Genome Medicine

RESEARCH

Open Access

Metronidazole response profiles of *Gardnerella* species are congruent with phylogenetic and comparative genomic analyses

Katherine A. Innamorati[1,2,3], Joshua P. Earl[1,2,3,8*], Shirley C. Barrera[3], Rachel L. Ehrlich[1,2,3], Josephine Aiyeku[1,2,3], Ari Gordon[1,2,3], Evan Powell[4], Adam C. Retchless[6], Azad Ahmed[1,2], Bhaswati Sen[1,2], Sergey Balashov[1,2,3], Joshua Chang Mell[1,2,3], Sharon L. Hillier[7] and Garth D. Ehrlich[1,2,3,5*]

Abstract

Background Bacterial vaginosis (BV) affects 20–50% of reproductive-age female patients annually, arising when opportunistic pathogens outcompete healthy vaginal flora. Many patients fail to resolve symptoms with a course of metronidazole, the current first-line treatment for BV. Our study was designed to identify genomic variation associated with metronidazole resistance among strains of *Gardnerella vaginalis* spp. (GV), a genus of biogenic-amine-producing bacteria closely associated with BV pathogenesis, for the development of a companion molecular diagnostic.

Methods Whole-genome sequencing and comparative genomic metrics, including average nucleotide identity and GC content, were performed on a diverse set of 129 GV genomes to generate data for detailed taxonomic analyses. Pangenomic analyses were employed to construct a phylogenetic tree and cluster highly related strains within genospecies. *G. vaginalis* spp. clinical isolates within our collection were subjected to plate-based minimum inhibitory concentration (MIC) testing of metronidazole ($n=60$) and clindamycin ($n=63$). DECIPHER and MAFFT were used to identify genospecies-specific primers associated with antibiotic-resistance phenotypes. PCR-based analyses with these primers were used to confirm their specificity for the relevant genospecies.

Results Eleven distinct genospecies based on standard ANI criteria were identified among the GV strains in our collection. Metronidazole MIC testing revealed six genospecies within a closely related phylogenetic clade contained only highly metronidazole-resistant strains (MIC ≥ 32 µg/mL) and suggested at least two mechanisms of metronidazole resistance within the eleven GV genospecies. All strains within the six highly metronidazole-resistant genospecies displayed susceptibility to clinically relevant clindamycin concentrations (MIC ≤ 2 µg/mL). A PCR-based molecular diagnostic assay was developed to distinguish between members of the metronidazole-resistant and mixed-response genospecies, which should be useful for determining the clade membership of various GV strains and could assist in the selection of appropriate antibiotic therapies for BV cases.

*Correspondence:
Joshua P. Earl
jpe39@dragons.drexel.edu
Garth D. Ehrlich
ge33@drexel.edu
Full list of author information is available at the end of the article

Conclusions This study provides comparative genomic and phylogenetic evidence for eleven distinct genospecies within the genus *Gardnerella vaginalis* spp., and identifies genospecies-specific responses to metronidazole, the first-line treatment for BV. A companion molecular diagnostic assay was developed that is capable of identifying essentially all highly metronidazole-resistant strains that phylogenetically cluster together within the GV genospecies, which is informative for antibiotic treatment options.

Keywords Comparative genomics, Pangenome, Supragenome, Distributed genome hypothesis, Antibiotic resistance, Bacterial vaginosis, Phylogenetics, Gardnerella vaginalis, Molecular diagnostic, Metronidazole resistance

Background

Bacterial vaginosis (BV) is a commonly occurring microbial dysbiosis of the lower female genital tract arising when opportunistic pathogens outcompete healthy vaginal flora [1, 2]. Affecting over 21.2 million individuals in the USA annually, patients with BV experience discharge, dysuria, malodor, and itching and are at an increased risk of post-cesarean endometritis, urinary tract infections, and HIV acquisition [3–9]. The Centers for Disease Control recommends topical or oral metronidazole, a nitroimidazole antibiotic, as the most common treatment for BV, followed by topical clindamycin [10]. Tinidazole and secnidazole are recommended as alternative regimens [11]. Antibiotic use resolves approximately 70–80% of BV cases, but yearly recurrences range from 50 to 70% [7, 8, 12–14].

The composition of human vaginal microbiomes in individuals of reproductive age are predominantly characterized by acid-producing bacteria such as members of the *Lactobacillus* and *Limosilactobacillus* genera or mixed anaerobe communities [15–17]. The major species of healthy vaginal microbiomes produce hydrogen peroxide, lactic acid, and bacteriocins that reduce vaginal pH, promote their own growth, and inhibit the growth of pathogens. Menstruation, pregnancy, and sexual intercourse can reduce the abundance of commensal microbes, which increases the pH and facilitates the growth of pathogens that produce biogenic amines such as putrescine, cadaverine, and trimethylamine. This further increases the vaginal pH and produces the malodorous symptoms of BV [1, 18]. One proposed causative species of BV is *Gardnerella vaginalis* spp. (GV), a rod-shaped, Gram-indeterminate, facultative anaerobe present in 90–97% of BV cases and in less than half of healthy vaginal microbiomes [19–22]. The GV genome also encodes several virulence factors for biofilm formation, nutrient scavenging, and biogenic amine production, providing a survival advantage in the acidic vaginal niche [19, 23, 24].

G. vaginalis spp. has been a species in taxonomic flux since its discovery in 1953. Initially characterized as *Haemophilus vaginalis* in 1955, it was recharacterized as *Corynebacterium vaginalis* then renamed again as *Gardnerella vaginalis* [21, 25, 26]. Early typing efforts identified a series of biotypes and genotypes that distinguished among subtypes of the "species" [27–30]. A previous investigation by our lab concluded that GV was not a single species, but rather a genus composed of at least four distinct clades each with its own characteristic genome size, core genome, and GC content [31]. This was followed by the development of qPCR primers that could differentiate with high confidence between the four clades, and correlation of cpn60 subtypes with each clade [32, 33]. Shortly after, metronidazole resistance (MIC $\geq$ 32 μg/mL) was determined to vary among clades; strains contained within clades 1 and 2 reported irregular metronidazole response profiles (clade 1 = 35% resistant, clade 2 = 7.1% resistant), whereas all strains within clades 3 and 4 were fully resistant to this first-line antibiotic [34].

Additional comparative genomic analyses such as average nucleotide identity (ANI) and digital DNA-DNA hybridization on larger GV datasets led to the discovery of nine to thirteen component "genomospecies" or "genospecies," providing an amended species definition and new species names for three of the well-defined species (*G. piotii*, *G. swidsinskii*, and *G. leopoldii*) [35, 36]. Since that time, additional studies have confirmed this diversity by analyzing genotypes of virulence factors or through evidence of lateral gene transfer, and a recent study has announced two additional named genospecies, *G. pickettii* and *G. greenwoodii*, highlighting a need for additional genus-level analyses [37].

Comparative bacterial genomic methodologies were originally developed to test the distributed genome hypothesis, in which it was posited that each bacterial strain within a species possesses a unique complement of non-core (distributed/accessory) genes [31, 38–51]. The results of these studies defined the bacterial species-level supragenome/pan-genome and ushered in a new era in phylogenetically supported taxonomy. The underlying technologies were then extended to study genus and family-level pangenomes [52–58]. The clinical fruits of these advances enabled gene-based diagnostics and therapeutics for infectious diseases. [53, 54, 59]

The association of bacterial gene presence or absence with clinical metadata can answer many questions regarding tropism, virulence, and antibiotic resistance [38–41, 43, 53, 54, 60, 61]. Should analysis of a species reveal that a distributed gene is "core" to a clinically distinct group of strains, that gene can serve as a biomarker to inform a patient's treatment. In this study, we aimed to improve the clade-level resolution of strains into genospecies within GV using comparative genomics to discriminate between highly metronidazole-resistant and mixed-metronidazole response genospecies.

Methods

Method details

Bacterial strains and growth conditions

All 129 strains of *Gardnerella vaginalis* spp. used in this project are listed in Additional file 1: Table S1- S2 and Table S5. Strains were streaked from glycerol stock onto 12-h pre-reduced A80 selective plates (Hardy Diagnostics, Santa Maria, CA, USA) and grown under anaerobic conditions (37 °C, anaerobic mixed gas canister [5.5% H_2, 10% CO_2, N_2] + N_2 canister) in an anaerobic chamber (BugBox, Baker Ruskinn, Sanford ME USA) for 3–10 days. Liquid cultures were cultivated in Bacto BHI (BD Biosciences, San Jose, CA, USA) or Casman Broth (HiMedia, Kennett Square, PA, USA) supplemented with 10% horse serum (Boston BioProducts, Ashland, MA, USA) and grown under anaerobic conditions.

Genome acquisition

Genomes analyzed in this project were acquired from several different sources. Clinical isolates ($n=57$) were acquired from the biorepository of one of us (SLH) at the University of Pittsburgh and Magee-Women's Research Institute in Pittsburgh, PA, USA. Vaginal, endometrial, and placental samples were collected from symptomatic and asymptomatic females with known pregnancy status, and rectal samples were collected from males (Additional file 1: Table S2). The bacterial isolates were obtained from individuals enrolled in nine different clinical trials or observational cohort studies. All participants provided written informed consent prior to the collection of their samples and provided future use consent for use of their samples or sample remnants. Samples in this biorepository are deidentified and cannot be linked with patient identifiers.

High-quality genomes ($n=59$) were downloaded from NCBI in addition to 13 strains originally described by Ahmed et al. [31], for a total of 129 strains included in this study [31, 35, 62–67]. Accession numbers and additional metadata for each downloaded strain, including non-published direct submissions to NCBI, can be found in Additional file 1: Table S1. Taxator-tk (v1.2), performed on all strains, was used to exclude genome sequences that were not considered *G. vaginalis* from further analysis [68].

Isolation of DNA

Strain DNA was isolated using the "Purification of Total DNA from Animal Tissues Spin-Column Protocol" from the DNeasy Blood and Tissue Kit (Qiagen, Venlo, NED) after pretreatment for Gram-positive bacteria in enzymatic lysis buffer (20 mM Tris–Cl, pH 8.0; 2 mM sodium EDTA; 1.2% Triton X-100; lysozyme at 20 mg/mL directly before use). DNA quality was confirmed by NanoDrop (ThermoFisher, Waltham, MA USA). Genomic DNA was visualized via gel electrophoresis run at 180 V on a 1% agarose gel stained with ethidium bromide.

WGS assembly

Genomes generated using Roche/454 Life Sciences GS FLX Titanium sequencing technology ($n=20$) were prepared, assembled, and annotated [31]. Briefly, individual fragment libraries (optimized to produce between 300 and 400 base pair average read length) were created from each strain. Preparation and pyrosequencing were performed as described in GS FLX Titanium emPCR and Sequencing Protocols (October 2008). The raw sequence reads for each strain were assembled into contigs using a Roche/454 Life Sciences GS de novo Newbler assembler (version 2.0.00.20 or 2.0.01.14) and the default parameters except for minimum overlap identity, which was adjusted to obtain the fewest contigs. Raw sequence reads were assembled into contigs with the software Newbler (version > =2.0.00.20). These strains were spread across the tree and did not make up any one genospecies in its entirety, indicating that 454 sequencing artifacts are not responsible for the genomic variation seen between the genospecies. The remaining strains ($n=37$) were sequenced as paired end runs at 2×150nt on an Illumina NextSeq500. Individual libraries were created with Nextera XT and multiplexed to target 100–300× coverage per sample (raw nucleotide sequence/predicted genome size). Illumina reads were assembled as previously described using a custom pipeline [69]. Briefly, reads were trimmed using Trimmomatic v0.30, overlapped when possible using COPE v1.1.2, error corrected with ALL-PATHS-LG, and assembled with Ray v2.2.0 [70–73].

Pan-genome analysis

Annotation and pan-genome analyses were carried out as previously described [46, 49, 74]. In short, all 129 assemblies (both novel and downloaded from NCBI) were annotated with Prokka v1.11 [75]. Homologous genes were clustered with Roary v3.5.1 [76]. The threshold used for gene clustering of homologs was set at 60% BlastP

identity. We examined a range of identity thresholds (at 5% increments) and 60% was determined to be the lowest threshold at which changing the threshold resulted in only a small change in the clustering (See Additional file 2: Fig S1).

Phylogenetic analysis

ANI was calculated for all strain pairs with pyani v0.2 using the ANIb method [77]. Twenty-five genomes of *Escherichia coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Bacillus cereus* were downloaded from RefSeq and run with ANIb methods for comparison. All analyses were performed using R Statistical software (v3.6.3) [78]. Data visualizations were performed with the ggplot2 package (v3.3.0) [79]. Self-comparisons were removed with the harrietr package (v0.2.4), then 1-ANI values were converted to a distance matrix and clustered with single linkage [80]. Core genes were aligned with PRANK (v.140603) and the resulting alignments were concatenated [81]. The phylogenetic tree was constructed from this core gene alignment with RAxML (v8.2.4) [82] and visualized using the ggtree package (v3.4.4) [83, 84].

Synteny investigation

Genes of interest were aligned in Mauve (snapshot v. 20,150,226) [85]. Genes were analyzed with BlastKOALA via the online KEGG platform and potential homology was studied via BLAST [86, 87]. Results were visualized in R using the gggenes (v0.5.1) package [88].

Antibiotic susceptibility tests and MIC determination

Strains from our clinical isolate collection and those included from the Ahmed et al. paper [31] were revived from − 80 °C glycerol stocks on A80 plates and re-plated once. Colonies were scraped and suspended in Casman broth; 200 µL of suspension was streaked in duplicate or triplicate on pre-reduced A80 plates. Metronidazole or clindamycin MIC strips (0.016–256 ng/µL, Liofilchem, Waltham, MA, USA) were added to plate centers and incubated under anaerobic conditions. MICs were recorded where the no-growth ellipse intersected the MIC strip on day 3 for clindamycin and day 10 for metronidazole, in accordance with reported clinical treatment practices [89, 90]. The median of triplicate readings was reported as the MIC; in duplicate readings, the higher number was reported. Metronidazole and clindamycin breakpoints were used as recommended by CLSI (Metronidazole: R (resistant) ≥ 32 µg/mL, I (intermediate) = 16 µg/mL, S (sensitive) ≤ 8 µg/mL; Clindamycin: R ≥ 8 µg/mL, I = 4 µg/mL, S ≤ 2 µg/mL) [91].

Scoary analysis

Scoary v1.6.16 was provided with a gene presence-absence table from Roary, a discrete trait file with susceptible and resistant phenotypes, and our concatenated core gene tree (see " Phylogenetic analysis") [92]. Results were subjected to Bonferroni and Benjamini–Hochberg multiple testing correction procedures, candidates were ranked by worst-case *p* values for different potential evolutionary histories, and the remaining significant genes were selected for analysis [93, 94].

Identification of candidate genes

A gene presence-absence table was constructed by Roary as described above (see " Phylogenetic analysis"). Genes present in every strain of highly metronidazole-resistant genospecies and absent in every strain of mixed-susceptibility genospecies were considered primer set candidates.

Primer set development

FASTA files for candidate genes were aligned with MAFFT v7, utilizing G-INS-1 and default parameters [95, 96]. Alignments were then provided to the DECIPHER "Design Primers" tool [97]. Genes requiring fewer than 4 distinct F and R primers were selected for oligo synthesis (IDT, Coralville, Iowa USA). Degenerate bases were incorporated if necessary. In silico PCR using USEARCH compared primers against genomic datasets [98].

Clade-specific PCR

Amplifications were performed as 25 µL reaction mixtures containing 10 µL genomic DNA. Reactions were prepared according to the specifications for GoTaq DNA Polymerase (Promega, Madison, WI, USA). Final concentrations of 1X reaction buffer, 10 mM dNTP (Promega, Madison, WI, USA), 1.0 µM forward and reverse primers, and 1.25 u polymerase were used. PCR was performed in a Mastercycler X50 Thermocycler (Eppendorf, Hamburg, DE). Cycling parameters began with a 2-min denaturation at 95 °C, followed by 35 cycles of 15 s denaturation at 95 °C, 15-s annealing at 55 °C, and 30-s elongation at 72 °C. A final extension for 7 min was performed at 72 °C, followed by a hold at 4 °C. Annealing temperatures for primers were identified by gradient PCR (See Additional file 2: Fig. S2). PCR products were visualized on 2% agarose gels stained with ethidium bromide.

Results

Phylogenetics and comparative genomics suggest eleven distinct genospecies of Gardnerella vaginalis spp.

Prior research into the genomic and phylogenetic structure of *G. vaginalis* spp. isolates revealed a minimum

of four clades within the taxa, a discovery supported by: > 0.5% differences in GC content; changes in the estimated core genome size from 27 to 91% when calculated within or among clades; sequence length variations ranging from 1.491 to 1.717 MB, and the observation that horizontal gene transfer (HGT) was frequently seen within, but not among the four clades [31].

We sequenced 57 additional genetically diverse clinical isolates of *G. vaginalis* spp, 36 of which were selected after prescreening a set of 50 strains by PCR for unique combinations of distributed genes from the four original clades and combined them with 72 previously characterized genomes to assemble a 129-strain pangenome that likely contains the majority of the clade (genospecies) structure within the GV genus (Additional file 1: Table S2-S4).

To determine nucleotide diversity among the GV genomes and estimate the number of distinct genospecies present in the pangenome, we calculated the difference in Average Nucleotide Identity (ANI) among all possible strain pairs ($n = 16{,}641$), using the ANIb (BLASTN+) method (Fig. 1a). ANIb calculates pairwise comparisons among all strains in a set using BLAST, and values of 95% or greater are considered indicative of a single genospecies [99]. After removal of self-comparisons, the median ANIb value among all strains was 84.1%. The two most genetically similar strains, GV_75712 and B473, had an ANIb value of 99.9%, while the two least similar strains (GED7760B and BS611) possessed an ANIb value of 79.0%. A comparison of the median ANI value to a 25-strain subset of several well-characterized bacterial species indicates that the median ANIb value for all GV strains is far outside the range for a single species (Fig. 1b). The presence of larger blocks of deep blue are indicative of strains pairs with highly similar ANI values, such as those in the top right encompassing strains GV_101, B719, UMB1686, and B512, or the grouping of 53 strains in the bottom left. We suggest that these blocks of highly genomically similar strains belong to distinct phylogenetic clades or genospecies (Fig. 1a). The ANI data were visualized as a dendrogram and compared to a phylogenetic tree composed of concatenated core genes from all 129 GV strains in the analysis set (Fig. 1c). A cutoff of 5% ANI difference was applied to the dendrogram, in line with the Konstantinidis ANI species definition [99]. Strains were considered to belong to different clades if a vertical branch connecting them crossed that threshold. This method identified eleven distinct genospecies.

A phylogenetic tree was constructed from alignments of 431 genes determined to be core (as defined below) to the GV genus (Fig. 2). The majority of isolates formed three major genospecies with two sub-species each, all six of which would be classified as individual species based on both the differences in their ANI values (< 95%) and the differences in their % GC content values (> 0.5%). Previous studies have split two of the observed "major" genospecies into component subspecies and presented evidence for the renaming of each. Specifically, genospecies 4 is divided into 4a (*G. pickettii*) and 4b (*G. piotii*), and genospecies 7 is divided into 7a (*G. swidsinskii*), and 7b (*G. leopoldii*) [35, 37]. Genospecies 1a remains the amended definition of *G. vaginalis*, but further testing will be required to determine if genospecies 1b is included within *G. vaginalis* or if renaming is required. Notably, if ANI is set at 94%, 8 total groupings are seen, and if the ANI threshold is set to 96%, the total number of groups rises to 16. This indicates that defining an exact number of species in the genus is complicated by high sensitivity to the specific ANI threshold. The exact number notwithstanding, this analysis demonstrates a minimum of eight separate genospecies within GV, corroborating previous findings [31, 35–37, 100, 101].

The core genome percentage is more representative of a species when GV is subset into individual genospecies

Further support for splitting GV into multiple genospecies was provided by examining how core genome membership changed when pangenomes of individual genospecies were calculated. The pangenome of GV in its entirety contained 3324 genes but only 615 genes were identified as core, comprising ~ 18% of the total gene content. A previous review of 295 bacterial species pangenome analyses revealed that the median core gene percentage for a pangenome calculated on at least 25 strains is 39 ± 21%, of which the core genome of GV is a full standard deviation below [53]. However, when individual genospecies pangenomes were recalculated for the three largest GV genospecies (1a/b, 4a/b, and 7a/b), the core genome of each genospecies increased, ranging in size from 44–51% (Fig. 3). These data further support the claim that GV is comprised of distinct groups of organisms that should be assigned to different species.

Metronidazole resistance is predominantly associated with distinct genospecies

Antibiotic susceptibility tests (AST) were performed on clinical isolates from ten of the eleven identified genospecies to determine quantitative levels of metronidazole susceptibility in GV clinical isolates (See Additional file 1: Table S5). Strains from genospecies 8 could not be acquired for this experiment. The minimum inhibitory concentration (MIC) of these strains to metronidazole was measured with test strips containing a gradient of 15 two-fold dilutions (0.016–256 μg/mL). Resistance to metronidazole was determined by breakpoints set by the

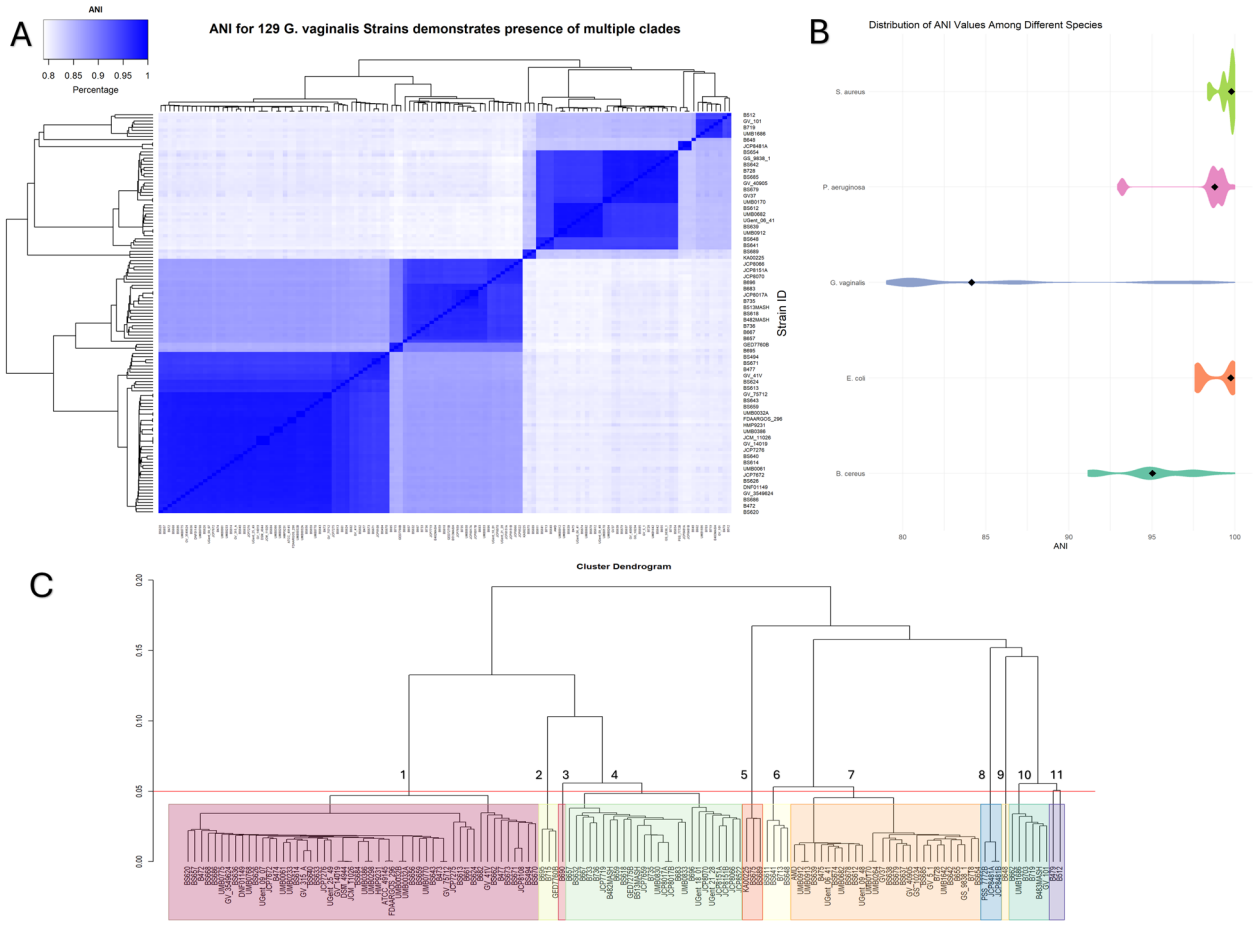

Fig. 1 Comparative genomic techniques suggest eleven distinct genospecies of *Gardnerella vaginalis* spp. **A** ANI approximates genome similarity using BLAST. Each of the 129 total strains of *G. vaginalis* are represented on each axis in the same order, where the diagonal represents comparison to self, and off-diagonal represents comparison to a different strain. The range of ANI from least to greatest is 79.0% (white) to 99.9% (dark blue). Cluster diagrams were built using single-linkage hierarchical clustering. **B** Violin plot of ANI values for 129 strains of *G. vaginalis* (median: 86 ± 0.07%) and 25 strains each of four well-characterized species: *Escherichia coli* (99.7 ± 1.0%), *Staphylococcus aureus* (99.8 ± 0.5%), *Pseudomonas aeruginosa* (98.8 ± 2.1%), and *Bacillus cereus* (95.0 ± 2.4%). Density of data points at a given percentage is indicated by the width of the plot. Median value is indicated by a black diamond. **C** Distance matrix of *G. vaginalis* strains calculated from 1-ANI values and hierarchically clustered using the average linkage method. A cutoff was drawn at 95% similarity (horizontal line) and the number of genospecies counted. Eleven genospecies were observed using this cutoff value. Comparison to genospecies seen in phylogeny is displayed in colored boxes and labeled with clade numbers from the phylogenetic tree in Fig. 2. See also Additional File 1: Table S1 – S4

Clinical & Laboratory Standards Institute (CLSI)- Resistance: MIC ≥ 32 μg/mL; Intermediate (antibiotic response depends on the dose given and the site of infection): MIC between 8 and 32 μg/mL; Sensitivity (or susceptibility): MIC ≤ 8 μg/mL (Fig. 4A) [91, 102, 103]. 49/60 (82%) of tested strains were classified as metronidazole resistant, and the remaining 11/60 strains (18%) were determined to be metronidazole intermediate or sensitive.

We superimposed the metronidazole MIC data onto the core-gene phylogenetic tree (Fig. 4B). Several genospecies are highly enriched for metronidazole resistance. All strains tested in genospecies 7a/7b, 9, 10, and 11 were highly resistant to metronidazole (22 of 22 strains (100%), MIC range 128–256 μg/mL, median MIC ≥ 256 μg/mL), with 95% of strains reporting an MIC ≥ 256 μg/mL (See Additional Fig. 5). Conversely, genospecies 1(a/b) reported mixed MIC responses (range 8–256 μg/mL). Forty percent of strains in genospecies 1 were classified as sensitive or intermediate responses (8–24 μg/mL). Full resistance was generally observed to split into two "classes", which we will label low (32–64 μg/mL, 36% of strains) or high (96–256 μg/mL, 24% of strains) resistance responses. Genospecies 1 reported a median MIC of 32 μg/mL, which is strikingly different from the median MIC observed in genospecies 7. Genospecies 4(a/b) also displayed varied metronidazole responses, reporting a

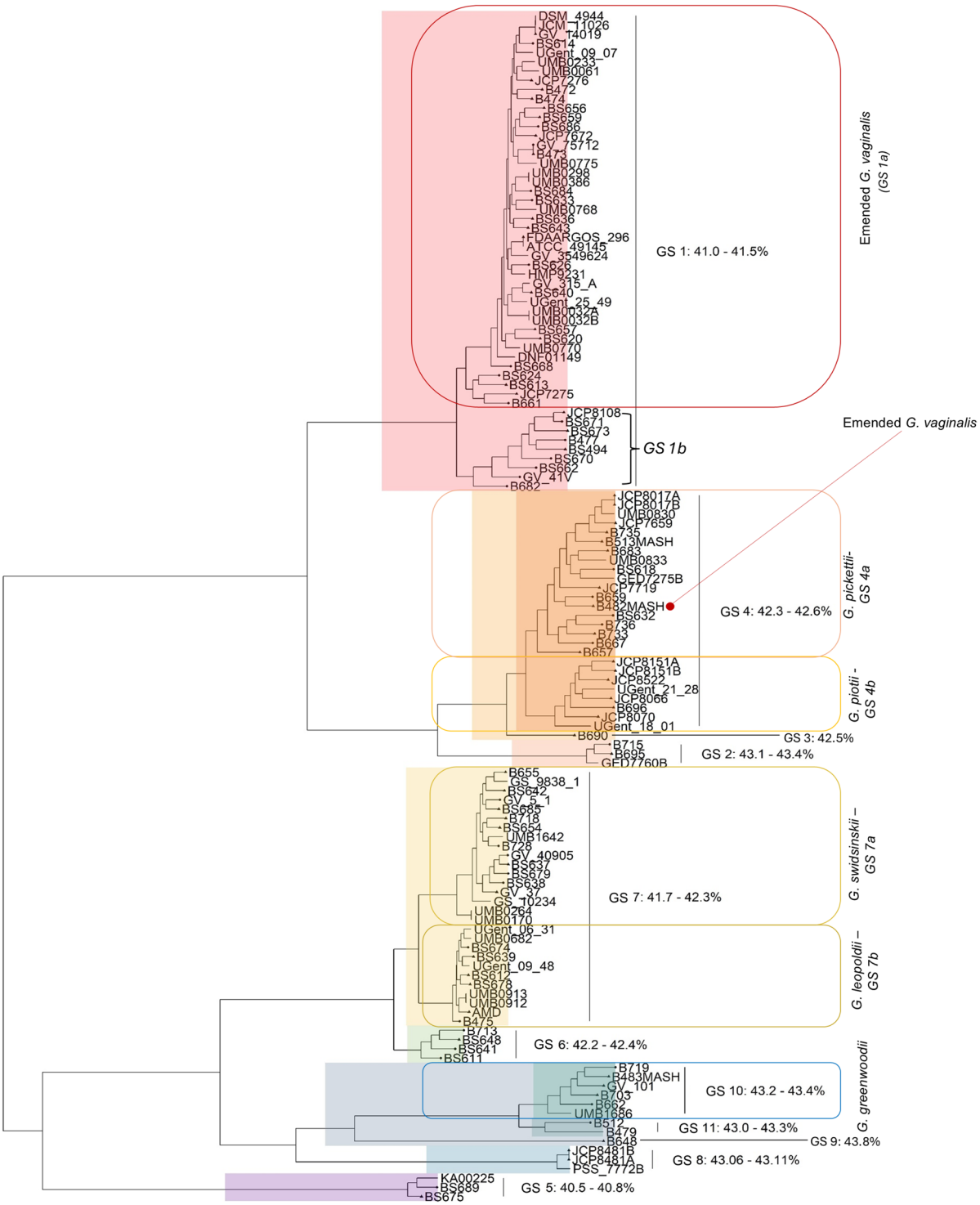

Fig. 2 Concatenated core-genome maximum likelihood tree displays the presence of eleven genetically distinct genospecies. The pangenome for 129 strains of *G. vaginalis* was calculated and 431 core genes were aligned with PRANK to create a phylogenetic tree with RAxML. Eleven individual genospecies (GS) are highlighted by color and marked with the range of GC content percentages within each. Singleton strains (genospecies 3 and 9) are included with the nearest genospecies with 5 or more strains and marked with a different shaded box. Genospecies designated as individual species by Vaneechoutte et al. [35] and Sousa et al. [37] are indicated by colored rings and labeled with their proposed species names. See also Additional File 1: Table S1-S2

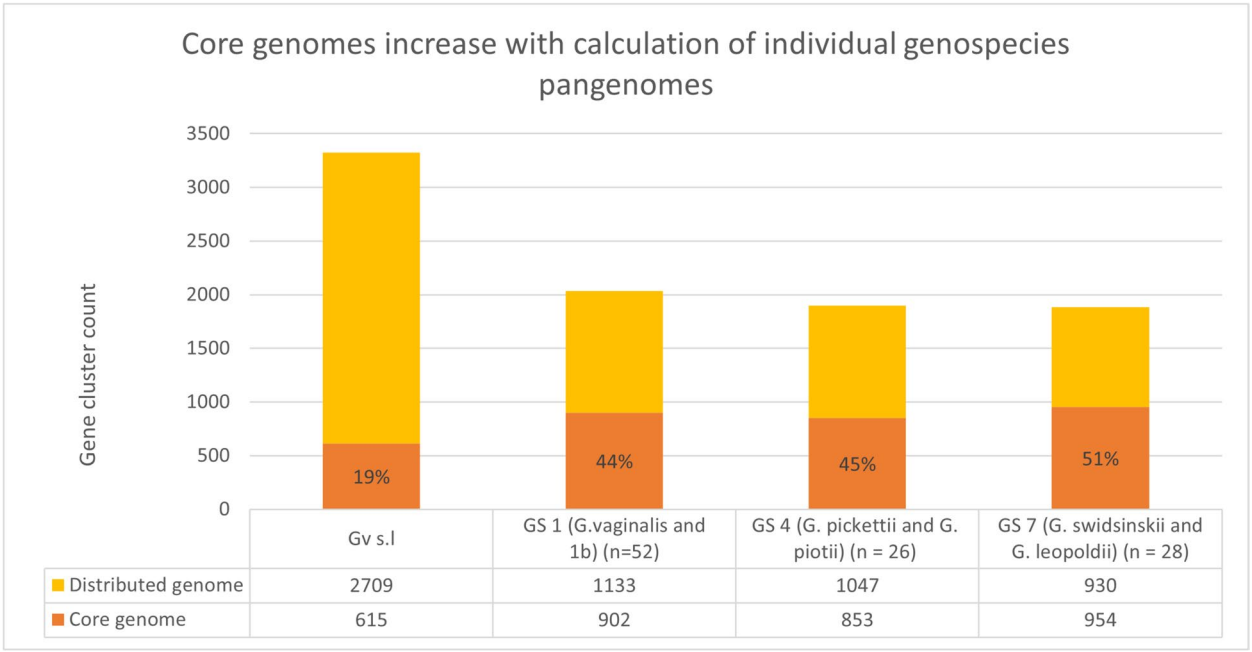

Fig. 3 Core genome percentage is more reflective of a single species after GV is separated into individual genospecies. The count of core genes (orange bar) of all 129 GV strains is less than half of the proportion of the independently calculated genospecies' pangenomes. See also Additional file 2: Fig. S1

median MIC (48 μg/mL) dramatically lower than genospecies 7(a/b) but higher than genospecies 1 (Table 1). The marked differences in resistance response in different genospecies could suggest differences in the mechanisms through which metronidazole resistance is conferred.

Examination of the phylogenetic tree (Fig. 4b) reveals genospecies 1–4 are primarily composed of metronidazole-intermediate or low-resistance strains. Using Scoary, one gene/operon was found to be significantly (negatively/positively) associated with resistance by Fisher's exact test but this may have been a spurious association due to population structure (data not shown). However, we examined the existence of three intermediate response strains in these genospecies with a metronidazole high-resistance closest neighbor. The gene cluster presence of these three strain pairs: B659/B482MASH (*G. pickettii*), BS620/BS657 (*G. vaginalis*), and B477/BS494 (GV genospecies 1b) were examined to identify any genes that were exclusive only to the metronidazole intermediate strains or the high-resistant ones. Across all three strain pairs, no gene clusters were found to be exclusive to the intermediate strains that were also not present in the high-resistant genomes. However, eleven genes were identified in all three high-resistant strains that were not present in their intermediate-response neighbors. Alignment of the genes present in these strains (Fig. 4C) revealed that the eleven genes were in a syntenic region. A high degree of similarity in the flanking regions was found in the corresponding intermediate-response strain pairs. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of these genes identified a putative DNA primase/helicase, putative transcriptional regulator, macrolide transport system ATP-binding/permease protein, and an MFS transporter—DHA3 family macrolide efflux protein. Five genes reported no assigned function. Interestingly, the MFS transporter is further defined as a drug transporter and an important antimicrobial resistance gene (KEGG identifier KO8217). The presence of this individual gene is not sufficient to convey metronidazole resistance as three additional strains in genospecies 1 (B661, BS613, and BS668) with intermediate metronidazole response were found to also contain members of this gene cluster, nor is it necessary for metronidazole resistance as it was missing in at least some resistant strains (e.g. BS641). This data also suggests that there may be multiple pathways to metronidazole resistance, and acquisition of this individual gene is not enough to become resistant to metronidazole.

Metronidazole-resistant strains are susceptible to clindamycin

We next asked whether GV isolates were susceptible to alternative antibiotics used for BV treatment. We plated 25 strains from genospecies 5–11, and 38 strains from genospecies 1–4. All 63 strains (100%) showed high sensitivity to clindamycin (CLSI MIC sensitivity

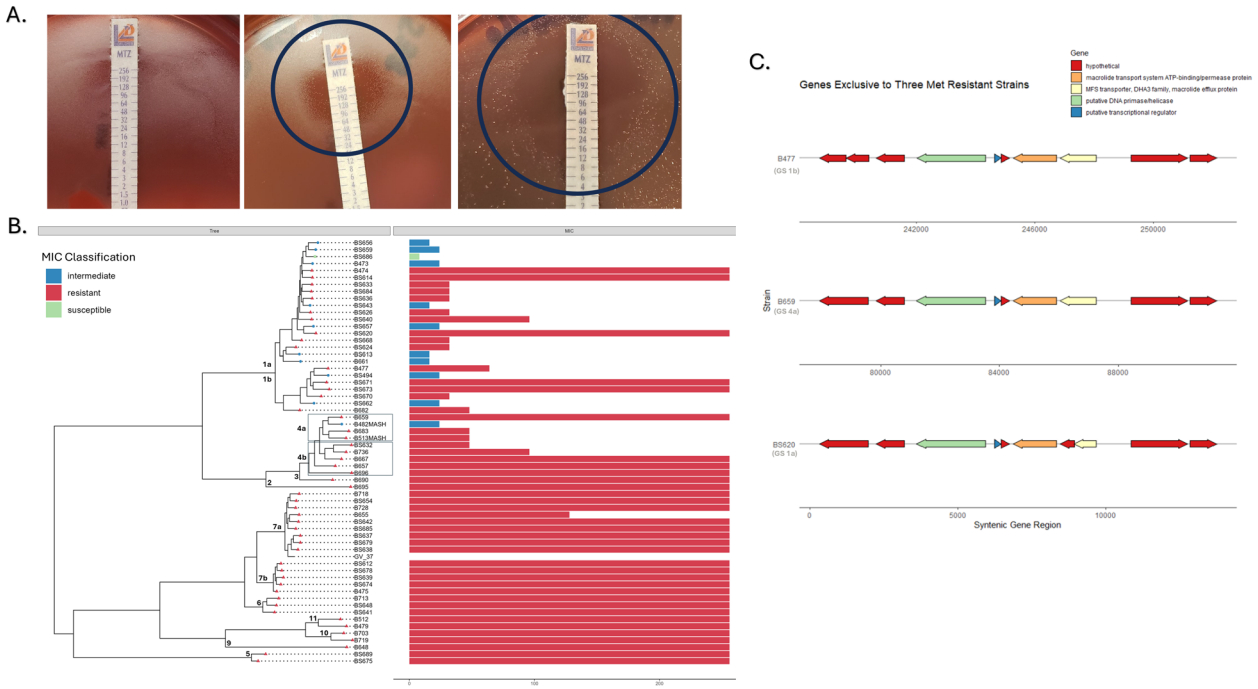

Fig. 4 GV genospecies display two major phenotypic responses to metronidazole, but analysis of genes does not uncover the mechanism of resistance. **A** Representative images of MIC assays. Each metronidazole MIC strip contains fifteen twofold dilution concentrations. An ellipse of inhibition forms at the minimum concentration of metronidazole required to stop GV growth. Left: Resistant strain with MIC > 256 µg/mL. Middle: Resistant strain with MIC of 64 µg/mL. Right: Susceptible strain with MIC of 8 µg/mL. **B** Concatenated core gene tree labeled with MIC data indicate the existence of genospecies correlated with antibiotic resistance. Results of MIC analyses were plotted against the tree. At least one strain from each genospecies reported a result except for genospecies 8. On the right, MIC data are shown for each tested strain. A color legend is available at the top left. Length of MIC bar corresponds with MIC value. Color corresponds to CLSI classification. **C** Alignment of syntenic regions identified in the metronidazole-resistant partner of three resistant/susceptible strain pairings within the mixed-resistance genospecies. Colored arrows represent each of the genes with one of five KEGG-annotated functions (see legend). Arrow direction corresponds with coding strand. See Additional file 1: Table S5

breakpoint ≤ 2 µg/mL) with MICs ranging from 0.016–0.5 µg/mL and a median MIC of 0.125 µg/mL [91]. Clindamycin treatment was equally effective on strains from genospecies 1–4 and genospecies 5–11 (Fig. 5).

Presence of the aruH gene is diagnostic to genospecies displaying high metronidazole resistance

The presence of two strikingly different metronidazole response phenotypes, delineated by genospecies, provided the impetus to develop a molecular diagnostic to identify strains within the metronidazole high resistance genospecies. Fourteen genes identified as "core" (present and confirmed via BLAST and Roary in every strain) to the entirety of genospecies 5–11 that were absent from the mixed-resistance genospecies (genospecies 1–4) were selected for the creation of diagnostic primers using DECIPHER, an online toolkit which designs primers based on multiple sequence alignments (See Additional file 1: Table S6) [97]. Primers were designed for five genes (*yfeA, LysM, aruH, bioY,* and *comEA),* and three hypothetical proteins (group_2343, group_2326, and group_1951) (See Table 2). Names or functions of each named gene were confirmed by blastx [87]. Successful universal PCR amplification among the highly resistant isolates was only accomplished with primers for the *aruH* gene, which catalyzes the transamination of L-arginine and pyruvate into ketoarginine and L-alanine [104]. *AruH* amplification was observed in all strains tested in genospecies 5, 6, 7a (*G. swidsinskii)*, 7b *(G. leopoldii)*, 10, and 11 (Fig. 6). Two bands were observed in most amplified strains, likely due to the presence of a gene paralog identified in the genomes of these strains. Genospecies 9, composed of a single strain (BS610/B648), did not support amplification with these primers. We were unable to acquire DNA from any genospecies 8 strains for testing, but in silico amplification with USEARCH was positive for the gene in all three genospecies 8 genomes (JCP8481B, JCP8481A, PSS_7772B) [98]. PCR amplification of *aruH* was not observed for any strain found in genospecies 1a (*G. vaginalis)*, 2, 3, 4a (*G. pickettii)*, or 4b (*G.*

Table 1 Range and median metronidazole MIC for tested GV clinical isolates for each genospecies and sub clade/genospecies, with associated species names where appropriate

Species name	Genospecies	Median MIC	Range (ng/μL)	Strains included
G. vaginalis	1a	32	8–256	18
	1b	48	24–256	7
	2	256	256	1
	3	256	256	1
G. pickettii	4a	48	24–256	8
G. piotii	4b	256	256	1
	5	256	256	2
	6	256	256	2
	7a	256	128–256	9
G. swidsinskii	7b	256	256	5
G. leopoldii	8	NA	NA	0
	9	256	256	1
G. greenwoodii	10	256	256	1
	11	256	256	2

piotii), which are associated with a mixed-sensitivity and low-resistance response to metronidazole (median MIC = 32–48 μg/mL) (Fig. 6). Further analysis will need to be done to confirm lack of amplification in genospecies 1b. Additionally, in silico PCR of our *aruH* primers against NCBI reference strains for three of the most common species colonizing the vaginal microbiome (*L. crispatus, L. iners, L. gasseri)* did not result in amplification, supporting the specificity of our primer set [105] (data not shown).

Discussion

Previously, our lab proposed that GV was best described as a genus composed of at least four distinct clades based on GC content, genome size, number of core genes, and rates of homologous recombination among the clades [31]. Since then, correlations between clade and metronidazole resistance were observed in two of the four original clades, and additional analyses have proposed amending the species definition of *Gardnerella vaginalis* leading to the newly described genospecies *G. leopoldii*, *G. piotii*, *G. swidsinskii*, *G. greenwoodii*, and *G. pickettii* [34, 35, 37].

In this study, we further characterized the *G. vaginalis* spp. phylogenetic clade structure and identified gene content and phenotypic differences among the multiple genospecies within what is now recognized as the GV genus using a large collection of diverse clinical isolates. Phenotypic assays of antibiotic susceptibility revealed that five distinct genospecies displayed complete resistance to metronidazole. Utilizing gene possession differences between high-resistance and sensitive/intermediate/low-resistance genospecies, we developed a simple PCR-based molecular diagnostic that can positively identify strains belonging to highly metronidazole-resistant genospecies to help guide clinical management, since all tested members of these resistant genospecies are clindamycin-sensitive.

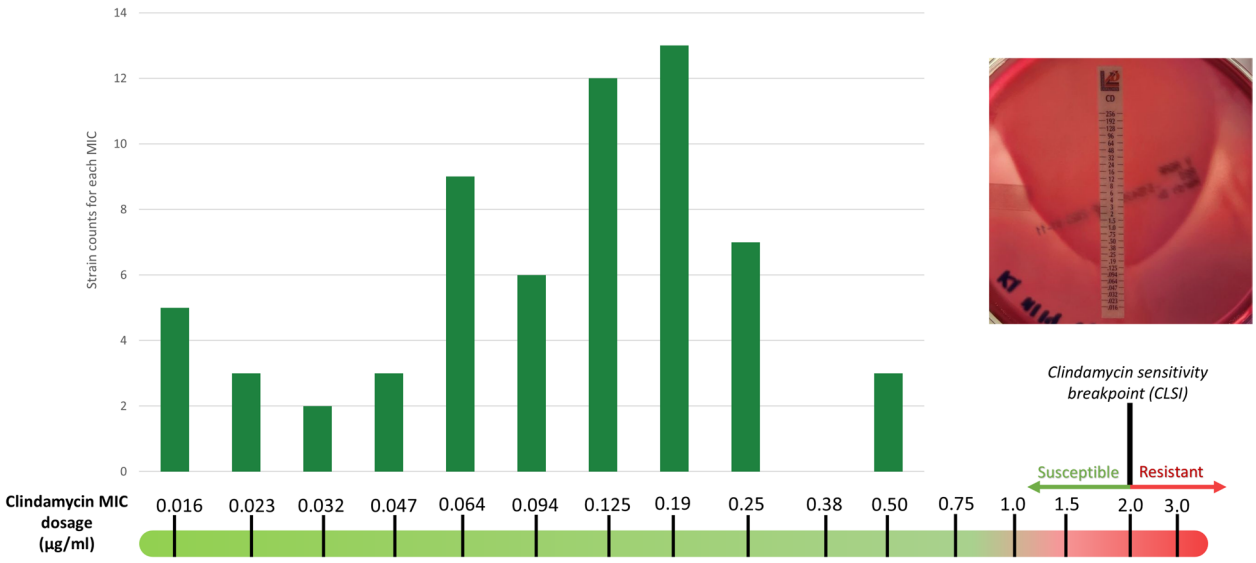

Fig. 5 Selected metronidazole-resistant strains are susceptible to clindamycin. Sixty-three strains from various genospecies were subjected to antibiotic susceptibility testing with clindamycin MIC strips. A bar plot is superimposed above the gradient of measurable MIC values, indicating the number of strains with each measured MIC. The clinical sensitivity breakpoint as determined by CLSI for clindamycin is marked with an arrow (2.0 μg/mL). A representative image is shown at right, demonstrating the ellipse of inhibition observed for a clindamycin-susceptible strain with an MIC of 0.125 μg/mL. See Additional file 1: Table S5

Table 2 Sequences of primers tested for genes present in genospecies 5–11 that were not present in genospecies 1a, 2, 3, or 4

Primer name	Sequence
yfeA_F1	GCA GGM TAY GAY GAT TGG GC
yfeA_R1	CCT GGA TCT TGY TGA CCY TGR TT
aruH_F1	GCY ATY ATT ATC AAY TCW CCW TGC
aruH_F2	CTG TAA TCT TAA ATT CAC CAG GA
aruH_R1	GCA TCA TTT CTG CAA TTT TGC
aruH_R2	GTG TAT CAT TTC CGC AAT TTT AC
g2326_F	CAT TGG GTC AGC TTG CTC C
g2326_R	GGT TGG ATA ATC GTT ACC GCC
g2343_F	CAY GAR CCK TGG TTY GAY GAA G
g2343_R	ART RAA AGG MGM GCG GAA WA
yfeA_F2_1	TGC GTA TCA CTG TTA AAA CTT TAG
yfeA_F2_2	CAT TAT TTG CGC GGC AA
yfeA_F2_3	ACC GTT CTT TGT GCA GTA A
yfeA_R2	AAK RTC RTC GCA AAT CTT MA
yfeA_F4	GCC AAA ATA CGC AGA ACA GC
yfeA_R4	TGG ATC TTG CTG ACC TTG GT
lysM_1_F1	ART TAT GCT GMW RCT ATT ACT CC
lysM_1_F2	TAC GCT TCA ACA ATT ACT CC
lysM_2_F1	TAT GCT GMW RCT ATT ACT CCT
lysM_2_F2	ACG CTT CAA CAA TTA CTC CA
lysM_1_and_2_R1	GGA ACA TTA ATA GAC TGT CCG AT
lysM_1_and_2_R2	GCA CAT TAA TAG ACT GTC CAA C
bioY2_F	TGG TTT AAC GTT GAG CTG GC
bioY2_R	CAG AAG CGT GCG AAA GTC TT
comEA2_F	CGG AAG TGA TGT AAG TGA CGG
comEA2_R	CAG CAG TTT TGG GTC CGA TT
g2343_F	GAT TCT TCG TTT GCT GAT TT
g2343_R1	TGC TTG CAA WWC CAA CWA AAG
g2343_R2	ACT TGC AAA ACC AAC GAA AG
g1951_F	GCG TAT GAG TTC AGC CAG TG
g1951_R	CCA ACA TGA CGA GCG GAA TT
g2321_F	ATA TCG CCA CGC ACT AAA GC
g2321_R	TTC GTG CAT TCC TTC AGC TG

Prior to the advent of whole genome sequencing (WGS), microbial species classification was fraught with many taxonomic errors and little consensus as to whether a phenotypic or cladistic taxonomy best captured the true relationships among bacterial strains [106–109]. Genomic approaches such as pangenomic analyses, core gene content, and ANI follow guidelines stemming from the pre-sequencing era "gold standard" of 70% DNA-DNA hybridization, which was a somewhat arbitrarily set percentage calibrated from an even earlier phenotype-driven species definition, which may or may not accurately reflect more than a half century of research on bacterial speciation and which varies among different bacterial groups [49, 52, 108]. While species classifications for non-sexually reproducing organisms can therefore be challenging, they prove helpful in a clinical setting. Therefore, bacteria need to be classified using a combination of methods that collectively provide sufficient resolution among related strains that require different clinical responses after infection. In practice this means utilizing a number of genomic parameters including comparative genomics, phylogeny, and measurement of HGT [110].

We endeavored to characterize the component GV genospecies and found ANI and phylogenetic support for eleven GV genospecies using a concatenated core genome. This expands on our previous work and is in close alignment with the findings of the Ghent University group, who reported each subtype of these major clades as an individual genospecies [35]. Minor shifts from the canonical species definition of 95% ANI result in a range of eight to sixteen species using ANIs of 94% and 96%, respectively. Pan-genomic analyses also indicate that separation into multiple genospecies is appropriate. The median core gene percentage in single-species pangenome projects containing 25 or more strains is 39% [53], whereas the core of all 129 GV strains in this study is 18%. Core genomes for individual GV genospecies are nearly triple that of GV in its entirety and are within one standard deviation of the median core gene percentage seen in single-species projects. Gene presence and absence data for the strain collection did not identify a single gene that could provide a single unifying mechanism for metronidazole resistance among the resistant strains in our data set (data not shown). Therefore, it is likely that multiple genes confer metronidazole resistance, which may differ between genospecies. This hypothesis is supported by two observations. First, we have identified the existence of strain pairs in genospecies 1 (1a: *G. vaginalis* and genospecies 1b) (B477/BS494; BS657/BS620) and genospecies 4 (4a: *G. pickettii,* 4b: *G. piotii)* (B659/B482MASH) in which one strain displays an intermediate response to metronidazole and the other presents significant metronidazole resistance. Analysis of genes found only in the highly resistant strains of each pair uncovered a syntenic region composed of eleven genes with varying functions that could be related to antibiotic resistance, including an MFS transporter—DHA3 family macrolide efflux protein, which is classified as an important antimicrobial resistance gene. This protein was originally identified in *Streptococcus pyogenes* (PubMed ID 8971709). A BLAST analysis of this gene identifies its presence in multiple *Streptococcus* genomes at < 99% identity. The MFS transporter has been identified to function in quorum sensing

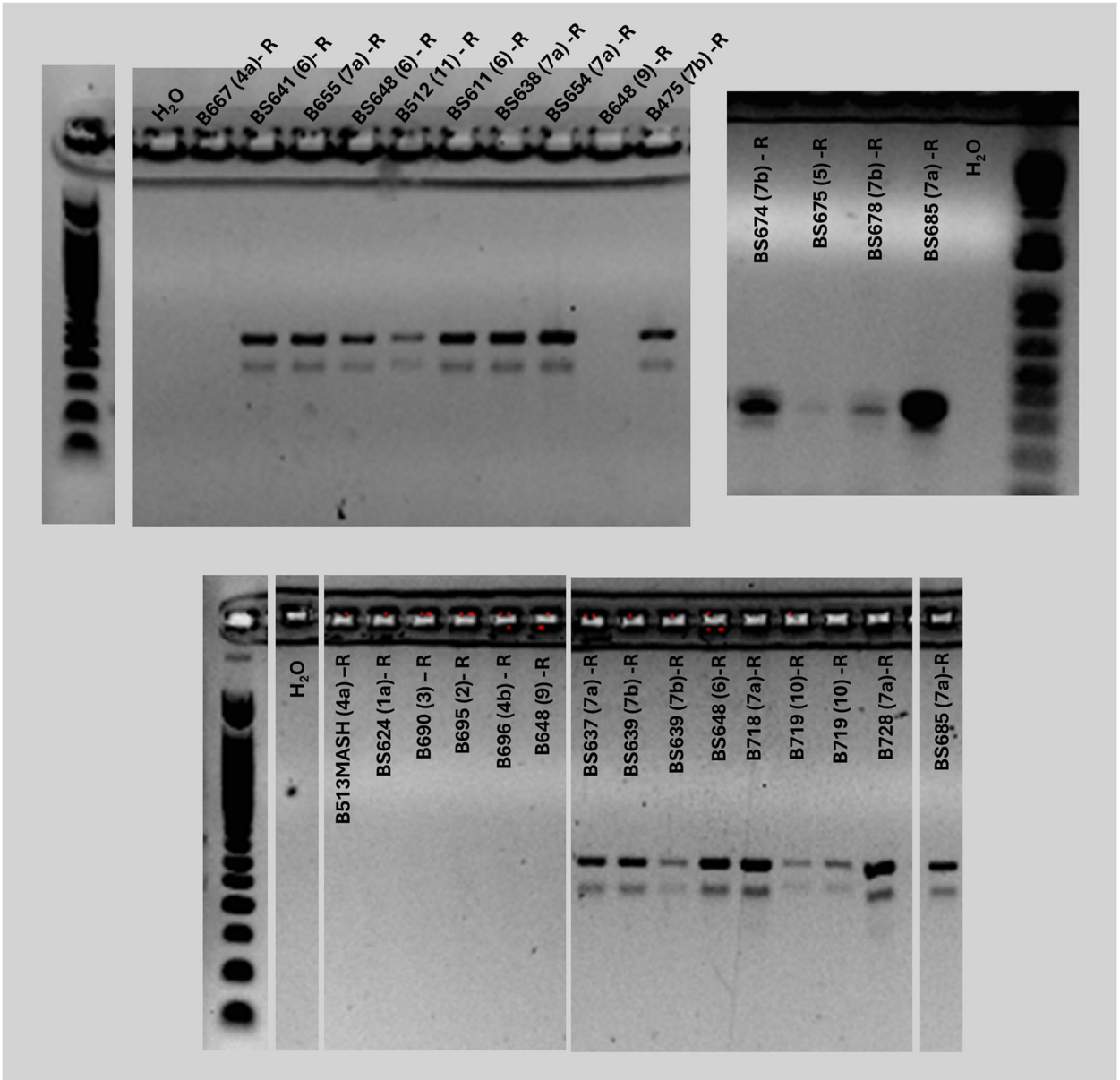

Fig. 6 *aruH* primers specifically amplify strains in metronidazole-resistant genospecies. A set of two forward and two reverse primers for *aruH*, encompassing the range of diversity seen in this gene among the various GV species, successfully discriminated between selected clinical isolates within the metronidazole-mixed response genospecies 1a, 2, 3, 4a, and 4b and the entirely metronidazole-resistant genospecies (5–11). Successful amplification coincided with membership in the highly metronidazole-resistant genospecies (5–11), with the exception of strain B648 (BS610), a resistant strain and the only member of genospecies 9. See also Table 2, Additional file 1: Table S6, and Additional file 2: Fig. S2

during biofilm formation in *S. pyogenes*, and HGT occurrences have been reported between different clades of GV, suggesting a potential mechanism through which this transporter could have been acquired [100, 111].

While the presence or absence of this specific gene was not a necessary or sufficient biomarker for metronidazole resistance, it is possible that expression levels of this gene could play a mechanistic role or help explain one component of how resistance occurs. Additionally, if a common mechanism of metronidazole resistance existed, we would have expected to observe patterns of gene presence or absence among the three sporadically highly resistant strains that were also common among strains of

highly metronidazole-resistant genospecies 5–11, but no clear patterns were distinguished.

Second, two significantly different metronidazole resistance phenotypes were seen. Resistance broadly fell into one of two categories: low-moderate resistance with an MIC of 32–96 μg/mL, or total (high) resistance, with an MIC of 128–256 μg/mL or higher. Both low-moderate and total resistance phenotypes were observed in the species that also reported metronidazole susceptible and intermediate strains (genospecies 1–4), but all except for one strain in the genospecies reported as fully metronidazole resistant (genospecies 5–11) displayed resistance greater than 256 μg/mL. These two strikingly different phenotypic patterns likely suggest multiple mechanisms through which metronidazole resistance is conferred in these species, one which leads to the strains eventually succumbing to a moderately high dose of the antibiotic, and another that does not. It therefore remains an ongoing effort to understand the full mechanistic reasons behind the resistance of these different genospecies to metronidazole treatment.

The observation of strikingly varied resistance response profiles, coupled with the confirmation that strains in genospecies 5–11 were highly susceptible to clindamycin, motivated us to design a simple PCR-based assay for detection of strains in the highly resistant genospecies. Primers have previously been designed to identify the GV genus through amplification of 16S or 23S rRNA, *cpn60*, or sialidase A, and a multiplex PCR has been designed that successfully distinguishes between the four original clades described by Ahmed et al. [29, 31, 32, 112–116]. This paper itself designed a sixteen-primer distributed-gene PCR that can identify members of each of the original clades. However, only seventeen strains were available during the design of the original PCR, which did not provide a sample size necessary to capture the full diversity of *G. vaginalis* spp. While it is informative, it does not differentiate between the higher-resolution clades, nor does it provide information about specifically metronidazole-resistant genospecies. Our new assay builds on the existing work by adding primers that are specific for antibiotic-resistant strains or species, in addition to those that distinguish the genus or specific genospecies.

In this study, 82% of all clinical isolates tested were metronidazole resistant, with highly resistant strains comprising 40% of the total. Development of a molecular diagnostic to identify which genospecies a patient's GV strain belongs to could change the current treatment modality for BV from guesswork to a more targeted approach, providing rapid relief for millions of patients who presently receive an antibiotic that may or may not treat their symptoms. Thirteen candidate genes were present in all seven highly resistant genospecies and absent in the remaining four mixed response/low resistance genospecies. Seven candidates were suitable for primer design, but only the *aruH* gene was consistently amplified across most of the resistant genospecies. *aruH* is involved in arginine catabolism and is associated with production of biogenic polyamines such as putrescine (KEGG: map00330) [117]. It is interesting to consider that the activity of *aruH* likely contributes to enhanced production of polyamines, which in turn provide these GV genospecies with the capability to increase the pH of the vaginal microbiome and lead to pathogenesis. The *aruH* gene is therefore informative of membership within the most highly resistant genospecies, and it would be reasonable to suggest that amplification of this gene in an unknown strain would suggest that the strain possesses metronidazole resistance. The *aruH* gene, however, is not necessarily responsible for metronidazole resistance, and further testing of the gene is necessary before making that claim. *aruH* is not present in any of the sporadically high-resistance strains in the metronidazole mixed-response genospecies 1–4; it therefore is not responsible for metronidazole resistance in those species. However, this observation, coupled with the evidence of multiple metronidazole response phenotypes, lends credibility to the hypothesis that metronidazole resistance can be conferred through different mechanisms. Given that our analysis does not find a single gene cluster that provides resistance in all the metronidazole-resistant strains, there is likely more than one way that resistance occurs, which could result in the varied resistance response phenotypes that are observed.

To pave the way for personalized diagnoses and treatments of BV, species differences must be rapidly and accurately identified. A PCR assay allows for the inexpensive classification of GV into two superclades/genera and can be used to rapidly rule out the prescription of metronidazole, one of the most common antibiotics used in BV treatment. This assay does have some limitations. It cannot currently identify low-resistance strains in genospecies 1a, 2, 3, or 4a/4b, and does not amplify the single strain in genospecies 9 which has a low prevalence and is phylogenetically distant from all other members of the genus. We also acknowledge that while the results we have achieved across genospecies are consistent, the robustness of our dataset could be improved with the acquisition of additional strains in several of the smaller genospecies. Future goals include expansion of our diagnostic set to discriminate between low-level resistant, intermediate response, or sensitive strains within the mixed response genospecies; inclusion of additional

antibiotic response patterns to our panel; and identification of genes responsible for the mechanisms of metronidazole resistance within this genus. This research led to a simple molecular diagnostic to identify highly metronidazole-resistant strains in a clinical setting and present an opportunity for strain-specific antibiotic treatment for patients with BV.

Conclusions

These results corroborate previous work recommending the reclassification of *Gardnerella vaginalis* spp. into a genus composed of multiple species [31, 35, 37]. Pangenomic analyses suggest a minimum of eleven genospecies are contained within the GV genus. Analysis of metronidazole susceptibility on GV clinical isolates identified the existence of a set of closely related genospecies with significantly higher resistance profiles (> 256 μg/mL) to metronidazole than other genospecies within the genus. Notably, all tested isolates were susceptible to clindamycin at clinically relevant levels (< 2 μg/mL), which may be cause for a shift in the current BV treatment paradigm. Lastly, a molecular diagnostic was designed which successfully amplifies a portion of the *aruH* gene in the highly metronidazole-resistant strains, allowing for easier identification of strains within those genospecies and opening the door to personalized medicine approaches for treating Bacterial Vaginosis.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-025-01446-4.

Additional file 1. Supplementary Tables. This file contains all supplementary tables, S1-S6. Table S1. GV strains downloaded from NCBI. Table S2. GV strains sequenced for this study. Table S3. Primers for diversity PCR screening. Table S4. Results of diversity PCR screening. Table S5. Metronidazole and clindamycin MIC results. Table S6. Candidate gene annotations.

Additional file 2. Supplementary Figures. This file contains supplementary figures S1 and S2. Figure S1. BlastP identity changes. Figure S2. Results of gradient PCR for *aruH* primer set.

Acknowledgements

Some of the contents of this manuscript have previously appeared in the first author's PhD dissertation [118]. KAI would like to thank Francisca Alarcon, Angie Albarouki, and Michael J. Tan for their assistance with performing the molecular diagnostic component of this work, and Kayla M. Socarras for her assistance with lab assays and proofreading.

Authors' contributions

All authors read and have approved the final manuscript. Conceptualization and Writing – Original Draft: K.A.I, J.P.E, and G.D.E; Methodology: K.A.I, J.P.E, J.C.M, S.L.H, and G.D.E; Software: J.P.E and R.L.E; Validation: K.A.I, E.P, and A.C.R; Resources: R.L.E, A.A, B.S, S.B, and S.L.H; Formal Analysis: K.A.I, S.C.B, J.A, E.P, and A.C.R; Investigation: K.A.I, S.C.B, J.A, A.G, E.P, A.R, A.A, and B.S; Data Curation: K.A.I, J.P.E, R.L.E, E.P, and A.R; Visualization: K.A.I and J.P.E; Writing—Review and Editing: K.A.I, J.P.E, R.L.E, A.G, S.B, J.C.M, S.L.H, and G.D.E; Supervision: J.P.E, J.C.M, and G.D.E; Funding Acquisition: G.D.E., K.A.I.

Funding

This work was supported by funds from the Institute of Molecular Medicine and Infectious Diseases at Drexel University and the Mary Dewitt Pettit, MD Fellowship. Additional funding was provided by the Oskar Fischer Project (a James Truchard Philanthropy), the Bill and Marian Cook Foundation, and NIH grants DC01428 and DK082316.

Data availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Garth D. Ehrlich (ge33@drexel.edu).

Materials availability

All unique resources (*G. vaginalis* spp. isolates) generated in this study may be available from the lead contact under a material transfer agreement.

Data and code availability

The datasets generated during and/or analyzed during the current study are available in DDBJ/ENA/GenBank under individual accession numbers which can be found in Additional file 1: Table S2. The dataset supporting the conclusions of this article is available under BioProject PRJNA1071662 (https://www.ncbi.nlm.nih.gov/bioproject/1071662)

This work did not use or generate new code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Sequences downloaded from NCBI can be found by accession number in Additional file 1: Table S1 [31, 35, 62–67].

Declarations

Ethics approval and consent to participate

The *Gardnerella vaginalis* spp. isolates included in this study were made available from the biorepository of one of the authors of this study (SLH) at the University of Pittsburgh and Magee-Women's Research Institute. The bacterial isolates were obtained from individuals enrolled in nine different clinical trials or observational cohort studies. All participants provided written informed consent prior to the collection of their samples and provided future use consent for use of their samples or sample remnants. The use of bacterial isolates in the present study does not constitute human subjects research because neither the repository holder nor the recipient of the isolates can link the isolates to any identifiable living individual. All links between the study participants, the bacterial isolates derived from their samples and their personal identifiers have been destroyed.

Consent for publication

Not applicable.

Competing interests

Authors KAI, JPE, and GDE have filed for a patent on the molecular diagnostic/primer sequences noted in the study. The remaining authors declare that they have no competing interests.

Author details

[1]Center for Advanced Microbial Processing, Drexel University College of Medicine, Philadelphia, PA, USA. [2]Center for Genomic Sciences, Drexel University College of Medicine, Philadelphia, PA, USA. [3]Department of Microbiology and Immunology, Drexel University College of Medicine, 245 N 15th St, Rm 5110, Philadelphia, PA, USA. [4]Magee-Women's Hospital of University of Pittsburgh Medical Center, Pittsburgh, PA, USA. [5]Department of Otolaryngology - Head and Neck Surgery, Drexel University College of Medicine, Philadelphia, PA, USA. [6]Center for Genomic Sciences, Allegheny Singer Research Institute, Pittsburgh, PA, USA. [7]University of Pittsburgh, Pittsburgh, PA, USA. [8]Earl Consulting, LLC, 3631 Wallace Street, Philadelphia, PA 19104, USA.

Received: 24 May 2024 Accepted: 25 February 2025
Published online: 25 March 2025

References

1. Bradshaw CS, Sobel JD. Current treatment of bacterial vaginosis-limitations and need for innovation. J Infect Dis. 2016;214 Suppl 1:S14–20. https://doi.org/10.1093/infdis/jiw159.
2. Hillier SL. Diagnostic microbiology of bacterial vaginosis. Am J Obstet Gynecol. 1993;169:455–9. https://doi.org/10.1016/0002-9378(93)90340-o.
3. Sweet RL. Role of bacterial vaginosis in pelvic inflammatory disease. Clin Infect Dis. 1995;20(Suppl 2):S271–5. https://doi.org/10.1093/clinids/20.supplement_2.s271.
4. Soper DE. Bacterial vaginosis and postoperative infections. Am J Obstet Gynecol. 1993;169:467–9. https://doi.org/10.1016/0002-9378(93)90343-h.
5. Gravett MG, Hummel D, Eschenbach DA, Holmes KK. Preterm labor associated with subclinical amniotic fluid infection and with bacterial vaginosis. Obstet Gynecol. 1986;67:229–37. https://doi.org/10.1097/00006250-198602000-00013.
6. Watts DH, Krohn MA, Hillier SL, Eschenbach DA. Bacterial vaginosis as a risk factor for post-cesarean endometritis. Obstet Gynecol. 1990;75:52–8.
7. Schwebke JR. Gynecologic consequences of bacterial vaginosis. Obstet Gynecol Clin North Am. 2003;30:685–94. https://doi.org/10.1016/s0889-8545(03)00086-x.
8. Muzny CA, Schwebke JR. Pathogenesis of bacterial vaginosis: discussion of current hypotheses. J Infect Dis. 2016;214 Suppl 1:S1–5. https://doi.org/10.1093/infdis/jiw121.
9. Koumans EH, Sternberg M, Bruce C, McQuillan G, Kendrick J, Sutton M, Markowitz LE. The prevalence of bacterial vaginosis in the United States, 2001–2004; associations with symptoms, sexual behaviors, and reproductive health. Sex Transm Dis. 2007;34:864–9. https://doi.org/10.1097/OLQ.0b013e318074e565.
10. Center for Disease Control. "Bacterial Vaginosis." Sexually Transmitted Infections Treatment Guidelines. 2021. https://www.cdc.gov/std/bv/stats.htm.
11. Núñez JT, Gómez G. Low-dose secnidazole in the treatment of bacterial vaginosis. Int J Gynaecol Obstet. 2005;88:281–5. https://doi.org/10.1016/j.ijgo.2004.11.028.
12. Eschenbach DA. Bacterial vaginosis: resistance, recurrence, and/or reinfection? Clin Infect Dis. 2007;44:220–1. https://doi.org/10.1086/509584.
13. Jones A. Bacterial vaginosis: a review of treatment, recurrence, and disparities. J Nurse Pract. 2019;15:420–3. https://doi.org/10.1016/j.nurpra.2019.03.010.
14. Schwebke JR, Desmond RA. A randomized trial of the duration of therapy with metronidazole plus or minus azithromycin for treatment of symptomatic bacterial vaginosis. Clin Infect Dis. 2007;44:213–9. https://doi.org/10.1086/509577.
15. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011;108(Suppl 1):4680–7. https://doi.org/10.1073/pnas.1002611107.
16. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, et al. A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. Int J Syst Evol Microbiol. 2020;70:2782–858. https://doi.org/10.1099/ijsem.0.004107.
17. Onderdonk AB, Delaney ML, Fichorova RN. The human microbiome during bacterial vaginosis. Clin Microbiol Rev. 2016;29:223–38. https://doi.org/10.1128/CMR.00075-15.
18. Fettweis, J.M., Brooks, J.P., Serrano, M.G., Sheth, N.U., Girerd, P.H., Edwards, D.J., Strauss, J.F., The Vaginal Microbiome Consortium, Jefferson, K.K., and Buck, G.A. Differences in vaginal microbiome in African American women versus women of European ancestry. Microbiology. 2014;160:2272–82. https://doi.org/10.1099/mic.0.081034-0.
19. Nelson TM, Borgogna J-LC, Brotman RM, Ravel J, Walk ST, Yeoman CJ. Vaginal biogenic amines: biomarkers of bacterial vaginosis or precursors to vaginal dysbiosis? Front Physiol. 2015;6: 253. https://doi.org/10.3389/fphys.2015.00253.
20. Eschenbach DA, Hillier S, Critchlow C, Stevens C, DeRouen T, Holmes KK. Diagnosis and clinical manifestations of bacterial vaginosis. Am J Obstet Gynecol. 1988;158:819–28. https://doi.org/10.1016/0002-9378(88)90078-6.
21. Gardner HL, Dukes CD. Haemophilus vaginalis vaginitis: a newly defined specific infection previously classified "nonspecific" vaginitis. Am J Obstet Gynecol. 1955;69:962–76. https://doi.org/10.1016/0002-9378(55)90095-8.
22. Leopold S. Heretofore undescribed organism isolated from the genitourinary system. U S Armed Forces Med J. 1953;4:263–6.
23. Nisha K, Antony B, Udayalaxmi J. Comparative analysis of virulence factors & biotypes of Gardnerella vaginalis isolated from the genital tract of women with & without bacterial vaginosis. Indian J Med Res. 2019;149:57–61. https://doi.org/10.4103/ijmr.IJMR_1674_16.
24. Patterson JL, Stull-Lane A, Girerd PH, Jefferson KK. Analysis of adherence, biofilm formation and cytotoxicity suggests a greater virulence potential of Gardnerella vaginalis relative to other bacterial-vaginosis-associated anaerobes. 2010. Preprint. https://doi.org/10.1099/mic.0.034280-0.
25. Zinnemann K, Turner GC. The taxonomic position of "Haemophilus vaginalis" [Corynebacterium vaginale]. J Pathol Bacteriol. 1963;85:213–9. https://doi.org/10.1002/path.1700850120.
26. 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;30:170–8. https://doi.org/10.1099/00207713-30-1-170.
27. Piot P, Van Dyck E, Peeters M, Hale J, Totten PA, Holmes KK. Biotypes of Gardnerella vaginalis. J Clin Microbiol. 1984;20:677–9. https://doi.org/10.1128/jcm.20.4.677-679.1984.
28. Benito R, Vazquez JA, Berron S, Fenoll A, Saez-Neito JA. A modified scheme for biotyping Gardnerella vaginalis. J Med Microbiol. 1986;21:357–9. https://doi.org/10.1099/00222615-21-4-357.
29. Ingianni A, Petruzzelli S, Morandotti G, Pompei R. Genotypic differentiation of Gardnerella vaginalis by amplified ribosomal DNA restriction analysis (ARDRA). FEMS Immunol Med Microbiol. 1997;18:61–6. https://doi.org/10.1111/j.1574-695X.1997.tb01028.x.
30. Paramel Jayaprakash T, Schellenberg JJ, Hill JE. Resolution and characterization of distinct cpn60-based subgroups of Gardnerella vaginalis in the vaginal microbiota. PLoS One. 2012;7: e43009. https://doi.org/10.1371/journal.pone.0043009.
31. Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, Powell E, Janto B, Eutsey R, Hiller NL, 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;194:3922–37. https://doi.org/10.1128/JB.00056-12.
32. Balashov SV, Mordechai E, Adelson ME, Gygax SE. Identification, quantification and subtyping of Gardnerella vaginalis in noncultured clinical vaginal samples by quantitative PCR. J Med Microbiol. 2014;63:162–75. https://doi.org/10.1099/jmm.0.066407-0.
33. Schellenberg JJ, Paramel Jayaprakash T, Withana Gamage N, Patterson MH, Vaneechoutte M, Hill JE. Gardnerella vaginalis subgroups defined by cpn60 sequencing and sialidase activity in isolates from Canada, Belgium and Kenya. PLoS One. 2016;11:e0146510. https://doi.org/10.1371/journal.pone.0146510.
34. Schuyler JA, Mordechai E, Adelson ME, Sobel JD, Gygax SE, Hilbert DW. Identification of intrinsically metronidazole-resistant clades of Gardnerella vaginalis. Diagn Microbiol Infect Dis. 2016;84:1–3. https://doi.org/10.1016/j.diagmicrobio.2015.10.006.
35. Vaneechoutte M, Guschin A, Van Simaey L, 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(3):679–87.
36. Potter RF, Burnham C-AD, Dantas G. In silico analysis of gardnerella genomospecies detected in the setting of bacterial vaginosis. Clin Chem. 2019;65:1375–87. https://doi.org/10.1373/clinchem.2019.305474.
37. 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;73. https://doi.org/10.1099/ijsem.0.006140.

38. Shen K, Wang X, Post JC, Ehrlich GD. Molecular and Translational Research Approaches for the study of Bacterial Pathogenesis in Otitis Media. In: Rosenfeld R, Bluestone CD, editors. Evidence-based Otitis Media. Hamilton: B.C. Decker Inc; 2003. p. 91–119. 2.
39. Ehrlich GD, Hu FZ, Post JC. Role for biofilms in infectious disease. In: Microbial biofilms. American Society of Microbiology; 2004. p. 332–358. https://doi.org/10.1128/9781555817718.ch18.
40. Ehrlich GD, Hu FZ, Shen K, Stoodley P, Post JC. Bacterial plurality as a general mechanism driving persistence in chronic infections. Clin Orthop Relat Res. 2005;20–24:20. https://doi.org/10.1097/00003086-200508000-00005.
41. Ehrlich GD, Ahmed A, Earl J, Hiller NL, Costerton JW, Stoodley P, Post JC, DeMeo P, Hu FZ. The distributed genome hypothesis as a rubric for understanding evolution in situ during chronic bacterial biofilm infectious processes. FEMS Immunol Med Microbiol. 2010;59:269–79. https://doi.org/10.1111/j.1574-695X.2010.00704.x.
42. Shen K, Antalis P, Gladitz J, Sayeed S, Ahmed A, Yu S, Hayes J, Johnson S, Dice B, Dopico R, et al. Identification, distribution, and expression of novel genes in 10 clinical isolates of nontypeable Haemophilus influenzae. Infect Immun. 2005;73:3479–91. https://doi.org/10.1128/IAI.73.6.3479-3491.2005.
43. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Scott Durkin A, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome." Proc Natl Acad Sci U S A. 2005;102:13950–5. https://doi.org/10.1073/pnas.0506758102.
44. Shen K, Sayeed S, Antalis P, Gladitz J, Ahmed A, Dice B, Janto B, Dopico R, Keefe R, Hayes J, et al. Extensive genomic plasticity in Pseudomonas aeruginosa revealed by identification and distribution studies of novel genes among clinical isolates. Infect Immun. 2006;74:5272–83. https://doi.org/10.1128/IAI.00546-06.
45. Erdos G, Sayeed S, Hu FZ, Antalis PT, Shen K, Hayes JD, Ahmed AI, Johnson SL, Post JC, Ehrlich GD. Construction and characterization of a highly redundant Pseudomonas aeruginosa genomic library prepared from 12 clinical isolates: application to studies of gene distribution among populations. Int J Pediatr Otorhinolaryngol. 2006;70:1891–900. https://doi.org/10.1016/j.ijporl.2006.06.016.
46. Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD. Characterization and modeling of the Haemophilus influenzae core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol. 2007;8: R103. https://doi.org/10.1186/gb-2007-8-6-r103.
47. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, Keefe R, Ehrlich NE, Shen K, Hayes J, et al. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol. 2007;189:8186–95. https://doi.org/10.1128/JB.00690-07.
48. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, et al. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol. 2010;11: R107. https://doi.org/10.1186/gb-2010-11-10-r107.
49. Boissy R, Ahmed A, Janto B, Earl J, Hall BG, Hogg JS, Pusch GD, Hiller LN, Powell E, Hayes J, et al. Comparative supragenomic analyses among the pathogens Staphylococcus aureus, Streptococcus pneumoniae, and Haemophilus influenzae using a modification of the finite supragenome model. BMC Genomics. 2011;12: 187. https://doi.org/10.1186/1471-2164-12-187.
50. Davie JJ, Earl J, de Vries SPW, Ahmed A, Hu FZ, Bootsma HJ, Stol K, Hermans PWM, Wadowsky RM, Ehrlich GD, et al. Comparative analysis and supragenome modeling of twelve Moraxella catarrhalis clinical isolates. BMC Genomics. 2011;12: 70. https://doi.org/10.1186/1471-2164-12-70.
51. Earl JP, de Vries SPW, Ahmed A, Powell E, Schultz MP, Hermans PWM, Hill DJ, Zhou Z, Constantinidou CI, Hu FZ, et al. Comparative genomic analyses of the Moraxella catarrhalis serosensitive and seroresistant lineages demonstrate their independent evolution. Genome Biol Evol. 2016;8:955–74. https://doi.org/10.1093/gbe/evw039.
52. Nistico L, Earl J, Hiller L, Ahmed A, Retchless A, Janto B, et al. Using the core and supra genomes to determine diversity and natural proclivities among bacterial strains. Appl Mol Microbiol Methods. 2014:105–20.
53. Innamorati KA, Earl JP, Aggarwal SD, Ehrlich GD, Hiller NL The bacterial guide to designing a diversified gene portfolio. In: The Pangenome. Springer International Publishing; 2020. p. 51–87. https://doi.org/10.1007/978-3-030-38281-0_3.
54. Hammond JA, Gordon EA, Socarras KM, Chang Mell J, Ehrlich GD. Beyond the pan-genome: current perspectives on the functional and practical outcomes of the distributed genome hypothesis. Biochem Soc Trans. 2020;48:2437–55. https://doi.org/10.1042/BST20190713.
55. Otani H, Udwary DW, Mouncey NJ. Comparative and pangenomic analysis of the genus Streptomyces. Sci Rep. 2022;12:18909. https://doi.org/10.1038/s41598-022-21731-1.
56. Kim M, Cha I-T, Lee K-E, Li M, Park S-J. Pangenome analysis provides insights into the genetic diversity, metabolic versatility, and evolution of the genus Flavobacterium. Microbiol Spectr. 2023;e0100323: e0100323. https://doi.org/10.1128/spectrum.01003-23.
57. Socarras KM, Haslund-Gourley BS, Cramer NA, Comunale MA, Marconi RT, Ehrlich GD. Large-scale sequencing of Borreliaceae for the construction of pan-genomic-based diagnostics. Genes (Basel). 2022;13: 1604. https://doi.org/10.3390/genes13091604.
58. Rajput A, Chauhan SM, Mohite OS, Hyun JC, Ardalani O, Jahn LJ, Sommer MO, Palsson BO. Pangenome analysis reveals the genetic basis for taxonomic classification of the Lactobacillaceae family. Food Microbiol. 2023;115: 104334. https://doi.org/10.1016/j.fm.2023.104334.
59. Tsalik EL, Bonomo RA, Fowler VG Jr. New molecular diagnostic approaches to bacterial infections and antibacterial resistance. Annu Rev Med. 2018;69:379–94. https://doi.org/10.1146/annurev-med-052716-030320.
60. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Curr Opin Microbiol. 2015;23:148–54. https://doi.org/10.1016/j.mib.2014.11.016.
61. Kress-Bennett JM, Hiller NL, Eutsey RA, Powell E, Longwell MJ, Hillman T, Blackwell T, Byers B, Mell JC, Post JC, et al. Identification and characterization of msf, a novel virulence factor in Haemophilus influenzae. PLoS ONE. 2016;11: e0149891. https://doi.org/10.1371/journal.pone.0149891.
62. Oshima K, Hisamatsu S, Toh H, Nakano A, Kiuchi M, Kuroyanagi H, Morita H, Hattori M. Complete genome sequence of Gardnerella vaginalis strain JCM 11026T, isolated from vaginal tracts of women. Genome Announc. 2015;3: e00286-15. https://doi.org/10.1128/genomeA.00286-15.
63. Yeoman CJ, Yildirim S, Thomas SM, Durkin AS, Torralba M, Sutton G, Buhay CJ, Ding Y, Dugan-Rocha SP, Muzny DM, et al. Comparative genomics of Gardnerella vaginalis strains reveals substantial differences in metabolic and virulence potential. PLoS ONE. 2010;5: e12411. https://doi.org/10.1371/journal.pone.0012411.
64. Bulavaitė A, Maier T, Pleckaityte M. Discrimination of Gardnerella species by combining MALDI-TOF protein profile, chaperonin cpn60 sequences, and phenotypic characteristics. Pathogens. 2021;10: 277. https://doi.org/10.3390/pathogens10030277.
65. Dröge J, Gregor I, McHardy AC. Taxator-tk: precise taxonomic assignment of metagenomes by fast approximation of evolutionary neighborhoods. Bioinformatics. 2015;31:817–24. https://doi.org/10.1093/bioinformatics/btu745.
66. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
67. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108:1513–8. https://doi.org/10.1073/pnas.1017351108.
68. Schuyler JA, Chadwick SG, Mordechai E, Adelson ME, Gygax SE, Hilbert DW. Draft genome sequence of a metronidazole-resistant Gardnerella vaginalis isolate. Genome Announc. 2015;3: e00992-15. https://doi.org/10.1128/genomeA.00992-15.
69. 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;6. https://doi.org/10.1128/mSphere.00154-21.
70. Harwich MD Jr, Alves JM, Buck GA, Strauss JF 3rd, Patterson JL, Oki AT, Girerd PH, Jefferson KK. Drawing the line between commensal and pathogenic Gardnerella vaginalis through genome analysis and

virulence studies. BMC Genomics. 2010;11: 375. https://doi.org/10.1186/1471-2164-11-375.
71. Lee AH-Y, Flibotte S, Sinha S, Paiero A, Ehrlich RL, Balashov S, Ehrlich GD, Zlosnik JEA, Mell JC, Nislow C. Phenotypic diversity and genotypic flexibility of Burkholderia cenocepacia during long-term chronic infection of cystic fibrosis lungs. Genome Res. 2017;27:650–62. https://doi.org/10.1101/gr.213363.116.
72. Liu B, Yuan J, Yiu S-M, Li Z, Xie Y, Chen Y, Shi Y, Zhang H, Li Y, Lam T-W, et al. COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. Bioinformatics. 2012;28:2870–4. https://doi.org/10.1093/bioinformatics/bts563.
73. Boisvert S, Laviolette F, Corbeil J. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol. 2010;17:1519–33. https://doi.org/10.1089/cmb.2009.0238.
74. Moleres J, Fernández-Calvet A, Ehrlich RL, Martí S, Pérez-Regidor L, Euba B, Rodríguez-Arce I, Balashov S, Cuevas E, Liñares J, et al. Antagonistic pleiotropy in the bifunctional surface protein FadL (OmpP1) during adaptation of Haemophilus influenzae to chronic lung infection associated with chronic obstructive pulmonary disease. MBio. 2018;9:e01176–218.
75. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9. https://doi.org/10.1093/bioinformatics/btu153.
76. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3. https://doi.org/10.1093/bioinformatics/btv421.
77. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods. 2015;8:12–24. https://doi.org/10.1039/C5AY02550H.
78. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. https://www.R-project.org/.
79. Wickham H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org. ISBN 978-3-319-24277-4.
80. Da Silva G. harrietr: Wrangle phylogenetic distance matrices and other utilities. R package version 0.2.3. 2017. https://CRAN.R-project.org/package=harrietr.
81. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;1079:155–70. https://doi.org/10.1007/978-1-62703-646-7_10.
82. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3. https://doi.org/10.1093/bioinformatics/btu033.
83. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. Ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. https://doi.org/10.1111/2041-210x.12628.
84. Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinformatics. 2020;69: e96. https://doi.org/10.1002/cpbi.96.
85. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14:1394–403. https://doi.org/10.1101/gr.2289704.
86. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31. https://doi.org/10.1016/j.jmb.2015.11.006.
87. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. https://doi.org/10.1016/s0022-2836(05)80360-2.
88. Wilkins D. _gggenes: Draw gene arrow maps in "ggplot2". CRAN; 2020. https://CRAN.R-project.org/package=gggenes.
89. Faught BM, Reyes S. Characterization and treatment of recurrent bacterial vaginosis. J Womens Health (Larchmt). 2019;28:1218–26. https://doi.org/10.1089/jwh.2018.7383.
90. Ahmed-Jushuf IH, Shahmanesh M, Arya OP. The treatment of bacterial vaginosis with a 3 day course of 2% clindamycin cream: results of a multicentre, double blind, placebo controlled trial. B V Investigators Group Genitourin Med. 1995;71:254–6. https://doi.org/10.1136/sti.71.4.254.
91. Clinical and Laboratory Standards Institute (CLSI). Methods for antimicrobial susceptibility testing of anaerobic bacteria. 9th ed. CLSI standard M11 (ISBN 978-1-68440-021-8 [Print]: ISBN 978-1-68440-022-5 [Electronic]). Wayne, Pennsylvania: Clinical and Laboratory Standards Institute; 2018. https://clsi.org/media/2577/m11-ed9_sample.pdf.
92. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 2016;17:238. https://doi.org/10.1186/s13059-016-1108-8.
93. Dunn OJ. Multiple comparisons among means. J Am Stat Assoc. 1961;56:52–64. https://doi.org/10.1080/01621459.1961.10482090.
94. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
95. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160–6. https://doi.org/10.1093/bib/bbx108.
96. Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 2013;41:W22–8. https://doi.org/10.1093/nar/gkt389.
97. Wright ES, Yilmaz LS, Ram S, Gasser JM, Harrington GW, Noguera DR. Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates. Environ Microbiol. 2014;16:1354–65. https://doi.org/10.1111/1462-2920.12259.
98. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1. https://doi.org/10.1093/bioinformatics/btq461.
99. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005;102:2567–72. https://doi.org/10.1073/pnas.0409727102.
100. Bohr LL, Mortimer TD, Pepperell CS. Lateral gene transfer shapes diversity of Gardnerella spp. Front Cell Infect Microbiol. 2020;10: 293. https://doi.org/10.3389/fcimb.2020.00293.
101. 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;87: e02188-20. https://doi.org/10.1128/AEM.02188-20.
102. Petrina MAB, Cosentino LA, Rabe LK, Hillier SL. Susceptibility of bacterial vaginosis (BV)-associated bacteria to secnidazole compared to metronidazole, tinidazole and clindamycin. Anaerobe. 2017;47:115–9. https://doi.org/10.1016/j.anaerobe.2017.05.005.
103. Hindler JA, Schuetz EAN. CLSI AST news update. https://clsi.org/media/3mrkwh4j/ast-news-update-volume-6-issue-1-april-2021_5-17-21.pdf.
104. Yang Z, Lu C-D. Characterization of an arginine: Pyruvate transaminase in arginine catabolism of *Pseudomonas aeruginos*a PAO1. J Bacteriol. 2007;189:3954–9. https://doi.org/10.1128/jb.00262-07.
105. Barrientos-Durán A, Fuentes-López A, de Salazar A, Plaza-Díaz J, García F. Reviewing the composition of vaginal microbiota: inclusion of nutrition and probiotic factors in the maintenance of eubiosis. Nutrients. 2020;12: 419. https://doi.org/10.3390/nu12020419.
106. Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. Nat Rev Microbiol. 2008;6:431–40. https://doi.org/10.1038/nrmicro1872.
107. Georgiades K, Raoult D. Defining pathogenic bacterial species in the genomic era. Front Microbiol. 2010;1: 151. https://doi.org/10.3389/fmicb.2010.00151.
108. Cohan FM. What are bacterial species? Ann Rev Microbiol. 2002. https://doi.org/10.1146/annurev.micro.56.012302.160634.
109. Moreno E. In search of a bacterial spedes definition. Revista de Biología Tropical. 1997:753–771.
110. Rosselló-Mora R, Amann R. The species concept for prokaryotes. FEMS Microbiol Rev. 2001;25:39–67. https://doi.org/10.1111/j.1574-6976.2001.tb00571.x.
111. Chang JC, Wilkening RV, Rahbari KM, Federle MJ. Quorum sensing regulation of a major facilitator superfamily transporter affects multiple streptococcal virulence factors. J Bacteriol. 2022;204: e00176-22. https://doi.org/10.1128/jb.00176-22.
112. Zariffard MR, Saifuddin M, Sha BE, Spear GT. Detection of bacterial vaginosis-related organisms by real-time PCR for Lactobacilli, Gardnerella

vaginalis and Mycoplasma hominis. FEMS Immunol Med Microbiol. 2002;34:277–81. https://doi.org/10.1111/j.1574-695X.2002.tb00634.x.
113. Menard J-P, Fenollar F, Henry M, Bretelle F, Raoult D. Molecular quantification of Gardnerella vaginalis and Atopobium vaginae loads to predict bacterial vaginosis. Clin Infect Dis. 2008;47:33–43. https://doi.org/10.1086/588661.
114. Santiago GLDS, Deschaght P, El Aila N, Kiama TN, Verstraelen H, Jefferson KK, Temmerman M, Vaneechoutte M. Gardnerella vaginalis comprises three distinct genotypes of which only two produce sialidase. Am J Obstet Gynecol. 2011;204(450):e1–7. https://doi.org/10.1016/j.ajog.2010.12.061.
115. Higashide S, Cho O, Matsuda Y, Ogishima D, Kurakado S, Sugita T. Loop-mediated isothermal amplification for the rapid detection of Gardnerella vaginalis. Microbiol Immunol. 2018;62:607–11.
116. Shipitsyna E, Krysanova A, Khayrullina G, Shalepo K, Savicheva A, Guschin A, Unemo M. Quantitation of all four Gardnerella vaginalis clades detects abnormal vaginal microbiota characteristic of bacterial vaginosis more accurately than putative G. vaginalis sialidase A gene count. Mol Diagn Ther. 2019;23:139–47. https://doi.org/10.1007/s40291-019-00382-5.
117. Yang F, Qiu K, Zhu Y, Zhang X, Yang T, Yi G, Xu M, Rao Z. Production of putrescine in metabolic engineering *Corynebacterium crenatum* by mixed sugar fermentation. ACS Sustain Chem Eng. 2022;10:14407–16. https://doi.org/10.1021/acssuschemeng.2c02602.
118. Innamorati K. Comparative genomics of the keystone mucosal pathogens Porphyromonas gingivalis and Gardnerella vaginalis reveal associations between virulence phenotypesand phylogenetic structures. PhD Dissertation. Drexel University; 2021. Available at: https://researchdiscovery.drexel.edu/esploro/outputs/doctoral/991014972849304721.

Publisher's Note