR000000000	*lsaC* (91.62)	0.5	0.06	0.5	1	0.06	0.25	2
		c15Ub_3	JAODXZ000000000	*lsaC* (91.62)	0.06	0.5	1	0.5	**2**	0.06	0.06
		c15Ub_5	JANUZA000000000	*lsaC* (91.62)	**2**	0.125	0.125	0.25	0.125	0.06	0.06
		c15Ub_15	JANUYY000000000	*lsaC* (91.62)	0.06	0.5	0.125	0.5	0.06	2	0.06
		c15VSa_39	JANUYX000000000	*lsaC* (91.62)	0.06	**1**	0.5	0.5	0.06	4	0.06
		c15VSb_4	JANUYW000000000	*lsaC* (91.62)	**1**	0.125	0.25	0.5	0.06	1	0.06
		c15VSb_14	JANUYU000000000	*lsaC* (91.62)	0.5	0.125	0.25	0.5	0.06	0.06	0.06
		c29Ua_32	JANUYN000000000	-	0.06	0.125	0.5	0.5	0.125	2	0.06
		c29Ub_7	JANUYM000000000	-	0.06	0.125	0.5	0.5	0.125	0.5	0.06
		c29VSa_19	JANUYL000000000	-	0.5	0.5	1	0.25	0.125	0.06	0.5
		c29VSb_3	JANUYK000000000	-	0.5	0.125	1	0.25	0.125	0.06	0.125
		c21Ua_54	JANUYR000000000	*lsaC* (91.69)	0.06	**1**	0.5	0.5	0.25	2	0.06
G. pickettii	HW	c10Ua_31	JANUYG010000000	*lsaC* (91.35)	0.5	**1**	**32**	1	0.25	**64**	0.5
		c12Ub_1	JANUZC000000000	*lsaC* (91.69)	0.06	0.5	1	1	0.5	0.5	0.06
		c23Ua_15	JANUYQ000000000	*lsaC* (91.55), *msrD* (100), *mefA* (99.92)	0.06	0.25	1	1	0.5	0.06	0.06
		c23Ua_38	JANUYP000000000	*lsaC* (91.55), *msrD* (100), *mefA* (99.92)	0.06	0.25	1	1	0.5	0.06	0.06
		c23Ub_4	JANUYO000000000	*lsaC* (91.69)	0.06	0.25	1	0.25	0.5	0.125	0.06
	OAB	c17Ua_112^{T}	GCF_029226345.1	*lsaC* (91.55), *msrD* (100), *mefA*	0.06	0.25	1	1	0.5	0.06	0.06

				(99.92)							
		c17Ua_118	JANUYT000000000	*lsaC* (91.69)	0.06	0.25	1	1	0.5	0.06	0.06
G. leopoldii	HW	c1Ub_6	JANUUT010000000	-	0.06	0.125	1	1	0.5	0.25	0.06
		c1Ub_40	JAOPFO000000000	-	0.5	0.125	0.5	2	0.125	0.125	0.125
		c31Ua_40	JANUYJ000000000	-	0.06	0.125	0.5	0.5	0.06	0.25	0.06
		c31VSa_14	JANUYI000000000	-	0.06	0.125	0.5	0.5	0.25	0.5	0.06
		c31VSa_35	JARFVP000000000	-	0.06	0.125	0.5	0.25	0.125	0.125	0.125
	OAB	c18Ua_110	JANUYS000000000	*lsaC* (91.62)	0.5	0.125	0.5	1	0.125	0.25	0.125
		c18Ua_112	JAPDSJ000000000	*lsaC* (91.62)	0.25	0.25	1	2	0.5	**8**	0.125
G. swidsinskii	HW	c10Ua_244	JARFVS000000000	*lsaC* (91.42)	**8**	0.06	0.5	1	0.06	0.25	0.125
		c15Ua_15	JAPDSI000000000	*lsaC* (91.42)	0.25	0.125	1	0.5	0.125	0.125	0.06
		c15Ua_23	JANUZB000000000	*lsaC* (91.42)	**1**	0.125	0.06	0.125	0.125	0.06	0.125
		c15Ub_13	JANUYZ000000000	*lsaC* (91.42)	0.25	0.06	0.125	0.125	0.125	0.06	0.125
		c15VSb_13	JANUYV000000000	*lsaC* (91.42)	ND	ND	0.5	0.5	0.06	0.06	0.125
		c15VSa_36	JARFVQ000000000	*lsaC* (91.42)	0.5	0.06	0.25	0.5	0.125	0.06	0.125
G. greenwoodii	HW	c31Ua_26^{T}	GCF_029207615.1	*lsaC* (91.49), *msrD* (100), *mefA* (99.92)	0.06	0.125	0.5	0.5	0.125	0.25	0.06

Abbreviations: CLI, clindamycin; ERY, erythromycin; HW, healthy woman; MEM, meropenem; MTZ, metronidazole; ND, not detected; OAB, woman diagnosed with overactive bladder; PEN, benzylpenicillin; PIP-TAZ, piperacillin-tazobactam; VAN, vancomycin. MICs presented in bold illustrate values above which therapy with the agent should be discouraged (25).

Discussion

The increasing recognition of *Gardnerella* spp. as a taxonomically diverse genus (1, 2) necessitates further characterization of AMR patterns and their clinical implications. In this study, we investigated the correlation between phenotypic MIC profiles and the presence of ARGs in *Gardnerella* isolates from healthy women and overactive bladder donors.

ARGs were identified in a significant proportion (72.7%) of isolates, with *lsaC* being the most prevalent (72.7%) and found in multiple species, which is consistent with previous studies reporting high *IsaC* detection in vaginal isolates (18, 21). Despite its association with resistance to lincosamides, streptogramin A, and pleuromutilins, only four isolates (12.2%) exhibited clindamycin MIC values exceeding the threshold at which treatment with these agents is discouraged (MIC > 0.5 mg/L), while other isolates carrying the gene had lower MICs. This finding is consistent with prior research showing that the presence of *lsaC* alone does not always correlate with high MIC values, suggesting potential regulatory influences on gene expression (18, 21).

The co-occurrence of *mefA* and *msrD* (12.1%) in *G. pickettii* and *G. greenwoodii* isolates raises concerns regarding potential macrolide resistance. However, all strains harboring these genes presented low MICs for erythromycin (MIC 0.06-2 mg/L). Similar findings have been reported, highlighting that *mefA* and *msrD* alone may not be sufficient to drive macrolide resistance without additional regulatory or structural modifications (18, 21).

Most isolates exhibited low MIC values to metronidazole, yet two strains (*G. pickettii* and *G. leopoldii*) displayed high MICs (MIC=64 mg/L and 8 mg/L, respectively). This finding is particularly relevant given the widespread use of metronidazole as a first-line treatment for BV, and prior studies have reported elevated MICs in *Gardnerella* (14, 15, 18). Zhang *et al*., (18) identified *yadH*, an ABC-type multidrug transport system gene, and *mcrA*, a 5-methylcytosine-specific restriction endonuclease gene, exclusively in metronidazole-high MIC isolates. However, these genes were not detected in our isolates, suggesting the involvement of alternative resistance mechanisms. Similarly, nitroimidazole resistance gene *nimJ*, associated with metronidazole resistance (20), *yadH*, *mcrA* (18), as well as alterations in OmpR and NadE proteins (26) were not identified in our strains. This absence suggests that resistance in *Gardnerella* may instead arise from alterations in oxidative stress response pathways, modifications in drug activation, or enhanced drug efflux systems. Further investigation is needed to elucidate the precise molecular mechanisms underlying metronidazole resistance in *Gardnerella* spp.

A high vancomycin MIC (32 mg/L) was observed in *G. pickettii* strain c10Ua_31, far exceeding previously reported MIC values (0.5-1 mg/L) (16, 18). The absence of known vancomycin resistance genes (*vanA, vanB, vanC*) suggests that intrinsic factors, such as modifications in

peptidoglycan biosynthesis, may contribute to this phenotype. Reduced transcription of *murF*, a gene involved in peptidoglycan precursor synthesis, have been linked to vancomycin resistance in other bacteria (27). Analysis of the MurF protein sequence in the *G. pickettii* isolate c10Ua_31 revealed a deletion of 16 amino acids, which was absent in all other strains. Future studies investigating this MurF deletion could help determine whether it is functionally linked to the observed vancomycin MIC elevation in *G. pickettii* isolate c10Ua_31.

One *G. pickettii* strain exhibited elevated MICs to multiple antibiotics (benzylpenicillin, vancomycin, and metronidazole), raising concerns about the potential for increased resistance in *Gardnerella* spp. Although multidrug-resistant *Gardnerella* isolates are uncommon, they have been detected in BV-associated biofilms, where resistance is often enhanced due to biofilm-mediated protection (13, 19). The emergence of multidrug-resistant strains highlights the importance of species-level identification and AST in clinical settings to ensure appropriate treatment strategies.

Conclusions

This study provides critical insights into AMR in *Gardnerella* spp., particularly in the context of species diversity and genotypic-phenotypic correlations. Despite the high prevalence of ARGs, MIC values remained low for most tested antibiotics, suggesting that *Gardnerella* spp. largely remain within susceptible ranges for current treatment options. However, the detection of isolates with high MIC values to metronidazole and vancomycin, and strain with elevated MICs to multiple antibiotics underscores the need for further investigation into resistance mechanisms.

Our findings highlight the need for continued AMR surveillance in *Gardnerella* spp., to better understand baseline resistance levels and potential reservoirs of AMR genes. Given the lack of established EUCAST epidemiological *cut-off* values for *Gardnerella*, future studies should evaluate alternative resistance mechanisms and prioritize phenotypic susceptibility testing to ensure accurate antimicrobial stewardship and effective treatment strategies.

Funding

This study 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. M.S. and F.G. were supported by FCT grants (SFRH/BD/05038/2020 and DL57/2016/CP1346/CT0034, respectively), T.G.R. by Unidade de Ciências Biomoleculares Aplicadas (UIDP/QUI/04378/2020), with the financial support of the FCT/ MCTES through national funds.

Contributor roles

M.S. performed all laboratory experiments and drafted the manuscript. F.G. contributed to study design, manuscript review and supervision. T.G.R. contributed to reviewing the drafted manuscript and supervision. L.P. contributed to project design, funding and manuscript review.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. 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.

2. Sousa M, Ksiezarek M, Perovic SU, Antunes-Lopes T, Grosso F, Ribeiro TG, Peixe L. 2023. *Gardnerella pickettii* sp. nov. (formerly *Gardnerella* genomic species 3) and *Gardnerella greenwoodii* sp. nov. (formerly *Gardnerella* genomic species 8) isolated from female urinary microbiome. *Int J Syst Evol Microbiol* 73:e005709.

3. Perovic SU, Ksiezarek M, Rocha J, Cappelli EA, Sousa M, Ribeiro TG, Grosso F, Peixe L. 2022. Urinary microbiome of reproductive-age asymptomatic European women. *Microbiol Spectr* 10(6), e0130822.

4. Hill JE, Albert AYK. 2019. Resolution and cooccurrence patterns of *Gardnerella leopoldii*, *G. swidsinskii*, *G. piotii*, and *G. vaginalis* within the vaginal microbiome. *Infect Immun* 87:e00532-19.

5. Hay PE, Lamont RF, Taylor-Robinson D, Morgan J, Ison C, Pearson J. 1994. Abnormal bacterial colonization of the genital tract and subsequent preterm delivery and late miscarriage. *BMJ* 308:295–298.

6. Menard JP, Mazouni C, Salem-Cherif I, Fenollar F, Raoult D, Boubli L, Gamerre M, Bretelle F. 2010. High vaginal concentrations of *Atopobium vaginae* and *Gardnerella vaginalis* in women undergoing preterm labor. *Obstet Gynecol* 115:134–140.

7. McKinnon LR, Achilles SL, Bradshaw CS, Burgener A, Crucitti T, Fredricks DN, Jaspan HB, Kaul R, Kaushic C, Klatt N, *et al.* 2019. The evolving facets of bacterial vaginosis: implications for HIV transmission. *AIDS Res Hum Retroviruses* 35:219–228.

8. López-Pintor JM, García-Fernández S, Ponce-Alonso M, Sánchez-Díaz AM, Ruiz-Garbajosa P, Morosini MI, Cantón R. 2021. Etiology and antimicrobial susceptibility profiles of anaerobic bacteria isolated from clinical samples in a university hospital in Madrid, Spain. *Anaerobe* 72:102446.

9. Veloo ACM, Tokman HB, Jean-Pierre H, Dumont Y, Jeverica S, Lienhard R, Novak A, Rodlogg A, Rotimi V, Wybo I, *et al.* 2020. Antimicrobial susceptibility profiles of anaerobic bacteria isolated from human clinical specimens within different European and surrounding countries: A joint ESGAI study. *Anaerobe* 61:102111.

10. Forbes JD, Kus JV, Patel SN. 2021. Antimicrobial susceptibility profiles of invasive isolates of anaerobic bacteria from a large Canadian reference laboratory: 2012–2019. *Anaerobe* 70, 102386.

11. Tripathi D, Saxena RK, Agarwal S. 2023. The pattern of antimicrobial resistance and the presence of the vaginolysin gene in *Gardnerella vaginalis* isolates from pregnant women. *Curr Med Res Pract* 13:172–175.

12. Tomusiak A, Strus M, Heczko PB. 2011. Lekowrailiwość szczepów *Gardnerella vaginalis* wyizolowanych z przypadlów bakteryjnej waginozy [Antibiotic resistance of *Gardnerella vaginalis* isolated from cases of bacterial vaginosis]. *Ginekologia polska*, 82(12), 900–904.

13. Qin H, Liu Y, Zhai Z, Xiao B. 2023. Biofilm-forming capacity and drug resistance of different *Gardnerella* subgroups associated with bacterial vaginosis. *Microorganisms* 11(9):2186.

14. Li T, Zhang Z, Wang F, He Y, Zong X, Bai H, Liu Z. 2020. Antimicrobial susceptibility testing of metronidazole and clindamycin against *Gardnerella vaginalis* in planktonic and biofilm formation. *Can J Infect Dis Med Microbiol* 1361825.

15. Schwebs T, Kieninger AK, Podpera Tisakova L, Oberbauer V, Berdaguer R, Mtshali A, Mzobe G, Rompalo A, Mindel A, Letsoalo M. 2024. Evaluation of metronidazole resistance of vaginal swab isolates from South African women treated for bacterial vaginosis. *Antibiotics (Basel)* 13(12), 1217.

16. Kharsany ABM, Hoosen AA, van den Ende J. 1993. Antimicrobial susceptibilities of *Gardnerella vaginalis*. *Antimicrob Agents Chemother* 37:2733–2735.

17. de Souza DMK, Diniz CG, Filho DSC, de Oliveira LMA, Coelho DM, Talha LS, Nascimento TC, Ferreira-Machado AB, Silva VL. 2016. Antimicrobial susceptibility and vaginolysin in *Gardnerella vaginalis* from healthy and bacterial vaginosis diagnosed women. *J Infect Dev Ctries* 10:913–919.

18. Zhang K, Lu M, Zhu X, Wang K, Jie X, Li T, Dong H, Li R, Zhang F, Gu L. 2022. Antibiotic resistance and pathogenicity assessment of various *Gardnerella* sp. strains in local China. *Front Microbiol* 13, 1009798.

19. Landlinger C, Oberbauer V, Tisakova LP, Schwebs T, Berdaguer R, Van Simaey L, Vaneechoutte M, Corsini L. 2022. Preclinical data on the *Gardnerella*-specific endolysin PM-477 indicate its potential to improve the treatment of bacterial vaginosis through enhanced biofilm removal and avoidance of resistance. *Antimicrob Agents Chemother* 66(5), e0231921.

20. Ruiz-Perez D, Coudray MS, Colbert B, Krupp K, Kumari H, Stebliankin V, Mathee K, Cook RL, Schwebke J, Narasimhan G, Madhivanan P. 2021. Effect of metronidazole on vaginal microbiota associated with asymptomatic bacterial vaginosis. *Access Microbiol* 3(5), 000226.

21. Sousa EG, Felice AG, Dominici FV, Jaiswal AK, Pedrosa MLC, Reis LP, Gomes LGR, Azevedo VAC, Soares SC. 2024. Unveiling resistance and virulence mechanisms under Darwinian positive selection for novel drug discovery for *Gardnerella vaginalis*. *Venereology* 3:120–135.

22. Ksiezarek M, Perovic SU, Rocha J, Grosso F, Peixe L. 2021. Long-term stability of the urogenital microbiota of asymptomatic European women. *BMC Microbiol* 21, 64.

23. Perovic SU, Ksiezarek M, Rocha J, Vale L, Silva C, Dinis P, Antunes TL, Peixe L. 2019. Time To Change Microbiological Approach To Overactive Bladder. *EAU19 - 34th Annual EAU Congress.* 18(1):e103, ISSN 1569-9056.

24. Olsson-Liljequist B, Nord CE. 1994. Methods for antimicrobial susceptibility testing of anaerobic bacteria. *Clin Infect Dis* 4:S293-6.

25. EUCAST. 2024. Guidance on when there are no breakpoints in breakpoint tables. https://www.eucast.org/guidance_no_breakpoints.pdf.

26. Schuyler JA, Mordechai E, Adelson ME, Gygax SE, Hilbert DW. 2015. Draft genome sequence of a metronidazole-resistant derivative of *Gardnerella vaginalis* strain ATCC 14019. *Genome Announc* 3:e01329-15.

27. Sobral RG, Jones AE, Des Etages SG, Dougherty TJ, Peitzsch RM, Gaasterland T, Ludovice AM, de Lencastre H, Tomasz A. 2007. Extensive and genome-wide changes in the transcription profile of *Staphylococcus aureus* induced by modulating the transcription of the cell wall synthesis gene *murF*. *J Bacteriol* 189:2376–2391.

4. General Conclusions

This study advances our understanding of *Gardnerella* species by exploring their genomic and functional diversity, revealing species- and strain-specific traits that underpin their ecological and clinical roles. Through an extensive pangenome analysis of 153 genomes, detailed virulence factor profiling, metabolic enzyme characterization, and antimicrobial resistance assessment, this thesis provides a multifaceted view of *Gardnerella*'s adaptability and pathogenicity, with implications for bacterial vaginosis (BV) and beyond.
Our pangenome analysis, encompassing six named species and eight genomic species, identified a conserved core of 505 genes alongside a vast accessory genome (90%), driving variability in virulence, metabolism, and resistance. Of note, 57% of public database genomes were misclassified despite 2019 taxonomic revisions, highlighting persistent identification challenges. Beyond taxonomy, we elucidated key virulence mechanisms (sialidases, vaginolysin), metabolic adaptations (glycogen degradation), and resistance profiles, integrating genomic and phenotypic data from symptomatic and asymptomatic women's isolates. These findings refine *Gardnerella*'s role in health and disease, offering a foundation for improved diagnostics and therapies.

The following sections outline the key findings of this study and their implications for understanding *Gardnerella*'s role in the female urogenital microbiome.

I - Advances in *Gardnerella* Taxonomy and Identification

Whole genome sequencing outperformed traditional 16S rRNA and *cpn60* sequencing methods, enabling precise delineation of *Gardnerella* species and the characterization of two novel taxa: *Gardnerella pickettii* (c17Ua_112^T) and *Gardnerella greenwoodii* (c31Ua_26^T). Their genetic proximity to *G. piotii* and *G. swidsinskii*, respectively, yet with distinct genomic signatures, underscores the inadequacy of subgroup- or clade-based classification. This work corrects misidentification in 57% of public genomes, emphasizing the need for WGS-based tools to clarify each species' ecological role, such as nutrient competition or host interactions, and clinical significance, areas warranting further study.

II - Virulence and Functional Characterization

We mapped sialidase diversity across 153 genomes, identifying NanH1 and NanH3 as ubiquitous, NanH2 as dysbiosis-linked, and NanH4 as a newly active intracellular enzyme. Structural studies suggest NanH2 and NanH3 dominate extracellular mucin degradation,

enhancing pathogenicity, while NanH1 and NanH4's roles remain less clear. The inhibition of NanH3 by DANA and Zanamivir suggests potential therapeutic strategies for targeting *Gardnerella*-associated virulence factors. Vaginolysin (VLY) was detected in 89.1% of strains, with a novel Type 3B variant being identified. Five distinct genetic platforms were associated with *vly*, revealing evolutionary complexity and potential regulatory links to transport genes. These insights pinpoint virulence targets and call for studies on VLY expression and host-specific effects.

III - Metabolic Insights and Adaptation Strategies

Carbohydrate metabolism emerged as a key adaptative trait, with *G. vaginalis* and *G. pickettii* excelling in glycogen utilization. We report the first identification of neopullulanase and the presence of GH13_20 + CBM34 in 51.2% of isolates, alongside glycogen debranching enzymes (GH13_11 + CBM48) across all species and α-amylase domains (GH13_32) with varying carbohydrate-binding modules in most strains. These enzymes enhance nutrient acquisition from glycogen, potentially shaping vaginal microbiome dynamics through competition or resource sharing. Further profiling is needed to quantify these metabolic advantages across species and strains.

IV - Antimicrobial Resistance and Clinical Implications

Six antibiotic resistance genes (*lsaC, mefA, msrD, aph(3')-Ia, tetM,* and *ermX*) were identified, with *lsaC* being the most prevalent, yet phenotypic MIC values remained low for most antibiotics. Exceptions - high metronidazole MICs in *G. pickettii* (64 mg/L) and G. *leopoldii* (8 mg/L), without known ARGs, together with a novel deletion in the gene coding for MurF in *G. pickettii* (vancomycin MIC 32 mg/L) - suggest undiscovered resistance mechanisms. The poor genotype-phenotype correlation complicates treatment, underscoring the urgency of establishing *Gardnerella*-specific breakpoints and surveillance protocols.

V - Future Directions

This study provides the basis for targeted research: 1) validating sialidase (e.g., Zanamivir) and potential VLY inhibitors *in vivo*, 2) quantifying glycogen metabolism's ecological impact via co-culture models, 3) investigating novel resistance mechanisms (e.g., MurF variants) with functional genomics, and 4) refining WGS-based diagnostics to replace outdated methods. Linking these findings to BV severity, host genetics, and microbiome interactions will enhance therapeutic precision, addressing gaps in current clinical practice.

5. References

1. Hay PE, Lamont RF, Taylor-Robinson D, Morgan J, Ison C, Pearson J. Abnormal bacterial colonisation of the genital tract and subsequent preterm delivery and late miscarage. BMJ (Clinical research ed.). 1994 308(6924), 295–298.

2. Menard JP, Mazouni C, Salem-Cherif I, Fenollar F, Raoult D, Boubli L, *et al.* High Vaginal Concentrations of *Atopobium vaginae* and *Gardnerella vaginalis* in Women Undergoing Preterm Labor. Obstet Gynecol. 2010 115(1).

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

4. 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 Sep 1;213(3):347.e1-347.e11.

5. Antunes-Lopes T, Vale L, Coelho AM, Silva C, Rieken M, Geavlete B, *et al.* The Role of Urinary Microbiota in Lower Urinary Tract Dysfunction: A Systematic Review. Eur Urol Focus. 2020 6(2), 361–369.

6. Leopold S. Heretofore undescribed organism isolated from the genitourinary system. US Armed Forces Med J. 1953 4(2), 263–266.

7. Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, *et al.* Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. J Bacteriol. 2012 Aug;194(15):3922–37.

8. Schellenberg JJ, Patterson MH, Hill JE. *Gardnerella vaginalis* diversity and ecology in relation to vaginal symptoms. Res Microbiol. 2017 Nov 1;168(9–10):837–44.

9. Cornejo OE, Hickey RJ, Suzuki H, Forney LJ. Focusing the diversity of *Gardnerella vaginalis* through the lens of ecotypes. Evol Appl. 2018 Mar 1;11(3):312–24.

10. 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 Mar 1;69(3):679–87.

11. Potter RF, Burnham CAD, Dantas G. *In silico* analysis of *Gardnerella* genomospecies detected in the setting of bacterial vaginosis. Clin Chem. 2019 65(11):1375–87.

12. Putonti C, Thomas-White K, Crum E, Hilt EE, Price TK, Wolfe AJ. Genome Investigation of Urinary *Gardnerella* Strains and Their Relationship to Isolates of the Vaginal Microbiota. mSphere. 2021 Jun 30;6(3).

13. Harwich MD, Alves JM, Buck GA, Strauss Iii JF, Patterson JL, Oki AT, *et al.* Drawing the line between commensal and pathogenic *Gardnerella vaginalis* through genome analysis and virulence studies. BMC Genomics. 2010 11,375.

14. Hill JE, Albert AYK. Resolution and Cooccurrence Patterns of *Gardnerella leopoldii, G. swidsinskii, G. piotii,* and *G. vaginalis* within the Vaginal Microbiome. Infect Immun. 2019 Nov 18;87(12):e00532-19

15. 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 Dec 1;21(1).

16. 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 Dec 1;9(1).

17. Brubaker L, Putonti C, Dong Q, Wolfe AJ. The human urobiome. Mamm Genome. 2021. 32(4), 232–238.

18. Berg G, Rybakova D, Fischer D, Cernava T, Vergès MCC, Charles T, *et al.* Microbiome definition re-visited: old concepts and new challenges. Microbiome. 2020 8,103.

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

20. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, *et al.* Structure, function and diversity of the healthy human microbiome. Nature. 2012 Jun 14;486(7402):207–14.

21. 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(3):871–6.

22. Lewis DA, Brown R, Williams J, White P, Jacobson SK, Marchesi JR, *et al.* The human urinary microbiome; bacterial DNA in voided urine of asymptomatic adults. Front Cell Infect Microbiol. 2013 Aug 15:3:41.

23. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, FitzGerald MP, *et al.* Evidence of uncultivated bacteria in the adult female bladder. J Clin Microbiol. 2012 Apr;50(4):1376–83.

24. Aragón IM, Herrera-Imbroda B, Queipo-Ortuño MI, Castillo E, Del Moral JSG, Gómez-Millán J, *et al.* The Urinary Tract Microbiome in Health and Disease. Eur Urol Focus. 2018 Jan;4(1):128-138.

25. Price TK, Wolff B, Halverson T, Limeira R, Brubaker L, Dong Q, *et al.* Temporal Dynamics of the Adult Female Lower Urinary Tract Microbiota. mBio. 2020 Apr 21;11(2):e00475-20.

26. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, *et al.* The female urinary microbiome: A comparison of women with and without urgency urinary incontinence. mBio. 2014 Jul 8;5(4).

27. 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 Aug;204(2):303–9.

28. 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 Sep 1;228:126–9.

29. 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 Jan 1;127(2):193–201.

30. Lewis FMT, Bernstein KT, Aral SO. Vaginal microbiome and its relationship to behavior, sexual health, and sexually transmitted diseases. Obstet Gynecol. 2017 129(4):643–54.

31. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, *et al.* Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011 Mar 15;108(suppl. 1):4680–7.

32. Liu F, Zhou Y, Zhu L, Wang Z, Ma LK, He Y, *et al.* Comparative metagenomic analysis of the vaginal microbiome in healthy women. Synth Syst Biotechnol. 2021 Jun 1;6(2):77–84.

33. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, *et al.* Temporal dynamics of the human vaginal microbiota. Sci Transl Med. 2012 May 2;4(132).

34. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, *et al.* Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci U S A. 2015 Sep 1;112(35):11060–5.

35. France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. Microbiome. 2020 Dec 1;8(1).

36. Komesu YM, Dinwiddie DL, Richter HE, Lukacz ES, Sung VW, Siddiqui NY, *et al.* Defining the relationship between vaginal and urinary microbiomes. Am J Obstet Gynecol. 2020 Feb;222(2):154.e1-154.e10.

37. Gottschick C, Deng ZL, Vital M, Masur C, Abels C, Pieper DH, *et al.* The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. Microbiome. 2017 Aug 14;5(1):99.

38. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system, Actinobacteria classis nov. Int J Syst Bacteriol. 1997;47(2):479–91.

39. Gardner HL, Dukes CD. *Haemophilus vaginalis* vaginitis: A newly defined specific infection previously classified “nonspecific” vaginitis. Am J Obstet Gynecol, 1955 69(5):962–76.

40. Zinnemann K, Turner GC. The taxonomic position of *Haemophilus vaginalis* (*Corynebacterium vaginale*). J Pathol Bacteriol. 1963 Jan 1;85(1):213–9.

41. Reyn A, Birch-Andersen A, Lapage SP. An Electron Microscope Study of Thin Sections of *Haemophilus Vaginalis* (Gardner And Dukes) and some Possibly Related Species. Can J Microbiol. 1966 Dec 1;12(6):1125–36.

42. Harper A N D JJ, Davis GHG. Cell Wall Analysis of *Gardnerella vaginalis* (*Haemophilus vaginalis*). Int J Syst Evol Microbiol. 1982 32(1).

43. Sadhu K, Domingue PAG, Chow AW, Nelligan J, Cheng N, Costerton JW. *Gardnerella vaginalis* has a gram-positive cell-wall ultrastructure and lacks classical cell-wall lipopolysaccharide. J Med Microbiol. 1989 29(3):229–35.

44. Turovskiy Y, Sutyak Noll K, Chikindas ML. The aetiology of bacterial vaginosis. J Appl Microbiol. 2011 May;110(5):1105-28.

45. Gardner HL. *Haemophilus vaginalis:* vaginitis after twenty-five years. Am J Obstet Gynecol 1980 Jun 1;137(3):385–91.

46. Greenwood JR, Pickett MJ. Transfer of *Haemophilus vaginalis* Gardner and Dukes to a New Genus, *Gardnerella: G. vaginalis* (Gardner and Dukes) comb. nov. Int J Syst Bacteriol. 1980 Jan 1;30(1):170–8.

47. Moss CW, Dunkelberg WE. Volatile and Cellular Fatty Acids of *Haemophilus vaginalis.* J Bacteriol. 1969 Oct;100(1):544–546.

48. Bulavaitė A, Maier T, Pleckaityte M. Discrimination of *Gardnerella* species by combining MALDI-TOF protein profile, chaperonin *cpn60* sequences and phenotypic characteristics. Pathogens. 2021 Mar 1;10(3):1–9.

49. Colwell RR, Mandel M, Anderson MD. Adansonian Analysis and Deoxyribonucleic Acid Base Composition of Some Gram-Negative Bacteria. J Bacteriol. 1964 Jun;87(6):1412–1422.

50. Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci. 2010;2(1):117–34.

51. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, *et al.* Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol. 2018 Jan 1;68(1):461–6.

52. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009 Nov 10;106(45):19126-31.

53. Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO. Application of tetranucleotide frequencies for the assignment of genomic fragments. Environ Microbiol. 2004 Sep 1;6(9):938–47.

54. Hou TY, Chiang-Ni C, Teng SH. Current status of MALDI-TOF mass spectrometry in clinical microbiology. J Food Drug Anal. 2019 Apr;27(2):404-414.

55. Rosca AS, Castro J, Sousa LGV, Cerca N. *Gardnerella* and vaginal health: The truth is out there. FEMS Microbiol Rev. 2020 Jan 1;44(1):73-105.

56. Vaneechoutte M. The human vaginal microbial community. Res Microbiol. 2017 Nov 1;168(9–10):811–25.

57. Tachedjian G, Aldunate M, Bradshaw CS, Cone RA. The role of lactic acid production by probiotic *Lactobacillus* species in vaginal health. Res Microbiol. 2017 Nov 1;168(9–10):782–92.

58. Fuochi V, Cardile V, Petronio Petronio G, Furneri PM. Biological properties and production of bacteriocins-like-inhibitory substances by *Lactobacillus* sp. strains from human vagina. J Appl Microbiol 2019 May 1;126(5):1541–50.

59. Vodstrcil LA, Twin J, Garland SM, Fairley CK, Hocking JS, Law MG, *et al.* The influence of sexual activity on the vaginal microbiota and *Gardnerella vaginalis* clade diversity in young women. PLoS One. 2017 Feb 24;12(2):e0171856.

60. Berman HL, Goltsman DSA, Anderson M, Relman DA, Callahan BJ. *Gardnerella* diversity and ecology in pregnancy and preterm birth . mSystems. 2024 Jun 18;9(6).

61. Lebeer S, Ahannach S, Gehrmann T, Wittouck S, Eilers T, Oerlemans E, *et al.* A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. Nat Microbiol. 2023 Nov 1;8(11):2183–95.

62. Onderdonk AB, Delaney ML, Fichorova RN. The human microbiome during bacterial vaginosis. Clin Microbiol Rev. 2016 Feb 10;29(2):223–38.

63. Perovic SU, Ksiezarek M, Rocha J, Cappelli EA, Sousa M, Ribeiro TG, *et al.* Urinary Microbiome of Reproductive-Age Asymptomatic European Women. Microbiol Spectr. 2022 Dec 21;10(6).

64. Albert AYK, Chaban B, Wagner EC, Schellenberg JJ, Links MG, Van Schalkwyk J, *et al.* A study of the vaginal microbiome in healthy Canadian women utilizing *cpn60*-based molecular profiling reveals distinct *Gardnerella* subgroup community state types. PLoS One. 2015 Aug 12;10(8).

65. Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. BMC Microbiol. 2011 11, 244.

66. Wolff BJ, Price TK, Joyce CJ, Wolfe AJ, Mueller ER. Oral probiotics and the female urinary microbiome: a double-blinded randomized placebo-controlled trial. Int Urol Nephrol. 2019 Dec 1;51(12):2149–59.

67. Pohl HG, Groah SL, Pérez-Losada M, Ljungberg I, Sprague BM, Chandal N, *et al.* The urine microbiome of healthy men and women differs by urine collection method. Int Neurourol J. 2020 24(1):41–51.

68. Hooks KB, O'Malley MA. Dysbiosis and its discontents. mBio. 2017 Sep 1;8(5).

69. Chen X, Lu Y, Chen T, Li R. The Female Vaginal Microbiome in Health and Bacterial Vaginosis. Front Cell Infect Microbiol. 2021 Apr 7:11:631972.

70. Murray L, Halpin J, Casserly B, O'Connell NH, Scanlon T. A pyo-hydropneumothorax with sepsis, secondary to *Gardnerella vaginalis* infection in a post-partum female. Respir Med Case Rep. 2019 Jan 1;26:189–92.

71. Datsomor EA, Zubach O, Prykuda N, Zinchuk A. A case report of septic gardnerellosis. IDCases. 2021 Jan 1;24.

72. Abdulla H, Chhina A. A Rare Case of *Gardnerella vaginalis* Causing Severe Sepsis from Ventilator Associated Pneumonia in a Patient with Traumatic Brain Injury. Am J Respir Crit Care Med 2022 205:A2262.

73. Hamod D. A New Aspect of Neonatal Infection to *Gardnerella vaginalis*. J Pediatr Neonatal Care. 2017 Feb 24;6(3).

74. Balks J, Grumaz S, Mazzitelli S, Neder U, Lemloh L, Melaku T, *et al*. Microbial cell-free DNA-sequencing as an addition to conventional diagnostics in neonatal sepsis. Pediatr Res. 2025 Feb;97(2):614-624.

75. Akamine CM, Chou A, Tavakoli-Tabasi S, Musher DM. *Gardnerella vaginalis* Bacteremia in Male Patients: A Case Series and Review of the Literature. Open Forum Infect Dis. 2022 Apr 8;9(10):ofac176.

76. Abbasian B, Shair A, O'gorman DB, Pena-Diaz AM, Brennan L, Engelbrecht K, *et al.* Potential Role of Extracellular ATP Released by Bacteria in Bladder Infection and Contractility. mSphere. 2019 Sep 4;4(5):e00439-19.

77. Bartoli S, Aguzzi G, Tarricone R. Impact on Quality of Life of Urinary Incontinence and Overactive Bladder: A Systematic Literature Review. Urology. 2010 Mar 1;75(3):491–500.

78. Shapiro KK, Brucker BM. Treatment of overactive bladder in men: Is it really different? Neurourol Urodyn 2022 Nov 1;41(8):1975–82.

79. Blaivas JG, Panagopoulos G, Weiss JP, Somaroo C. Validation of the Overactive Bladder Symptom Score. J Urol. 2007 Aug;178(2):543-7.

80. Homma Y, Yoshida M, Seki N, Yokoyama O, Kakizaki H, Gotoh M, *et al.* Symptom assessment tool for overactive bladder syndrome-overactive bladder symptom score. Urology. 2006 Aug;68(2):318–23.

81. Shim JS, Kim JH, Choi H, Park JY, Bae JH. Diagnostic tool for assessing overactive bladder symptoms: Could the international prostate symptom storage subscore replace the overactive bladder symptom score? Int Neurourol J. 2016 20(3):209–13.

82. Sun Q, Li L, Zhou H, Wu Y, Gao Y, Wu B, Qiu Y, *et al.* The detection of urinary viruses is associated with aggravated symptoms and altered bacteriome in female with overactive bladder. Front. Microbiol. 2022 13:984234.

83. Zhou Z, Qiu Y, Li K, Sun Q, Xie M, Huang P, *et al.* Unraveling the impact of *Lactobacillus* spp. and other urinary microorganisms on the efficacy of mirabegron in female patients with overactive bladder. Front Cell Infect Microbiol. 2022 Nov 14;12.

84. 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 Jul 27;6.

85. Carnes MU, Siddiqui NY, Karstens L, Gantz MG, Dinwiddie DL, Sung VW, *et al.* Urinary microbiome community types associated with urinary incontinence severity in women. Am J Obstet Gynecol. 2024 Mar 1;230(3):344.e1-344.e20.

86. Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol 1991 Feb;29(2):297-301.

87. Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, *et al.* The clinical urine culture: Enhanced techniques improve detection of clinically relevant microorganisms. J Clin Microbiol. 2016 May 1;54(5):1216–22.

88. Yoo JJ, Song JS, Kim W Bin, Yun J, Shin HB, Jang MA, *et al. Gardnerella vaginalis* in Recurrent Urinary Tract Infection Is Associated with Dysbiosis of the Bladder Microbiome. J Clin Med. 2022 May 1;11(9).

89. O'brien VP, Joens MS, Lewis AL, Gilbert NM. Recurrent *Escherichia coli* Urinary Tract Infection Triggered by *Gardnerella vaginalis* Bladder Exposure in Mice. J Vis Exp. 2020 Dec 4:(166):10.3791/61967.

90. Robinson LS, Schwebke J, Lewis WG, Lewis AL. Identification and characterization of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium *Gardnerella vaginalis*. JBC. 2019 Apr 5;294(14):5230–45.

91. Donders GGG, Bellen G, Grinceviciene S, Ruban K, Vieira-Baptista P. Aerobic vaginitis: no longer a stranger. Res Microbiol. 2017 Nov 1;168(9–10):845–58.

92. Achkar JM, Fries BC. *Candida* infections of the genitourinary tract. Clin Microbiol Rev. 2010 Apr;23(2):253-73.

93. Edwards T, Burke P, Smalley H, Hobbs G. *Trichomonas vaginalis:* Clinical relevance, pathogenicity and diagnosis. Crit Rev Microbiol. 2016 May;42(3):406-17.

94. Condori-Catachura S, Ahannach S, Ticlla M, Kenfack J, Livo E, Anukam KC, *et al.* Diversity in women and their vaginal microbiota. Trends Microbiol 2025 Jan 20:S0966-842X(24)00328-7.

95. Muzny CA, Schwebke JR. Asymptomatic Bacterial Vaginosis: to Treat or Not to Treat? Curr Infect Dis Rep. 2020 Dec;22(12):32.

96. Peebles K, Velloza J, Balkus JE, McClelland RS, Barnabas R V. High Global Burden and Costs of Bacterial Vaginosis: A Systematic Review and Meta-Analysis. Sex Transm Dis 2019 46(5).

97. Redelinghuys MJ, Geldenhuys J, Jung H, Kock MM. Bacterial Vaginosis: Current Diagnostic Avenues and Future Opportunities. Front Cell Infect Microbiol. 2020 Aug 11:10:354.

98. 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 Journal. 2019 May 1;13(5):1306–17.

99. Hardy L, Jespers V, Dahchour N, Mwambarangwe L, Musengamana V, Vaneechoutte M*, et al.* Unravelling the bacterial vaginosis-associated biofilm: A multiplex

Gardnerella vaginalis and *Atopobium vaginae* fluorescence *in situ* hybridization assay using peptide nucleic acid probes. PLoS One. 2015 Aug 25;10(8).

100. Morrill S, Gilbert NM, Lewis AL. *Gardnerella vaginalis* as a Cause of Bacterial Vaginosis: Appraisal of the Evidence From *in vivo* Models. Front Cell Infect Microbiol. 2020 Apr 24:10:168.

101. Munch MM, Strenk SM, Srinivasan S, Fiedler TL, Proll S, Fredricks DN. *Gardnerella* Species and Their Association With Bacterial Vaginosis. J Infect Dis. 2024 Jul 15;230(1):e171–81.

102. Himschoot L, Mulinganya G, Rogier T, Bisimwa G, Kampara F, Kujirakwinja Y, *et al.* Prevalence and clinical correlates of *Gardnerella* spp., *Fannyhessea vaginae, Lactobacillus crispatus* and *L. iners* in pregnant women in Bukavu, Democratic Republic of the Congo. Front Cell Infect Microbiol. 2025 14, 1514884.

103. Castro J, Rosca AS, Cools P, Vaneechoutte M, Cerca N. *Gardnerella vaginalis* Enhances *Atopobium vaginae* Viability in an *in vitro* Model. Front Cell Infect Microbiol. 2020 Mar 4;10.

104. Amsel R, Totten PA, Spiegel CA, Chen KCS, Eschenbach D, Holmes KK. Nonspecific vaginitis: Diagnostic criteria and microbial and epidemiologic associations. Am J Med. 1983 Jan 1;74(1):14–22.

105. Blankenstein T, Lytton SD, Leidl B, Atweh E, Friese K, Mylonas I. Point-of-care (POC) diagnosis of bacterial vaginosis (BV) using VGTest^{TM} ion mobility spectrometry (IMS) in a routine ambulatory care gynecology clinic. Arch Gynecol Obstet. 2015 Aug 10;292(2):355–62.

106. Mylonas I, Bergauer F. Diagnosis of Vaginal Discharge by Wet Mount Microscopy: A Simple and Underrated Method. Obstet Gynecol Sur. 2011 Jun;66(6):359-68.

107. Madhivanan P, Krupp K, Li T, Ravi K, Selezneva J, Srinivas V, *et al.* Performance of bvblue rapid test in detecting bacterial vaginosis among women in mysore, India. Infect Dis Obstet Gynecol. 2014 Jan 12;2014:908313.

108. Cartwright CP, Lembke BD, Ramachandran K, Body BA, Nye MB, Rivers CA, *et al.* Comparison of nucleic acid amplification assays with BD affirm VPIII for diagnosis of vaginitis in symptomatic women. J Clin Microbiol. 2013 Nov;51(11):3694–9.

109. Abou Chacra L, Drouet H, Ly C, Bretelle F, Fenollar F. Evaluation of Various Diagnostic Strategies for Bacterial Vaginosis, Including a New Approach Based on MALDI-TOF Mass Spectrometry. Microorganisms. 2024 Jan 1;12(1).

110. Saadaoui M, Singh P, Ortashi O, Al Khodor S. Role of the vaginal microbiome in miscarriage: exploring the relationship. Cell Infect Microbiol. 2023 Sep 13:13:1232825.

111. Schuster HJ, Bos AM, Himschoot L, van Eekelen R, Matamoros SPF, de Boer MA, *et al.* Vaginal microbiota and spontaneous preterm birth in pregnant women at high risk of recurrence. Heliyon. 2024 May 30;10(10).

112. Wee BA, Thomas M, Sweeney EL, Frentiu FD, Samios M, Ravel J, *et al.* A retrospective pilot study to determine whether the reproductive tract microbiota differs between women with a history of infertility and fertile women. Aust N Z J Obstet Gynaecol. 2018 Jun;58(3):341-348.

113. Tabatabaei N, Eren AM, Barreiro LB, Yotova V, Dumaine A, Allard C, Fraser WD. Vaginal microbiome in early pregnancy and subsequent risk of spontaneous preterm birth: a case-control study. BJOG. 2019 126(3), 349–358.

114. Xu L, Huang L, Lian C, Xue H, Lu Y, Chen X, *et al.* Vaginal microbiota diversity of patients with embryonic miscarriage by using 16S rDNA high-throughput sequencing. Int J Genomics. 2020 Nov 20;2020:1764959.

115. Al-Memar M, Bobdiwala S, Fourie H, Mannino R, Lee YS, Smith A, *et al.* The association between vaginal bacterial composition and miscarriage: a nested case–control study. BJOG 2020 Jan 1;127(2):264–74.

116. Brotman RM, Klebanoff MA, Nansel TR, Yu KF, Andrews WW, Zhang J, *et al.* Bacterial vaginosis assessed by Gram stain and diminished colonization resistance to incident gonococcal, chlamydial, and trichomonal genital infection. J Infect Dis. 2010 Dec 15;202(12):1907-15.

117. Cherpes TL, Meyn LA, Krohn MA, Lurie JG, Hillier SL. Association between Acquisition of Herpes Simplex Virus Type 2 in Women and Bacterial Vaginosis. Clinical Infectious Diseases. 2003 Aug 1;37(3):319–25.

118. Gallo MF, Macaluso M, Warner L, Fleenor ME, Hook EW, Brill I, *et al.* Bacterial Vaginosis, Gonorrhea, and Chlamydial Infection Among Women Attending a Sexually Transmitted Disease Clinic: A Longitudinal Analysis of Possible Causal Links. Ann Epidemiol. 2012 Mar;22(3):213–20.

119. McKinnon LR, Achilles SL, Bradshaw CS, Burgener A, Crucitti T, Fredricks DN, *et al.* The Evolving Facets of Bacterial Vaginosis: Implications for HIV Transmission. AIDS Res Hum Retroviruses. 2019 Mar;35(3):219-228.

120. Gosmann C, Anahtar MN, Handley SA, Farcasanu M, Abu-Ali G, Bowman BA, *et al. Lactobacillus*-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women. Immunity. 2017 Jan 17;46(1):29–37.

121. Usyk M, Zolnik CP, Castle PE, Porras C, Herrero R, Gradissimo A, *et al.* Cervicovaginal microbiome and natural history of HPV in a longitudinal study. PLoS Pathog. 2020 Mar 26;16(3):e1008376.

122. Kovachev SM. Cervical cancer and vaginal microbiota changes. Arch Microbiol. 2020 Mar 1;202(2):323–7.

123. Mitra A, MacIntyre DA, Marchesi JR, Lee YS, Bennett PR, Kyrgiou M. The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: What do we know and where are we going next? Microbiome. 2016 Nov 1;4(1):58.

124. Gillet E, Meys JFA, Verstraelen H, Bosire C, De Sutter P, Temmerman M, *et al.* Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: A meta-analysis. BMC Infect Dis. 2011 Jan 11;11.

125. Shvartsman E, Perciani CT, Richmond MEI, Russell JNH, Tough RH, Vancuren SJ, *et al. Gardnerella* subgroup dominant microbiomes are associated with divergent cervicovaginal immune responses in a longitudinal cohort of Kenyan women. Front Immunol. 2023 Jan 16;13.

126. Mtshali A, Ngcapu S, Mindel A, Garrett N, Liebenberg L. HIV susceptibility in women: The roles of genital inflammation, sexually transmitted infections and the genital microbiome. J Reprod Immunol. 2021 Jun:145:103291.

127. Masson L, Barnabas S, Deese J, Lennard K, Dabee S, Gamieldien H, *et al.* Inflammatory cytokine biomarkers of asymptomatic sexually transmitted infections and vaginal dysbiosis: a multicentre validation study. Sex Transm Infect. 2019 Feb 1;95(1):5.

128. Castro J, Alves P, Sousa C, Cereija T, França Â, Jefferson KK, *et al.* Using an *in-vitro* biofilm model to assess the virulence potential of Bacterial Vaginosis or non-Bacterial Vaginosis *Gardnerella vaginalis* isolates. Sci Rep. 2015 Jun 26;5.

129. Price JT, Vwalika B, France M, Ravel J, Ma B, Mwape H, *et al.* HIV-associated vaginal microbiome and inflammation predict spontaneous preterm birth in Zambia. Sci Rep. 2022 Dec 1;12(1).

130. Martin Jr. HL, Richardson BA, Nyange PM, Lavreys L, Hillier SL, Chohan B, *et al.* Vaginal Lactobacilli, Microbial Flora, and Risk of Human Immunodeficiency Virus Type 1 and Sexually Transmitted Disease Acquisition. J Infect Dis. 1999 Dec 1;180(6):1863–8.

131. Bohr LL, Mortimer TD, Pepperell CS. Lateral Gene Transfer Shapes Diversity of *Gardnerella* spp. Front Cell Infect Microbiol. 2020 Jun 23;10.

132. Tarracchini C, Lugli GA, Mancabelli L, Milani C, Turroni F, Ventura M. Assessing the Genomic Variability of *Gardnerella vaginalis* through Comparative Genomic Analyses: Evolutionary and Ecological Implications. Appl Environ Microbiol. 2020 Jan 1;87(1):1–16.

133. Zhang K, Lu M, Zhu X, Wang K, Jie X, Li T, *et al.* Antibiotic resistance and pathogenicity assessment of various *Gardnerella* sp. strains in local China. Front Microbiol. 2022 Sep 26;13.

134. Khan S, Vancuren SJ, Hill JE. A Generalist Lifestyle Allows Rare *Gardnerella* spp. to Persist at Low Levels in the Vaginal Microbiome. Microb Ecol. 2021 82, 1048–1060.

135. Felice AG, Sousa EG, Dominici FV, Azevedo VA de C, Soares S de C. Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis:* An *In Silico* Approach. Venereology. 2023 Sep 30;2(4):132–46.

136. Dillard LR, Glass EM, Lewis AL, Thomas-White K, Papin JA. Metabolic Network Models of the *Gardnerella* Pangenome Identify Key Interactions with the Vaginal Environment. mSystems. 2023 Feb 23;8(1).

137. Yeoman CJ, Yildirim S, Thomas SM, Durkin AS, Torralba M, Sutton G, *et al.* Comparative genomics of *Gardnerella vaginalis* strains reveals substantial differences in metabolic and virulence potential. PLoS One. 2010 Aug 26;5(8):e12411.

138. Ruiz-Hernández UE, Pelcastre-Rodriguez LI, Cabrero-Martínez OA, Hernández-Cortez C, Castro-Escarpulli G. Analysis of CRISPR-Cas systems in *Gardnerella* suggests its potential role in the mechanisms of bacterial vaginosis. Comput Biol Chem. 2020 Dec 1;89.

139. Castro J, Jefferson KK, Cerca N. Genetic Heterogeneity and Taxonomic Diversity among *Gardnerella* Species. Trends Microbiol. 2020 Mar 1;28(3):202–11.

140. Udayalaxmi J, Bhat GK, Kotigadde S. Biotypes and virulence factors of *Gardnerella vaginalis* isolated from cases of bacterial vaginosis. Indian J Med Microbiol. 2011 29(2):165–8.

141. Mohammadzadeh R, Kalani BS, Kashanian M, Oshaghi M, Amirmozafari N. Prevalence of vaginolysin, sialidase and phospholipase genes in *Gardnerella vaginalis* isolates between bacterial vaginosis and healthy individuals. Med J Islam Repub Iran. 2019 33(1).

142. Ma X, Wang X, Ye S, Liu J, Yuan H, Wang N. Biofilm and pathogenic factor analysis of *Gardnerella vaginalis* associated with bacterial vaginosis in Northeast China. Front Microbiol. 2022 Dec 22;13.

143. Pelayo P, Hussain FA, Werlang CA, Wu CM, Woolston BM, Xiang CM, *et al. Prevotella* are major contributors of sialidases in the human vaginal microbiome. Proc Natl Acad Sci U S A. 2024 Sep 3;121(36):e2400341121.

144. Lewis WG, Robinson LS, Gilbert NM, Perry JC, Lewis AL. Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted actinobacterium *Gardnerella* vaginalis. JBC. 2013 Apr 26;288(17):12067–79.

145. Gilbert NM, Lewis WG, Li G, Sojka DK, Lubin JB, Lewis AL. *Gardnerella vaginalis* and *Prevotella bivia* Trigger Distinct and Overlapping Phenotypes in a Mouse Model of Bacterial Vaginosis. Journal of Infectious Diseases. 2019 Aug 30;220(7):1099–108.

146. von Nicolai H, Hammann R, Salehnia S, Zilliken F. A newly discovered sialidase from *Gardnerella vaginalis*. Zentralbl Bakteriol Mikrobiol Hyg A. 1984 Oct;258(1):20-6.

147. Kurukulasuriya SP, Patterson MH, Hill JE. Slipped-Strand Mispairing in the Gene Encoding Sialidase NanH3 in *Gardnerella* spp. Infect Immun. 2021 Feb 16;89(3):e00583-20.

148. Govinden G, Parker JL, Naylor KL, Frey AM, Anumba DOC, Stafford GP. Inhibition of sialidase activity and cellular invasion by the bacterial vaginosis pathogen *Gardnerella vaginalis*. Arch Microbiol. 2018 Sep 1;200(7):1129–33.

149. Lewis WG, Robinson LS, Perry J, Bick JL, Peipert JF, Allsworth JE, *et al.* Hydrolysis of secreted sialoglycoprotein immunoglobulin a (IgA) in *ex vivo* and biochemical models of bacterial vaginosis. JBC. 2012 Jan 13;287(3):2079–89.

150. Hata K, Koseki K, Yamaguchi K, Moriya S, Suzuki Y, Yingsakmongkon S, *et al.* Limited inhibitory effects of oseltamivir and zanamivir on human sialidases. Antimicrob Agents Chemother. 2008 Oct;52(10):3484–91.

151. Garcia EM, Serrano MG, Edupuganti L, Edwards DJ, Buck GA, Jefferson KK. Sequence comparison of vaginolysin from different *Gardnerella* species. Pathogens. 2021 Feb 1;10(2):1–15.

152. Ragaliauskas T, Plečkaitytė M, Jankunec M, Labanauskas L, Baranauskiene L, Valincius G. Inerolysin and vaginolysin, the cytolysins implicated in vaginal dysbiosis, differently impair molecular integrity of phospholipid membranes. Sci Rep. 2019 Dec 1;9(1).

153. Cortés-Sarabia K, Bautista-Santos I, Cisneros-Sarabia A, Parra-Rojas I, Estrada-Moreno AK, Flores-Alfaro E, *et al. Gardnerella vaginalis* Vaginolysin (VLY)-Derived MAP8 Peptide (VLY-MAP8) Induced the Production of Egg Yolk IgY Antibodies that Inhibit Erythrocytes Lysis. Int J Pept Res Ther. 2021 Mar 1;27(1):413–20.

154. Pleckaityte M, Janulaitiene M, Lasickiene R, Zvirbliene A. Genetic and biochemical diversity of *Gardnerella vaginalis* strains isolated from women with bacterial vaginosis. FEMS Immunol Med Microbiol. 2012 Jun;65(1):69–77.

155. Pleckaityte M. Cholesterol-Dependent Cytolysins Produced by Vaginal Bacteria: Certainties and Controversies. Front Cell Infect Microbiol. 2020 Jan 10:9:452.

156. Gelber SE, Aguilar JL, Lewis KLT, Ratner AJ. Functional and phylogenetic characterization of vaginolysin, the human-specific cytolysin from *Gardnerella vaginalis*. J Bacteriol. 2008 Jun;190(11):3896–903.

157. Korchev YE, Bashford CL, Pederzolli C, Pasternak CA, Morgan PJ, Andrew PW, Mitchell TJ. A conserved tryptophan in pneumolysin is a determinant of the characteristics of channels formed by pneumolysin in cells and planar lipid bilayers. Biochem J. 1998 Feb 1;329 (Pt 3):571-7.

158. Pawley DC, Dikici E, Deo SK, Raccamarich P, Fischl MA, Alcaide M, *et al.* Rapid Point-of-Care Test Kit for Bacterial Vaginosis: Detection of Vaginolysin and Clue Cells Using Paper Strips and a Smartphone. Anal Chem. 2022 Aug 23;94(33):11619–26.

159. Bostwick DG, Woody J, Hunt C, Budd W. Antimicrobial resistance genes and modelling of treatment failure in bacterial vaginosis: clinical study of 289 symptomatic women. J Med Microbiol. 2016 May;65(5), 377-386.

160. Qin H, Xiao B. Research Progress on the Correlation Between *Gardnerella* Typing and Bacterial Vaginosis. Front Cell Infect Microbiol. 2022 Mar 25:12:858155.

161. World Health Organization. Recommendations for the treatment of *Trichomonas vaginalis, Mycoplasma genitalium, Candida albicans,* bacterial vaginosis and human papillomavirus (anogenital warts). Geneva: World Health Organization. 2024 ISBN 978-92-4-009637-0.

162. Saricaoglu T, Ayaz CM, Hazirolan G, Tokgozoglu AM, Uzun O. Prosthetic Joint Infection Caused by *Gardnerella vaginalis:* Case Report and Review of the Literature. Clin Microbiol Infect. 2022 Feb 14;4(1):72–5.

163. Macedo-Viñas M, Petitpierre N, Cherkaoui A, Wirth G, Kaiser L. *Gardnerella vaginalis* as a Cause of Abscesses and Bacteremias: A Narrative Review Apropos of a Clinical Case. Infect Dis Clin Pract 2017 25(5) 243-249.

164. Mastromarino P, Mosca L. Bacterial vaginosis: A review on clinical trials with probiotics. 2013. New Microbiol. 2013 Jul;36(3):229-38.

165. Wu S, Hugerth LW, Schuppe-Koistinen I, Du J. The right bug in the right place: opportunities for bacterial vaginosis treatment. NPJ Biofilms Microbiomes. 2022 Dec 1;8(1).

166. Bhandari P, Hill JE. Transport and Utilization of Glycogen Breakdown Products by *Gardnerella* spp. from the Human Vaginal Microbiome. Microbiol Spectr. 2023 Apr 13;11(2).

167. Bhandari P, Tingley J, Abbott DW, Hill JE. Glycogen-Degrading Activities of Catalytic Domains of a-Amylase and a-Amylase-Pullulanase Enzymes Conserved in *Gardnerella* spp. from the Vaginal Microbiome. J Bacteriol. 2023 Feb 1;205(2).

168. Nunn KL, Clair GC, Adkins JN, Engbrecht K, Fillmore T, Forney LJ. Amylases in the Human Vagina. mSphere. 2020 Dec 9;5(6):e00943-20.