Host and Bacterial Functions in Bacterial Vaginosis Elucidated by Metaproteomics

Elliot M. Lee

A dissertation
submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading Committee:
David Fredricks, Chair
W. Conrad Liles
Christopher Johnston

Program authorized to offer degree:
Microbiology

University of Washington

Abstract

Host and Bacterial Functions in Bacterial Vaginosis Elucidated by Metaproteomics

Elliot M. Lee

Chair of Supervisory Committee:
David Fredricks
Department of Microbiology

Bacterial vaginosis (BV) is a highly prevalent dysbiosis of the vaginal microbiota which causes a variety of unpleasant symptoms and places patients at higher risk of adverse sequelae. BV is a complex condition, and many of the host and bacterial functions which contribute to its development and recurrence are unknown. This dissertation describes an optimized metaproteomic analysis of cervicovaginal lavage samples from women with and without BV to identify host and bacterial proteins that may contribute to BV. Using this approach, we uncovered new potential synergistic interactions between BV-associated bacteria (BVAB) based on glutamate and identified a possible host response to increased concentrations of free heme in BV. We also demonstrated a novel syntrophic interaction between *Dialister micraerophilus* and *Fannyhessea vaginae* to increase putrescine biosynthesis, likely through cross-feeding of the arginine metabolite ornithine. Despite past reports that human amylase is primarily responsible for breaking down vaginal glycogen into fermentable carbohydrates, we identified glycogen-degrading enzymes from *L. crispatus* and *G. vaginalis* in samples from both BV- and BV+ study participants. This observation led to our discovery that a wide range of BVAB, but only *L. crispatus* and *L. iners* among the commensal *Lactobacillus* spp., can directly metabolize glycogen. Finally, we described the construction of an *E. coli-Gardnerella* shuttle vector that can be applied for genetic manipulation of this genus. This work contains novel insights into BV and opens new avenues to study the biology of vaginal bacteria, with implications for treatment and prevention of this condition.

Table of Contents

Acknowledgements

I'd first like to thank my mentor Professor David Fredricks for giving me the freedom to pursue my interests and trusting me to explore all the exciting rabbit holes I found in our data. I'm grateful for his many thoughtful critiques that have made me a better scientist and communicator, and all our delightful conversations in his office poring over new data and speculating about the vaginal microbiota.

I also need to thank the members of the Fredricks Lab for sharing their incredible expertise and technical knowledge on all things molecular- and micro-biology. Thank you to Sujatha, Tina, DJ, Matt, Sean, Stephanie, Congzhou, Sue, and especially Susan for teaching me, then very patiently re-teaching me everything I know.

My most heartfelt thanks to the Microbiology grad students for their friendship which has made my five years of PhD the best of my life. Thanks to Lyndsey for always being available for a wine night, Ciara for always keeping the fun going longer, and Robin for the commiseration during late nights in lab. And many thanks to everyone else for making sure I always made it home.

I'd like to thank my collaborators for trusting me with their data and their invaluable expertise – Chris, Dakota, Elsa, Sam, Danijel, and Brooke.

And most importantly, I'd like to thank my parents, Nancy and Mike. I would not have gone half as far or become half the person I am without their endless love and support. Thank you for signing me up for every activity, supporting my every interest, and taking as much delight in my life as I do.

This work was funded by NIH grant R01AI061628 and Department of Energy contract DE-AC05-76RLO-1830. The funders had no role in study design, data collection and interpretation, or the decision to submit any part of this work for publication.

List of Tables and Figures

Chapter 1: Background

Vaginal Physiology

The vagina is the roughly four inch-long muscular organ that connects the exterior vulva to the interior cervix in the human female reproductive tract[1]. While bacterial load in the vagina is lower than other body sites such as the mouth or gut, large communities of bacteria colonize the vagina with concentrations up to 10^{10} 16S rRNA copies per vaginal swab[2]. These bacteria must contend with a unique environment which is shaped both by host physiology and their own metabolism.

The vaginal lumen is generally anaerobic, with normal oxygen concentrations around 1%[3]. The interior walls of the vagina are composed of cornified epithelial cells[4,5]. Unlike the outermost layers of the epidermis, the vaginal epithelium does not form a water-tight barrier, allowing fluid to leak into the vaginal lumen from capillaries, which constitutes some of the liquid portion of vaginal fluid[6]. Following puberty, the vaginal walls thicken to approximately 28 cell layers deep[7]. These cells turn over quickly, with basal cells reaching the apical layer in less than 96 hours[8]. The rapid turnover of epithelial cells contributes to one of the vagina's most important functions – excluding pathogens from the upper reproductive tract. Cervical cells also secrete mucus which helps clear microbes as it passes down the female reproductive tract through the vagina[9]. In addition to its viscid properties, this mucus contains antimicrobial peptides and immunoglobulins which trap potentially pathogenic organisms and inhibit their growth[10,11].

Acid is another major barrier for vaginal pathogens. Normal vaginal pH is less than 4.0, which is toxic to a wide range of bacteria, viruses, and parasites[12,13]. Commensal bacteria are thought to be the main source of this antimicrobial acid[14]. Unlike other types of cornified cells, vaginal epithelial cells are packed with glycogen[15]. Commensal lactobacilli consume carbohydrates released by the breakdown of this glycogen, fermenting them into organic acids

which acidify the vagina. In addition to bacterial fermentation, proton pumps on vaginal epithelial cells likely contribute to vaginal acidity by actively pumping protons into the vaginal lumen[16].

While the micro-environment experienced by vaginal bacteria is relatively stable, the vaginal environment is subject to unique perturbations. Most notably, protein- and iron-rich blood regularly passes through the vagina for roughly 5 days of an approximately 28 day menstrual cycle (with substantial variability in menses/cycle duration and blood volume between individuals)[17,18]. In addition to this change in available nutrients, menses can also neutralize vaginal pH for an extended period of time[19]. To a similar, though lesser extent, sexual contact can also perturb the vaginal environment. Semen from unprotected penile-vaginal sex introduces a large quantity of carbohydrates and protein into the vaginal lumen[20]. In order to protect sperm from acid, semen also has a high buffering capacity to neutralize vaginal pH[21]. Additionally, sexual contact between women who have sex with women (WSW) has the potential to transmit large numbers of potentially pathogenic microbes between partners[22]. All these factors have the potential to alter the environmental and microbiological landscape of the vagina.

The Vaginal Microbiota in Eubiosis

Humans, like all other animals, live in close association with microbes. These relationships are often mutually beneficial, with microbes carrying out important functions in exchange for nutrients and a felicitous environment[23]. Such “eubiotic” consortia of microbial organisms contribute positively to the health of their host, both in the taxonomic composition of their communities and the functions they perform[24]. This is in contrast to a “dysbiotic” community, where resident microbes carry out functions which negatively impact host health and contribute to disease.

In the vagina, a eubiotic community is characterized by dominance of *Lactobacillus* spp., often with a single species making up >50% of the bacteria present in the microbiota[25].

Approximately 90% of women with a eubiotic community have a vaginal microbiota dominated by either *L. crispatus* or *L. iners,* though *L. gasseri, L. jensenii, L. mulieris,* and *L. vaginalis* are also commonly present[26-29]. Lactobacilli are Gram-positive, rod-shaped bacteria with the ability to tolerate acidic pHs that would be deadly to many other organisms[30]. They primarily ferment carbohydrates into lactate, which is a relatively strong organic acid that can reduce the pH of their environment below 4.0[12,31,32]. This characteristic has made them invaluable in the production of diverse fermented food products, and also helps them contribute to the pathogen exclusion function of the vagina[33,34]. A low environmental pH kills diverse pathogens including *Neisseria gonorrhoeae, Chlamydia trachomatis,* and *Trichomonas vaginalis,* and inactivates HIV and HSV virions[35-40]. Additionally, lactic acid is more antimicrobial than a low pH alone[40,41]. It is unclear why lactic acid is especially antimicrobial, but past studies have shown the molecule strips protective lipopolysaccharide off the outer membrane of Gram-negative bacteria, and suggests the protonated form may be able to cross cellular membranes, directly acidifying a cell's cytoplasm[42,43]. Lactic acid also inhibits vaginal *Candida* switching from their benign yeast form to the pathogenic hyphae which cause vulvovaginal candidiasis[44].

Although lactate production is one of the primary means by which commensal lactobacilli promote vaginal health, some strains also produce antibacterial bacteriocins[45]. These small proteins likely act as pore-forming toxins which disrupt the membrane integrity of target organisms[46]. Additionally, some commensal lactobacilli encode bacteriolysins, enzymes which attack the cell walls of other bacteria[47]. While their antibacterial effect is not as dramatic as the lactic acid produced by *Lactobacillus* spp., these proteins have also been shown to inhibit the growth of vaginal bacteria associated with dysbiosis[48].

While eubiotic vaginal communities are characterized by dominance of *Lactobacillus*, not all species in this genus are equally beneficial. The vaginal microbiota is most stable when *L. crispatus* is the majority community member, whereas communities dominated by *L. iners* are

twice as likely to spontaneously shift to dysbiosis[49-51]. Although it is generally associated with health, *L. iners* is also present in approximately 86% of women with a dysbiotic vaginal microbiota[25]. Studies suggest these differences may be due to unique proteins expressed by *L. iners* such as the host-targeting cytotoxin inerolysin[52], but additional research is required to fully understand what microbiological characteristics make an optimal eubiotic vaginal microbiota.

Introduction to Bacterial Vaginosis

Bacterial vaginosis (BV) is the most common dysbiotic condition of the vaginal microbiota[53]. In BV, the homogenous, *Lactobacillus*-dominated bacterial biota shifts to a much more heterogenous state, with BV-associated bacteria (BVAB) causing unpleasant symptoms including excessive vaginal discharge and a characteristic "fishy" odor[54,55]. BV can be diagnosed clinically by fulfillment of Amsel criteria or in the laboratory by analyzing a Gram stain of vaginal fluid and calculating a Nugent score. The four Amsel criteria are 1) thin, homogenous vaginal discharge, 2) vaginal pH >4.5, 3) release of a rotten-fish odor when 10% potassium hydroxide is added to vaginal fluid, and 4) presence of clue cells in vaginal fluid – epithelial cells covered in a dense bacterial biofilm[56]. A patient is determined to have BV if they present with three or more Amsel criteria. Nugent scoring focuses on the bacteria present in an individual's vagina rather than symptoms. To determine Nugent score, a Gram stain is performed on a smear of vaginal fluid and the number of Gram-positive rods, small Gram-variable coccobacilli, and curved Gram-variable rods are counted[57]. A score is then assigned based on the relative number of these different morphotypes with 0 – 3 indicating no BV, 4 – 6 an intermediate microbiota, and 7 – 10 the presence of BV. Although Nugent score is the gold-standard for BV research, Amsel criteria are used clinically for diagnosis and treatment of BV.

BV is an extremely common condition. Although reports of prevalence differ based on the population sampled, estimates for the proportion of reproductive-aged women with BV as determined by Nugent score range from 10% – 50%[53,58-62]. The largest study of BV in the United

States found 29% of reproductive-aged women were BV+ by Nugent score, with substantial variation by race. 51% of study participants who identified as Black were BV+, while only 23% of non-Hispanic Whites were BV+[53]. BV also disproportionately affects other disadvantaged groups, with women who have sex with women (WSW) and women in poverty at greater risk for BV[63,64]. In addition to the unpleasant symptoms associated with condition, BV is associated with increased risk for various adverse sequelae including pelvic inflammatory disease, preterm birth, and acquisition of various STIs including HIV[65-70]. Although a high proportion of individuals with BV are asymptomatic (63% – 84% depending on the study[53,63]), even asymptomatic BV is associated with an increased risk of adverse sequelae. BV also has significant social-emotional impacts on women, with malodor being a primary cause of embarrassment and shame that negatively impact their social and sexual lives[71]. These feelings often lead women to try self-help remedies such as douching or using scented feminine hygiene products which can exacerbate BV[72-76].

BV is treated by oral or vaginal administration of antibiotics, most commonly clindamycin or metronidazole[77]. Antibiotic treatment is generally very effective, resolving ~90% of cases within one month[78]. Recurrence is extremely high, however, with approximately 60% of patients experiencing a recurrence of BV within one year of treatment[79]. Various supplemental treatments have been proposed to augment antibiotic therapy including supplemental lactic acid, boric acid, lactoferrin, and probiotic lactobacilli, but trials with these therapies have had little effect on rates of cure or recurrence[80-85]. More recently, transplants of vaginal fluid from healthy women have been proposed to treat BV, and a small, prospective study showed some promise to treat women with intractable BV[86]. However, even this treatment required multiple transplants in some patients, highlighting the need for a better understanding of the dynamics of the vaginal microbiota to improve treatments for BV.

The Microbiology and Etiology of Bacterial Vaginosis

BV is characterized by a diversification of the vaginal microbiota and a decreased relative abundance of lactobacilli[54]. Anaerobes like *Gardnerella, Prevotella, Sneathia,* and *Fannyhessea* increase in abundance, though usually no single species makes up the majority of the community. *Gardnerella vaginalis* was thought to be the causative agent of BV when the condition was first described in the 1950's[87]. However, subsequent studies (with dubious ethics) found that vaginal fluid from women with BV was much more likely to cause BV in a healthy individual than a pure culture of *G. vaginalis* alone, providing early evidence that a community of BV-associated bacteria (BVAB) are important for BV development[88].

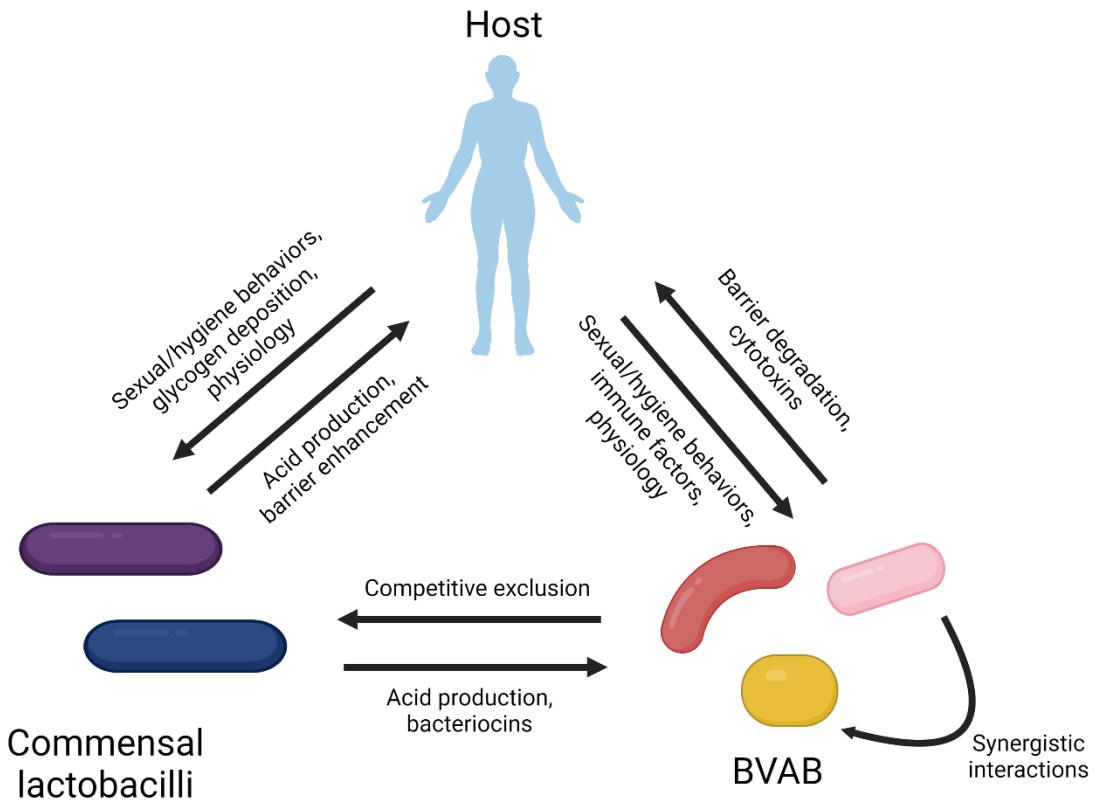
The polymicrobial nature of BV may be a result of complementary functions performed by different species of BVAB. Isolates of *Gardnerella* and *Prevotella* are known to produce sialidases, enzymes that cleave sialic acid residues off glycoproteins, degrading mucus and inactivating immune proteins[89-92]. Bacteria in these genera also make proteases which degrade host proteins, releasing amino acids that can be catabolized for energy and nitrogen[93,94]. Many strains of *Gardnerella* also encode a cholesterol-dependent cytolysin, vaginolysin, that is cytotoxic to host cells[95,96]. Curiously, *Lactobacillus iners* encodes a similar toxin, called inerolysin[52,97]. Some strains of *Gardnerella* are also prolific biofilm-formers, which promotes adherence to the vaginal epithelium and can increase bacterial tolerance to antibiotics and lactic acid[98]. Finally, some BVAB synthesize amino acids into foul-smelling compounds such as putrescine, cadaverine, and trimethylamine[99-101], although it is unclear which specific organisms catalyze these reactions.

Diverse synergistic interactions are also known to occur between BVAB. The biofilms formed by *Gardnerella* can house other organisms, and *Fannyhessea vaginae* have enhanced growth in biofilm with *Gardnerella*[102,103]. *Prevotella bivia* releases ammonia when catabolizing peptides, which *Gardnerella* can use as a source of nitrogen for building biomass[104,105]. *P. bivia*

can also promote the growth of *Peptostreptococcus anaerobius* by releasing excess amino acids from peptides which the latter organism needs to grow[106].

Despite decades of research into BV and elucidation of some bacterial functions in the vaginal microbiota, it remains unclear why eubiotic communities shift to BV, and why BV frequently recurs after treatment. This is likely a factor of the complex nature of BV, with interactions between the host, commensal lactobacilli, and diverse communities of BVAB all influencing vaginal microbiota composition (Fig. 1). Some groups have proposed that *Gardnerella* plays a vital role in the etiology of BV[107], but even if this were the case, it remains unclear what other organisms are necessary for BV to develop. Others have suggested that biofilms of BVAB which persist during antibiotic treatment[108], or organisms that have colonized the sexual organs of patients' partners[109], could serve as a reservoir for virulent organisms. However, supplemental therapies to target these sources of BVAB have had little success[83,110]. A better understanding of the host and microbial factors which influence BV would be helpful for developing better therapies to treat BV and prevent recurrences.

Figure 1. Host and bacterial interactions that shape the vaginal microbiota.



BV is a complex condition influenced by numerous host and bacterial factors. Interactions between the host, commensal lactobacilli, and the various BVAB present in the vagina all shape the composition of the bacterial community. Made with BioRender.com.

Presentation of Main Questions

Most microbiological research on BV has focused on a small number of organisms, namely *Lactobacillus, Gardnerella,* and *Prevotella*. These organisms are commonly some of the most abundant in the vaginal microbiota and investigations into their biology have uncovered functions and synergistic interactions that likely play a large role in shaping the vaginal community. Numerous functions and interactions likely remain undiscovered, however, especially for other BVAB which have received less attention.

This work describes an investigation using untargeted metaproteomic analysis to probe bacterial and host functions in the vagina, and subsequent laboratory experiments to elucidate

how these functions may relate to BV. Chapter 2 describes the process of optimizing protein databases in order to maximize protein data generated by metaproteomic analysis of cervicovaginal lavage (CVL) samples from study participants with eubiotic and BV vaginal microbiotas. Chapter 3 examines the results of that optimized metaproteomic analysis to identify host and bacterial functions which may contribute to the stability of eubiotic and BV bacterial communities, as well as shifts between them. Chapter 4 presents an investigation into glycoside hydrolase enzymes made by vaginal bacteria, which arose from the identification of pullulanase enzymes from *L. crispatus* and *Gardnerella* in the CVL metaproteomic data. Finally, Chapter 5 details the creation of a novel genetic system in *Gardnerella*, representing the first step toward genetic manipulation of these organisms. In total, this work describes new methods for studying BV and uncovers novel interactions between and among BVAB, commensal lactobacilli, and their host.

Chapter 2: Optimization of Metaproteomics Databases

Background

Metaproteomics is a powerful technique for uncovering biological functions in complex samples, with the additional benefit that proteins can be tied to specific taxa, revealing the functions of different organisms in a community[111]. Despite these advantages, the volume of data generated by metaproteomic analysis lags behind other omics techniques such as metagenomics and metatranscriptomics[112-115]. This is in part a result of the methods by which metaproteomic analysis is performed. Analysis begins by purifying proteins out of a sample and digesting them with a protease which has a well-defined cleavage sequence (i.e. trypsin)[116]. The resultant peptides are then analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS), which generates peptide mass spectra. Counterintuitively, even though throughput for mass spectrometry is much lower than for nucleotide sequencing-based technologies, data analysis is currently considered the primary bottleneck in mass spectrometry-based metaproteomics[117].

To identify sample proteins, peptide mass spectra generated by LC-MS/MS are fed into a search algorithm which compares them against theoretical spectra generated by performing an *in silico* digest of protein sequences from a user-supplied database[118-120]. Using statistical methods, the search algorithm identifies the theoretical spectrum which best matches each individual sample spectrum and assigns the resultant peptide-to-spectrum matches (PSMs) a significance value. To account for the multiple comparisons problem that arises from comparing each sample spectrum to the numerous peptides from a database, the search algorithm also compares the sample spectra to peptides from a "decoy" database of random protein sequences[121]. Matches to theoretical spectra from the decoy database are known false positives, so the proportion of decoy matches can be used to calculate the false discovery rate (FDR) for a database search. Search FDR and the significance value of an individual PSM are then used to calculate a *q* value[122]. The *q* value indicates the likelihood a sample spectrum was matched to a

database peptide by random chance, and therefore whether it is statistically significant. Finally, peptides from significant PSMs are linked back to their antecedent database protein(s) to determine what proteins are present in the sample. The large search space metaproteomics algorithms must cover makes data analysis computationally intensive and slow relative to data acquisition[123]. Therefore, improved methods for identification of sample spectra are needed to improve metaproteomics analysis.

Optimizing protein sequence databases is a promising approach for improving metaproteomics analysis. Database construction poses a challenging balancing problem. A sample spectrum can only be identified if there is an exact match present in the database; a single amino acid difference may result in a low *q* value and the spectrum being excluded from further analysis, or worse, misidentification of the spectrum[124]. This fact incentivizes researchers to include as many proteins as possible in the sequence database, however, larger databases have a higher FDR, so including more proteins can drive some PSMs below the threshold for significance and reduce the amount of data generated by analysis[125]. Although protein database construction is a complex and important part of metaproteomics data analysis, relatively few studies have investigated how to build an optimal database, and those that have primarily examined simple artificial communities of bacteria or samples from the gut microbiome[126-129].

This chapter describes a paper authored by Elliot Lee investigating the effect of using different strategies to build protein databases on the results of metaproteomic data analysis for cervicovaginal lavage (CVL) samples[130] to test the hypothesis that sample-matched databases populated with proteins translated from metagenomic sequencing of the samples will generate the most PSMs. This work was completed in collaboration with Sujatha Srinivasan, Samuel Purvine, Tina Fiedler, Owen Leiser, Sean Proll, Samuel Minot, Brooke Deatherage Kaiser, and David Fredricks. My contributions to this paper included performing experiments described in Figures 2 – 14 and Tables 1 – 6, as well as writing and preparing the final manuscript.

Results

Participant Characteristics

A cross-sectional case-control set of 20 vaginal samples from women with bacterial vaginosis (BV) and 10 samples from women without BV were selected at random from the parent study of 220 women from Seattle with and without BV for analysis by 16S rRNA gene sequencing, metagenomic sequencing, and metaproteomics by LC-MS/MS[25]. One of the BV-negative samples was excluded from further analysis due to the presence of a polymer or detergent that compromised the sample's analysis by mass spectrometry. 41% (12/29) of study participants identified themselves as Black and 48% (14/29) identified as White (Table 1).

Table 1. Study participant characteristics.

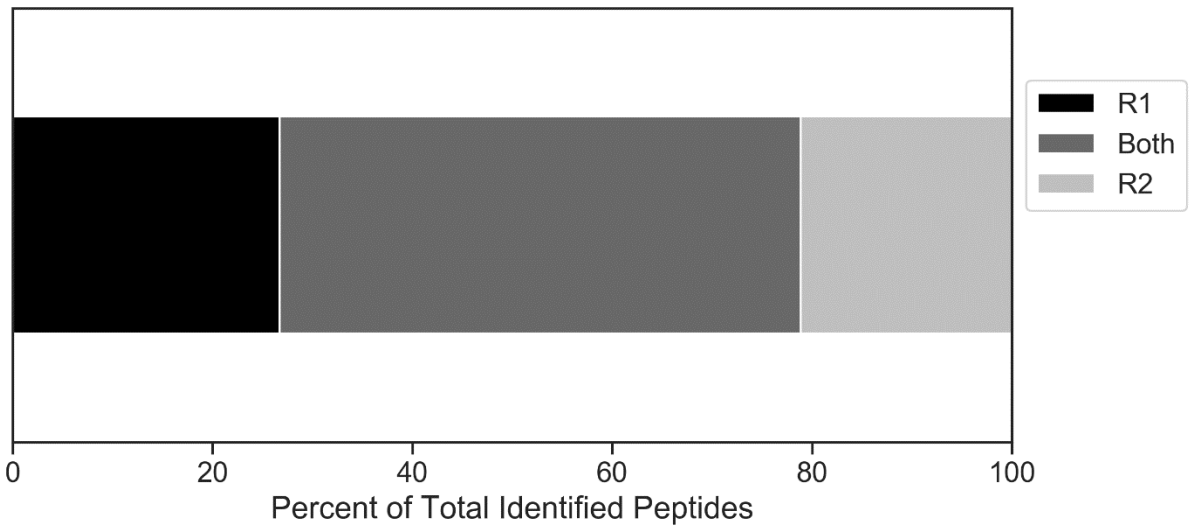
	All Participants	[a]BV-	BV+
[b]*N*	[c]30	9	20
Age range Mean	19-56 29.3	23-56 32.1	19-42 28.1
[d]Race/ethnicity Black White Other N/A	 12 (41.4%) 14 (48.3%) 2 (6.90%) 1 (3.40%)	 3 (33.3%) 6 (66.7%) 0 (0.00%) 0 (0.00%)	 9 (45.0%) 8 (40.0%) 2 (10.0%) 1 (5.00%)
Nugent Score Range	 0-10	 0-3	 7-10
Thin Homogeneous Vaginal Discharge	21 (72.4%)	3 (33.3%)	18 (90.0%)
Clue Cells	19 (65.5%)	0 (0.00%)	19 (95.0%)
pH Range	 4.0-5.8	 4.0-5.0	 5.0-5.8
Positive Whiff Test	20 (69.0%)	0 (0.00%)	20 (100%)

[a]Bacterial vaginosis was diagnosed using Amsel clinical criteria.
[b]*N* indicates the number of participants.
[c]The sample from one of the 10 BV− participants was contaminated with an unknown polymer which compromised metaproteomic analysis. Thus, this sample was excluded from further analysis and only nine BV− samples were considered.
[d]Other races included Asian/Filipino (one participants) and Native Hawaiian/Pacific Islander (one participants). One participant chose not to specify their race.

Identified peptides differ by technical replicate

We analyzed each sample by LC-MS/MS in two separate injection replicates. In a preliminary analysis of this data, we found that 26.6% (7,001/26,285) of all unique peptides were identified only in the first run of the samples, 21.1% (5,559/26,285) were identified only in the second run, and 52.2% (13,725/26,285) were identified in both runs, reflecting the stochasticity of metaproteomic analysis[131] (Fig. 2). To maximize usable protein data, we combined both runs of each sample for all further analyses.

Figure 2. Percentage of peptides identified only in the first replicate of the samples, only in the second replicate, or in both replicates.



In a preliminary analysis, spectra from the first and second replicates of the samples were searched against 16S_Sample-Matched databases. The percent of unique peptides sequences associated with a PSM where $q < 0.01$ identified in only one of the replicates, or both, are shown in the figure.

Small, sample-specific databases outperform broad databases when analyzing peptides in cervicovaginal lavage samples

We sought to compare different strategies for database construction by analyzing metaproteomic data with six different types of databases: three built with publicly available sequences, two populated with proteins translated from metagenomic sequencing of the samples,

and one hybrid that combined a sample-matched public sequence database and translated metagenomic database. The specific contents of each database are described in Methods and Figure 3. Briefly, each database contained all SwissProt human protein sequences (release 2019_02) plus 16 contaminants commonly introduced during metaproteomic analysis[132]. We built a Global database with all bacterial and fungal proteins available from NCBI RefSeq (RefSeq release 97). To build databases specific to the vaginal microbiome, we characterized the bacterial communities present in the samples by both 16S rRNA gene sequencing and shotgun metagenomic sequencing. In a preliminary analysis, we identified peptides from taxa present in the samples at a relative abundance as low as 0.1% according to 16S rRNA gene sequencing. Therefore, we built separate 16S_Sample-Matched databases for each sample, populated with all NCBI RefSeq proteins available for the bacteria present in the sample at >0.1% relative abundance. We then built a single 16S_Pooled database that included all bacterial proteins included in the individual 16S_Sample-Matched databases. To ensure the 16S_Pooled database better represented the broad vaginal microbiota, it also included RefSeq proteins from the common vaginal fungi *Alternia alternata, Candida albicans, Nakaseomyces glabrata, Candida tropicalis, Pichia kudravzevii,* and *Saccharomyces cerevisiae*, along with the common eukaryotic parasite *Trichomonas vaginalis*. For the translated metagenomic databases, the Shotgun_Pooled database contained the translated proteins from all bacterial open reading frames (ORFs) identified across all samples in the dataset, while the Shotgun_Sample-Matched databases only contained the translated proteins from its corresponding sample. Finally, we built a separate Hybrid_Sample-Matched database for each sample that combined the proteins in the 16S_Sample-Matched and Shotgun_Sample-Matched databases. For each database, we performed a two-step target-decoy database search as described in Methods. We evaluated the performance of the databases using the metrics time and cost for computational processing. Because protein spectral count is used to make statistical comparisons in metaproteomic

analysis, we also compared the number of significant human and bacterial PSMs generated by searches of each database.

Figure 3. Contents of the public sequence and translated metagenomic sequencing databases.

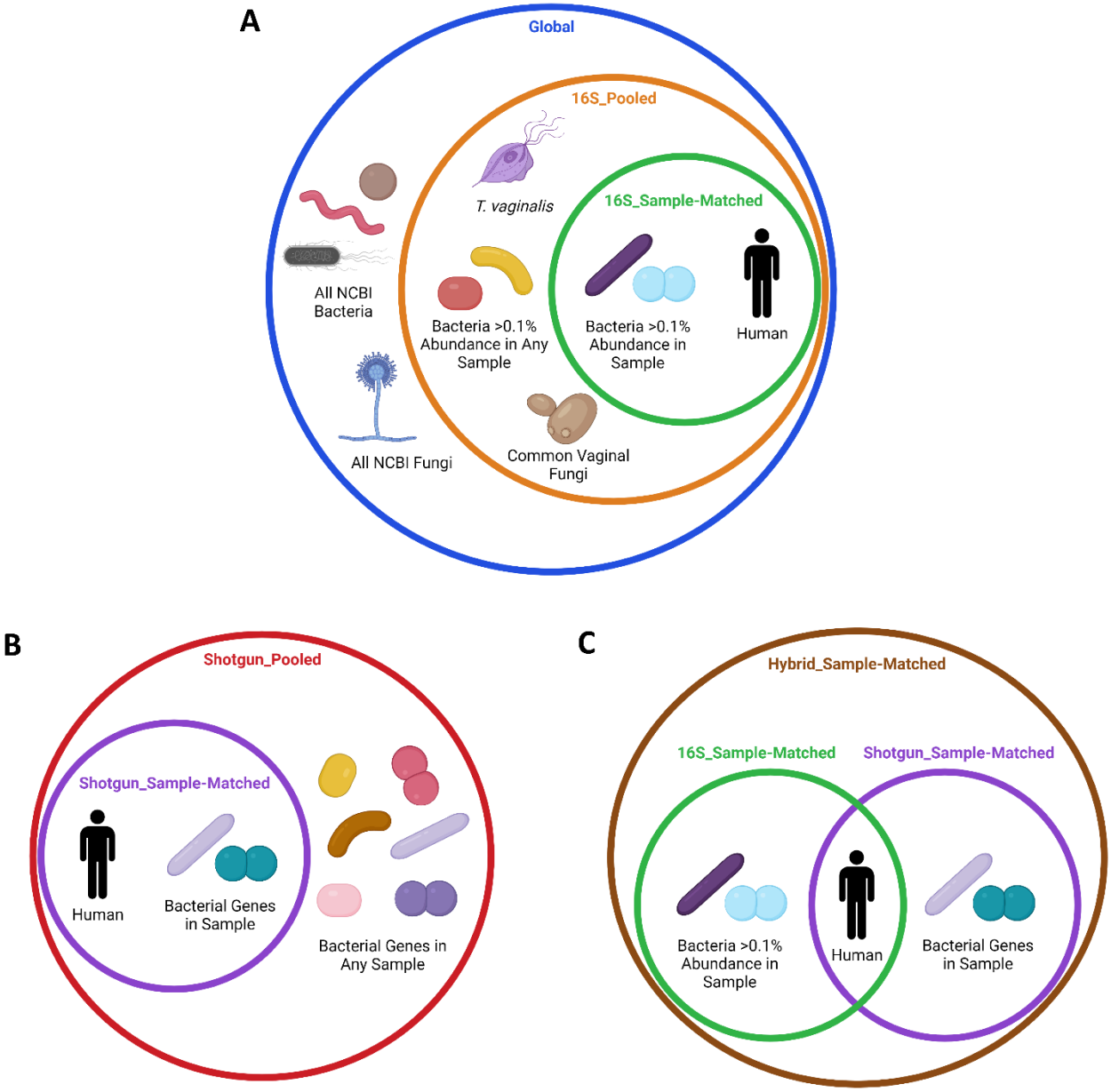
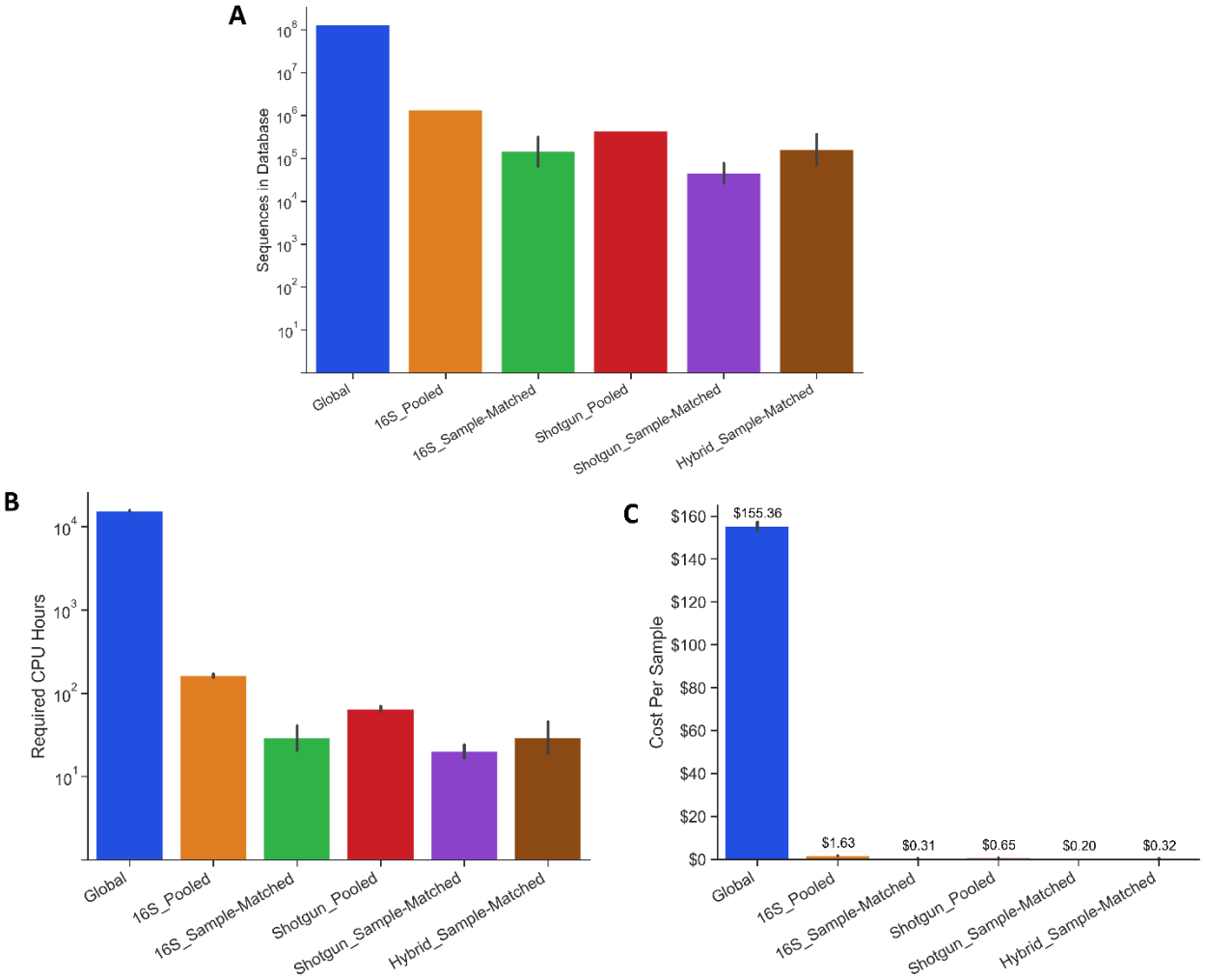


Illustration of proteins included in the tested databases. A) Proteins used to make the public sequence databases. B) Proteins used to make the translated shotgun metagenomic databases. C) Proteins used to make the hybrid public sequence/translated shotgun metagenomic databases. Created with Biorender.com.

The Global database was by far the largest with 131,886,982 total protein sequences (Fig. 4A). The 16S_Pooled database was two orders of magnitude smaller, containing 1,345,203 sequences. 16S_Sample-Matched databases ranged in size from 32,054 to 401,128 sequences, depending on the microbial diversity of their matched sample. The Shotgun_Pooled database contained 443,291 sequences, while the Shotgun_Sample-Matched databases ranged from 21,736 to 111,694 sequences. The Hybrid_Sample-Matched databases contained between 32,587 and 474,108 proteins. Because the Global database was so large and computationally expensive to search, we randomly selected a subset of six samples (two BV negative samples and four BV positive samples to maintain the proportion of the initial sample set) to search against it and compare to the other database types. The computational power required to perform a two-step search against each database type was proportional to their initial size, with the Global database requiring approximately 100-fold more central processing unit (CPU) hours than the 16S_Pooled database (Fig. 4B). Based on standard pricing for Amazon Web Services cloud computing at time of writing (2023) these CPU requirements equate to an average of $155.36 to perform a two-step search on one sample using the Global database, $1.63 with the 16S_Pooled database, $0.31 with a 16S_Sample-Matched database, $0.65 with the Shotgun_Pooled database, $0.20 with a Shotgun_Sample-Matched database, or $0.32 with a Hybrid_Sample-Matched database (Fig. 4C).

Figure 4. Comparison of database types: size, computational power, cost.

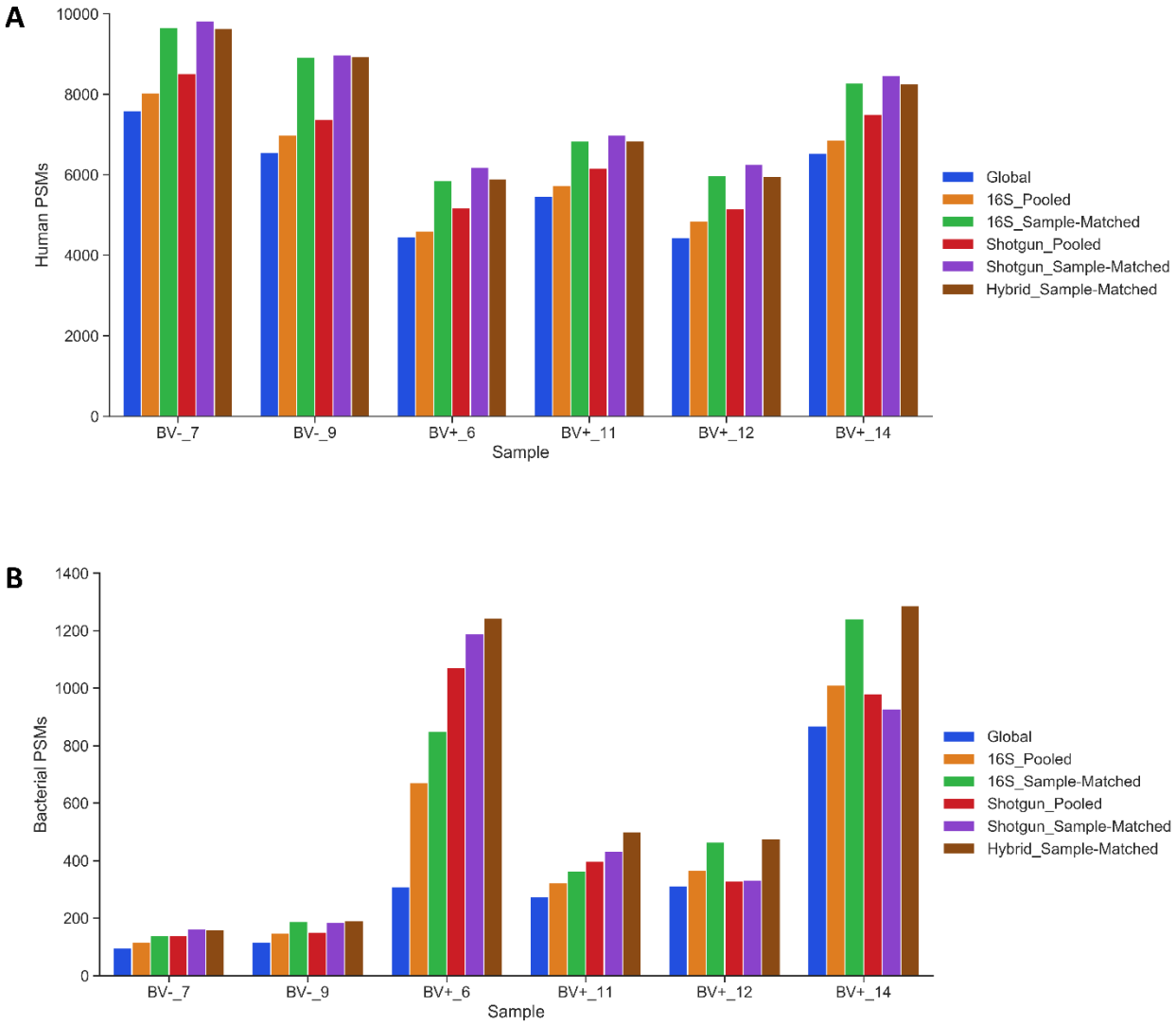


Comparison of database types: size, computational power, and cost. A) Comparison of the number of protein sequences in each database type. Lines show standard deviation in cases where a separate database was built for each sample. B) Total CPU hours required to perform a two-step search using each database type with the MS-GF+ search program. Lines show standard deviation to search the subset of six samples. C) Cost to perform a two-step search using each database type. Numbers represent the average cost to run the subset of six samples. Lines show standard deviation to search the subset of six samples. Graphs show mean ± standard deviation.

We analyzed the number of significant fungal PSMs generated by searches of Global and 16S_Pooled databases, as defined by a q value <0.01, since these were the only databases that included fungal proteins. Searches of the subset of six samples for both databases identified a very small number of significant fungal PSMs, 23 on average for Global and 1 for 16S_Pooled. Because the number of fungal proteins in the samples was so low, we focused the rest of our

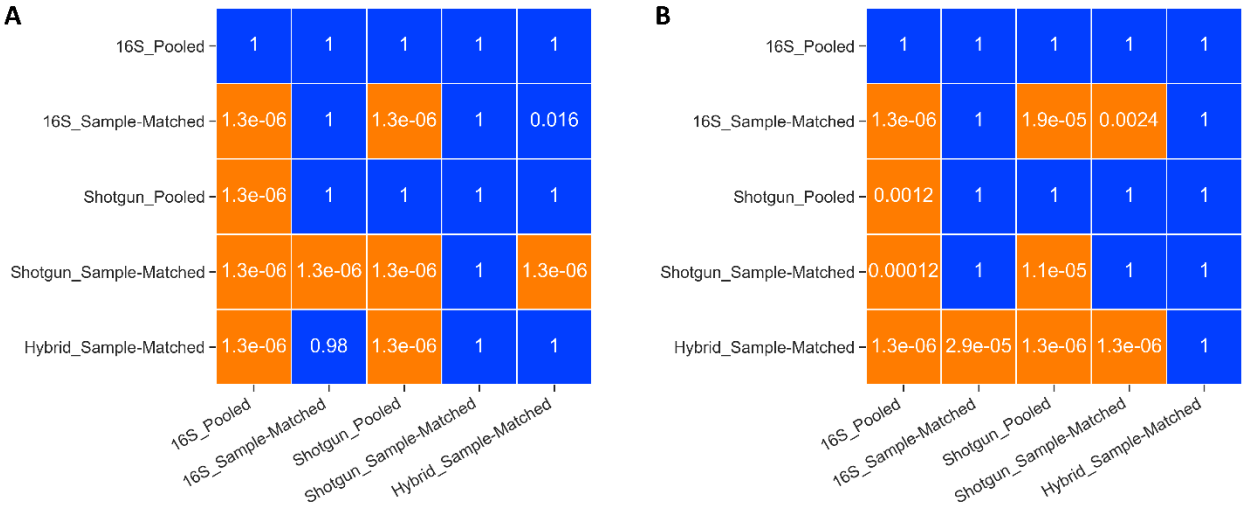
analysis on human and bacterial PSMs. Database size was a major factor in the number of significant human and bacterial PSMs generated, as the relatively small 16S_Sample-Matched databases generated the most significant human and bacterial PSMs of the three public sequence databases (Fig. 5A and 5B). This pattern held for all 29 vaginal samples. We performed a pairwise comparison of each database type (excluding the Global database) to determine which databases generated more significant human and bacterial PSMs (Fig. 6A and 6B) and calculated relative identification rates for each database to visualize these comparisons (Fig. 7A and 7B). Shotgun_Sample-Matched databases identified significantly more human and bacterial PSMs than the Shotgun_Pooled database, while the 16S_Sample-Matched databases similarly outperformed Global and 16S_Pooled. The relative performance of public sequence versus translated metagenomic databases varied by PSM type. The smaller Shotgun_Sample-Matched databases generated significantly more human PSMs than all other databases, but 16S_Sample-Matched databases significantly outperformed them in terms of bacterial PSMs. Hybrid_Sample-Matched database searches combined novel PSMs from both database types, identifying as many or more bacterial PSMs than the 16S_Sample-Matched databases in 24 of 29 samples (83%), and only slightly fewer human PSMs.

Figure 5. Comparison of significant human or bacterial PSMs generated by six database types.



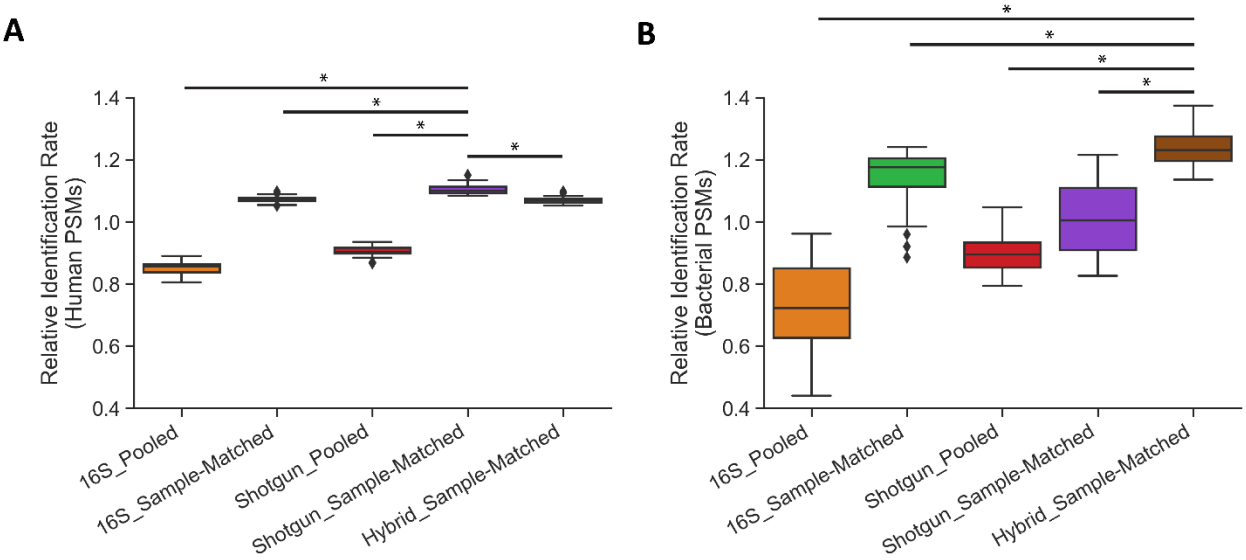
Number of significant A) human and B) bacterial PSMs generated by searching the subset of six samples using each database type.

Figure 6. Comparison of database performance across all samples.



All 29 CVL samples were searched against the five listed database types and one-sided Wilcoxon signed-rank tests were performed to determine whether the database listed on the row generated significantly more A) human or B) bacterial PSMs than the database listed under the column. The p-value for each test is shown in the cell. Comparisons that were significant ($P < 0.01$) are shaded orange while non-significant comparisons are shaded blue.

Figure 7. Relative identification rates between database types.



Relative identification rates between database types. Relative identification rate for each sample database search was calculated by taking the number of significant A) human or B) bacterial PSMs identified in a particular sample using the listed database, then dividing this number by the average significant PSMs of the same type identified in that sample across all five database searches. Bars above the data represent pair-wise comparisons of significant PSMs generated that were significantly different according to a Wilcoxon signed-rank tests ($P < 0.01$). These bars are only depicted for the database that performed best for the given PSM type.

We also compared the average percent of bacterial PSMs that only matched one or more proteins from a single genus or species to assess the taxonomic performance of different databases (Table 2). Public sequence databases outperformed the translated metagenomic databases when assigning PSMs at the genus level; however, all databases performed similarly at the species level, assigning between roughly 40% and 50% of bacterial PSMs to a single species.

Table 2. Percentage of significant PSMs per sample identified to specific taxa by searches of different protein databases.

Database	Genus	Species
16S_Pooled	86.0%	42.9%
16S_Sample-Matched	95.8%	50.1%
Shotgun_Pooled	48.9%	46.3%
Shotgun_Sample-Matched	50.0%	49.4%
Hybrid_Sample-Matched	89.1%	43.8%

Average percent of bacterial PSMs matched to proteins from only one genus or species by searches of the listed database type.

To determine whether generating additional significant PSMs translated into more biological insights, we compared the number of differentially abundant Gene Ontology (GO) terms associated with proteins between BV+ and BV− samples[133,134]. We used EggNOG-Mapper to collect GO terms for each protein[135,136]. Searches that generated a greater number of significant PSMs also identified more functions that were significantly differentially abundant by BV status (Table 3). Results from Shotgun_Sample-Matched searches, which identified the most human PSMs in all samples, led to the most significantly differentially abundant human functions. Similarly, results from the Hybrid_Sample-Matched database searches led to the most significant differentially abundant bacterial functions.

Table 3. Differences in significantly differentially abundant functional annotations between BV- and BV+ samples by database type.

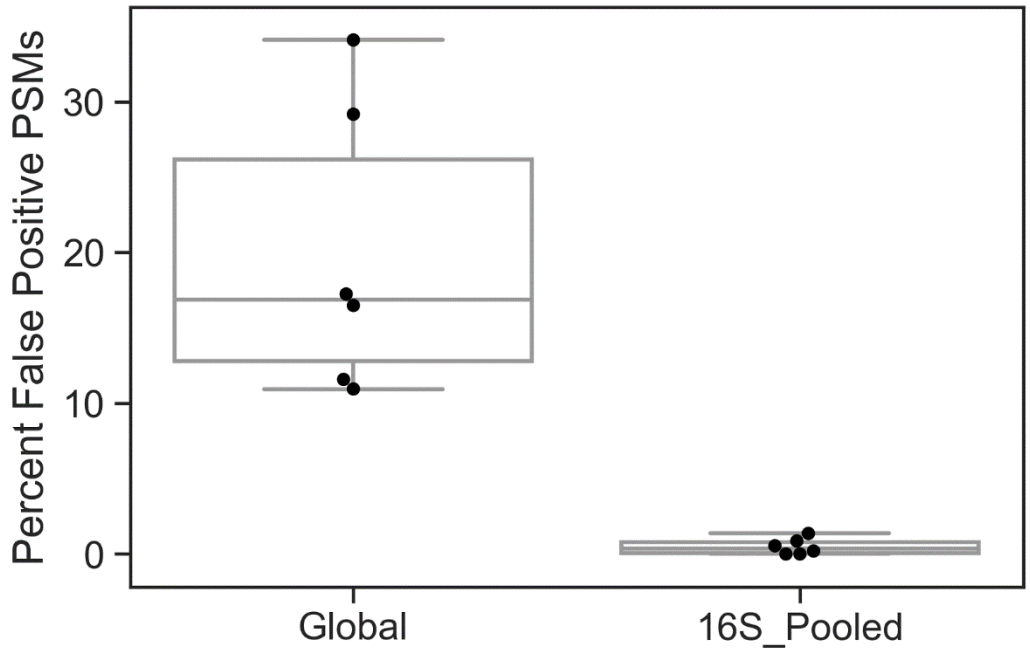
Database Type	Significantly Different Human Functions	Significantly Different Bacterial Functions
16S_Pooled	1,713	386
16S_Sample-Matched	1,881	460
Shotgun_Pooled	1,710	377
Shotgun_Sample-Matched	1,935	411
Hybrid_Sample-Matched	1,903	479

Number of Gene Ontology (GO) numbers that were significantly differentially abundant between BV- and BV+ samples based on the results of different database searches. Test for significance was carried out by Mann-Whitney U test for individual proteins/functional annotations using a significance of $P < 0.01$.

An extremely broad protein database is prone to produce false-positive identifications, while a focused database containing taxa expected to be in the community is more accurate

The Global and 16S_Pooled databases contained sequences from many microbial taxa that were not present in a given sample, so we evaluated whether this impacted the accuracy of their searches. We analyzed the significant bacterial PSMs identified by the Global and 16S_Pooled databases and determined what percent exclusively matched proteins from taxa at <0.1% abundance (based on 16S rRNA gene sequencing) in the sample, likely representing false positives. Of the bacterial PSMs detected by the 16S_Pooled database search, less than 0.5% on average exclusively matched genera below 0.1% relative abundance in the sample (Fig. 8). In contrast, an average of 20% of significant bacterial PSMs identified by searches of the global database only matched genera that were not likely present in the sample.

Figure 8. Presumptive false positive hits in Global and 16S_Pooled database searches.



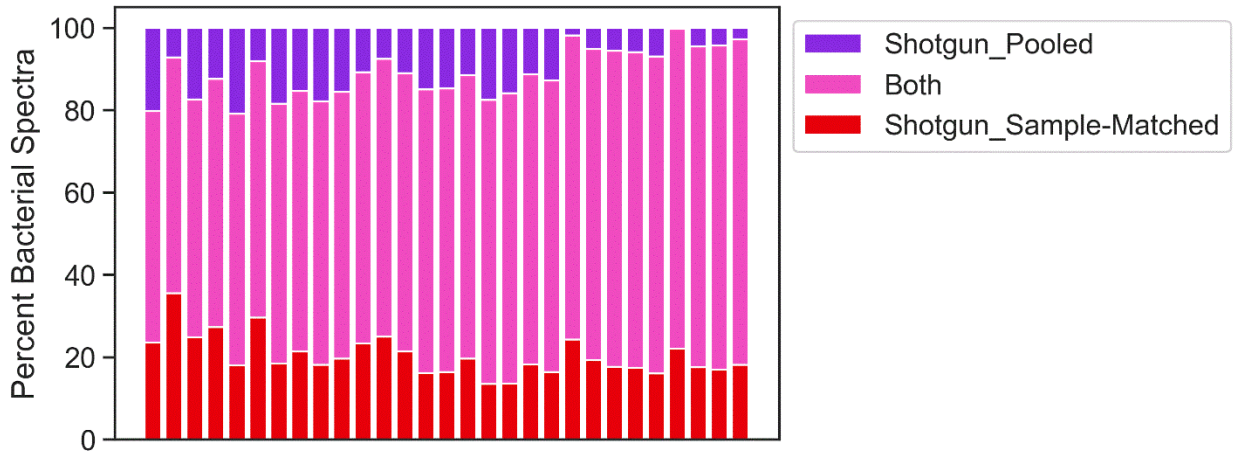
Each data point represents the percent of PSMs in a given sample which MS-GF+ matched to protein(s) belonging only to bacteria known to not be present in the sample, as determined by a relative abundance <0.1% by 16S rRNA sequencing. Results shown from the subset of six samples searched using the Global database.

Shotgun_Single databases outperform the Shotgun_Pooled database because their smaller size reduces the threshold of significance for spectrum identifications

Shotgun_Sample-Matched database searches identified more bacterial PSMs than the Shotgun_Pooled database in significantly more samples ($P < 0.0001$). This result could either indicate that the additional sequences in the Shotgun_Pooled database do not lead to novel spectrum identifications or that the smaller size of the Shotgun_Sample-Matched databases requires a lower statistical threshold for individual PSMs. To determine whether the pooled approach identified any new spectra, we investigated bacterial spectra given a significant identification by only one of the Shotgun databases, or both (Fig. 9). All samples had at least one bacterial spectrum identified only by the Shotgun_Pooled database search, showing that additional protein sequences resulted in identifications of novel spectra. We also evaluated

spectra successfully identified by Shotgun_Sample-Matched database searches but missed by those of the Shotgun_Pooled database. For 92% of these spectra, MS-GF+ assigned them the same peptide sequence regardless of the database searched, but with a *q* value greater than 0.01 for the Shotgun_Pooled searches. This indicates that the larger size of the Shotgun_Pooled database pushed these spectra below the threshold for significance and reduced the overall search performance.

Figure 9. Overlap of spectra identified in searches of Shotgun_Pooled and Shotgun_Sample-Matched databases.



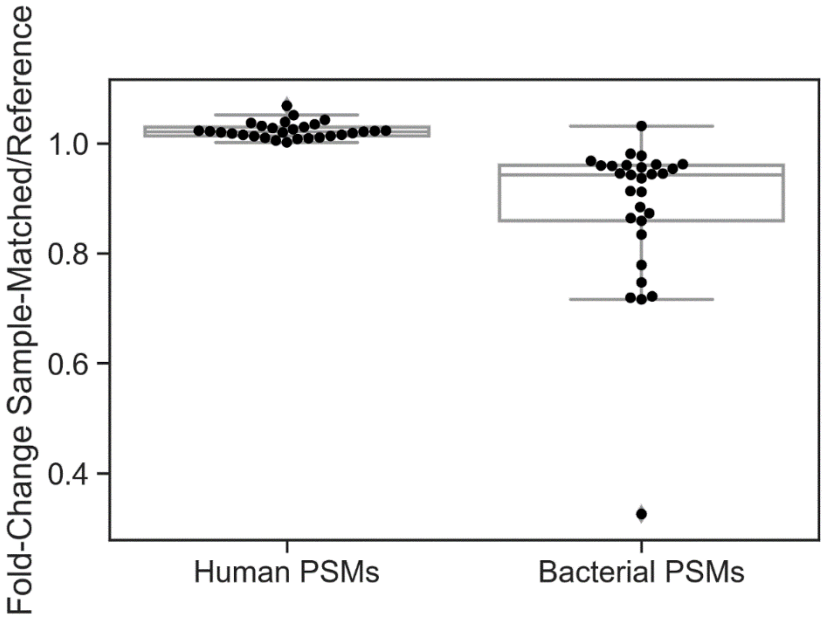
Percent of bacterial spectra in each sample identified only be searching the Shotgun_Sample-Matched database (red), Shotgun_Pooled database (purple) or identified in both database searches (pink).

Minimizing the number of bacterial proteins in a public sequence database slightly increases significant human PSMs but decreases identified bacterial spectra

Counterintuitively, one of the primary determinants of the number of significant human PSMs generated by a database search may be the number of bacterial proteins in the database, as many more bacterial proteins had to be included in the database to account for the heterogeneity of the bacterial proteome. Therefore, we tested whether minimizing the number of bacterial proteins in a database would increase the number of significant human PSMs it

generated. For each sample, we built a 16S_Reference database that included proteins from each bacterial species present in a sample at >0.1% relative abundance but only from the reference strain of that species as listed by NCBI[137]. Searches of these pared-down databases on average resulted in approximately 2% more significant human spectrum identifications than the 16S_Sample-Matched databases, but searches of the 16S_Sample-Matched databases resulted in approximately 19% more significant bacterial spectrum identifications, on average (Fig. 10). There was a large range in the fold-change difference in bacterial PSMs between the two databases, however, and in one sample the 16S_Reference search identified fewer than 40% of bacterial PSMs identified by the 16S_Sample-Matched search.

Figure 10. Fold-change in statistically significant human and bacterial PSMs identified by the 16S_Sample-Matched database searches compared to searches of 16S_Reference databases.

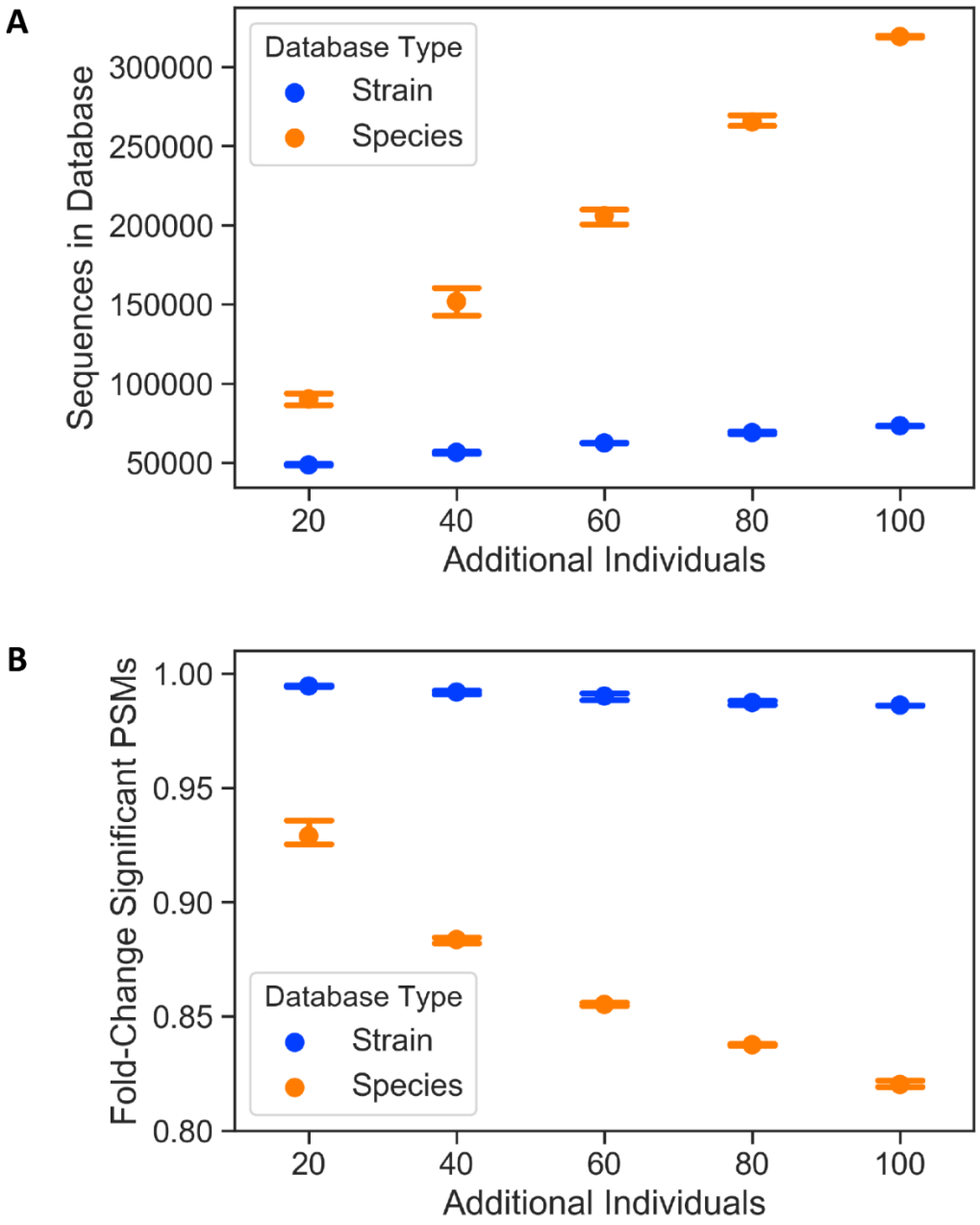


Each data point represents the ratio of significant PSMs identified by the 16S_Sample-Matched database compared to the 16S_Reference database in a sample.

Adding proteins to a database from additional strains of a species has a less negative effect than adding proteins from additional species due to greater protein heterogeneity

While our comparisons of the Global, 16S_Pooled, and 16S_Sample-Matched databases showed that database size is a major factor in search performance, our tests of the 16S_Reference databases indicated that including proteins from many strains of the species present in a sample also increased significant bacterial PSMs generated by a search. To test how adding additional strains or species to a database impacts search performance, we constructed a set of test databases by iteratively adding protein sequences from additional random bacterial species from the Human Microbiome Project[138] or strains of *Lactobacillus crispatus* to a baseline database. Due to overlapping protein sequences of the different *L. crispatus* strains, databases with additional species increased in size much faster compared to the databases with additional strains (Fig. 11A). We performed one-step searches using these databases on six of the proteomic samples: three low-diversity samples dominated by *L. crispatus* (>50% relative abundance) and three high-diversity samples not dominated by *L. crispatus* (<50% relative abundance). The number of significant PSMs decreased much faster for the databases with additional species than the databases with additional strains, reaching a ~17% reduction when protein sequences from 100 different species were added (Fig. 11B).

Figure 11. Effect of additional protein sequences from strains or species on database search performance.

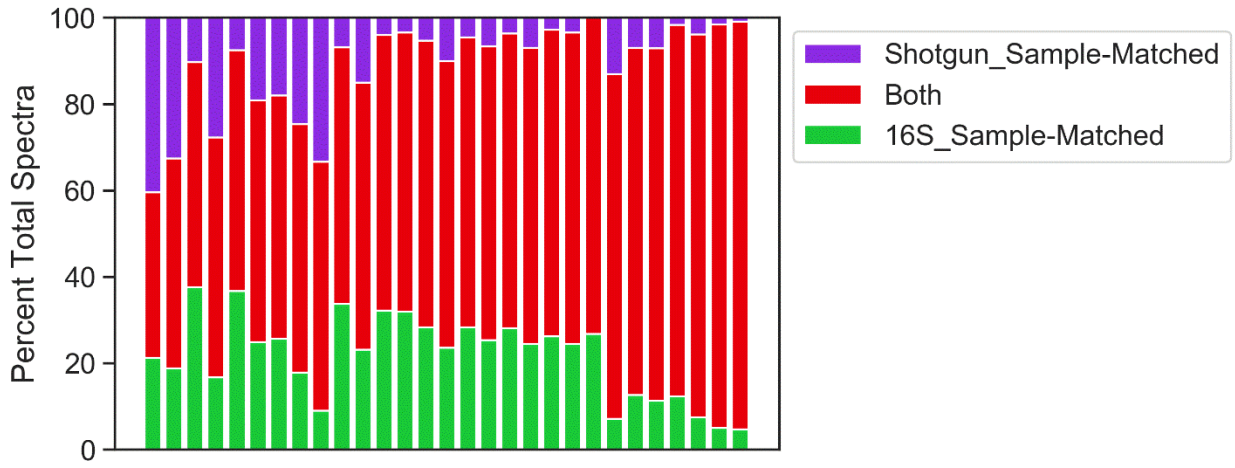


Effect of additional protein sequences from strains or species on database search performance. Dots represent mean values, and bars represent standard error. Databases were built by randomly adding a set number of individuals from a list of 103 strains of *L. crispatus* or 103 different bacterial species taken from the human microbiome sequencing project. A) Total number of sequences in the database. Data represents three randomly generated databases at each size. B) Fold-change in number of significant PSMs generated as compared to the baseline database. The value for each individual database was the mean fold-change for all six samples searched. Data represents three randomly generated databases at each size.

Public sequence databases identify many peptides from highly abundant taxa while translated shotgun sequencing databases excel at identifying peptides from uncultivated species

In most samples, searches of both the 16S_Sample-Matched and Shotgun_Sample-Matched databases identified many bacterial spectra that were missed by the other database type (Fig. 12).

Figure 12. Overlap of spectra identified in searches of 16S_Sample-Matched and Shotgun_Sample-Matched databases.



Percent of bacterial spectra in each sample identified only by searching its 16S_Sample-Matched database (green), Shotgun_Sample-Matched database (purple), or identified in both database searches (red).

We investigated the taxonomic identifiers associated with these spectra to determine whether the different database types were better at identifying peptides from certain groups of bacteria (Table 4). Many of the spectra that were exclusively identified by the 16S_Sample-Matched database were associated with taxa that were in high abundance in the samples, including *Gardnerella, Prevotella, Fannyhessea,* and *Lactobacillus*. The majority of these spectra were associated with *Gardnerella*, which is a well-sequenced, highly diverse genus that is frequently the dominant community member in BV. This indicates that shotgun metagenomic

sequencing can miss genes from major bacterial community members that are well represented in public sequence databases, likely as a result of low inherent sampling depth in metagenomic sequencing. Of those spectra exclusively identified by the Shotgun_Sample-Matched database searches, the largest number were associated with the taxonomic identifiers “Unknown” or “Terrabacteria.” Therefore, many of the bacterial genes identified by shotgun metagenomic sequencing could not be confidently resolved to a single group of bacteria. However, more than 300 spectra exclusively identified by the Shotgun_Sample-Matched database were assigned the taxa label “Clostridiales bacterium KA00274.” This bacterium is closely related to BV-associated bacterium 2 (BVAB2), a bacterium with no currently published genomes. These results demonstrate a situation where shotgun metagenomic sequencing identified bacterial peptides present in a sample but absent from public repositories.

Table 4. Taxa associated with spectra identified in 16S_Sample-Matched database searches and missed by Shotgun_Sample-Matched search, and *vice versa*.

16S_Sample-Matched		Shotgun_Sample-Matched	
Assigned Taxa	Total Spectra	Assigned Taxa	Total Spectra
Gardnerella	4929	*Gardnerella*	271
Prevotella timonensis	278	*Prevotella timonensis*	30
Prevotella buccalis	153	*Prevotella*	129
Fannyhessea vaginae	145	*Fannyhessea vaginae*	13
Megasphaera lornae	125	*Megasphaera lornae*	8
Prevotella disiens	125	*Prevotella disiens*	12
Prevotella bivia	117	*Prevotella bivia*	8
Lactobacillus crispatus	114	*Lactobacillus crispatus*	11
Lactobacillus iners	109	*Lactobacillus iners*	30
Prevotella amnii	56	*Prevotella amnii*	4
Porphyromonas uenonis	50	*Porphyromonas*	4
Lactobacillus jensenii	37	Bacilli	3
Peptoniphilus lacrimalis	34	*Peptoniphilus*	7
Candidatus *Lachnocurva vaginae*	31	*Lachnospiraceae*	3
Mobiluncus mulieris	30	*Mobiluncus mulieris*	15

Peptostreptococcus anaerobius	24	*Peptostreptococcus anaerobius*	1
Dialister micraerophilus	21	*Dialister micraerophilus*	2
Megasphaera hutchinsoni	18	*Megasphaera*	28
Prevotella bergensis	17	*Prevotella bergensis*	4
Mobiluncus curtisii	14	*Mobiluncus curtisii*	1
Finegoldia magna	13	Bacteria	15
Dialister sp	12	*Dialister* sp. type 2	3
Sneathia vaginalis	10	*Sneathia*	8
Peptoniphilus grossensis	10	Unknown	1667
Porphyromonas sp	8	*Clostridiales* bacterium KA00274	345
Parvimonas micra	7	Terrabacteria group	308
Prevotella intermedia	5	*Prevotellaceae*	8
Arcanobacterium haemolyticum	5	*Arcanobacterium* sp. S3PF19	4
Aerococcus christensenii	5	*Aerococcus christensenii*	6
Anaerococcus prevotii	5	*Anaerococcus lactolyticus*	2
Mobiluncus holmesii	5	*Mobiluncus*	11
Lactobacillus reuteri	4	*Lactobacillus*	50
Mageeibacillus indolicus	4	*Mageeibacillus indolicus*	7
Sneathia sanguinegens	4	*Tissierellia*	18
Sutterella sp	3	*Atopobium*	10
Fusobacterium nucleatum	3	*Leptotrichiaceae*	6
Ezakiella massiliensis	2	*Veillonellaceae* bacterium DNF00751	3
Prevotella colorans	2	*Actinomycetaceae*	2
Bacteroides thetaiotaomicron	2	*Bacteroidales*	42
Bacteroides faecis	2	*Bacteroidales* bacterium WCE2008	2
Eggerthella-like	2	*Coriobacteriales* bacterium DNF00809	4
Veillonella atypica	1	*Veillonellaceae*	46
Bifidobacterium dentium	1	*Bifidobacteriaceae*	31
Campylobacter ureolyticus	1	*Campylobacter*	1
Anaerococcus vaginalis	1	*Anaerococcus*	2
Megasphaera micronuciformis	1	*Peptostreptococcus*	2
Streptococcus agalactiae	1	*Streptococcaceae*	2
Sutterella wadsworthensis	1	*Bifidobacterium*	3
Candidatus TM7	1	*Veillonella*	2
Porphyromonas endodontalis	1	*Tissierellia* bacterium KA00581	2
		Gemella asaccharolytica	2
		Atopobiaceae	1
		Tissierellia incertae sedis	1
		Bacteroidia	1
		Coriobacteriales	1
		Lachnospiraceae bacterium	1

		Olsenella	1

The number of spectra across all samples that were matched to each taxonomic identifier and were identified by the 16S_Sample-Matched search and missed by searching the corresponding Shotgun_Sample-Matched search, or *vice versa.*

Results from optimized database searches identified a large number of unique proteins while remaining in line with past metaproteomic studies of the vaginal microbiome

Past metaproteomic studies of vaginal samples have primarily used the Uniprot/TrEMBL database as a reference for their bacterial proteins (Table 5). However, we found that a more tailored approach outperformed large, very broad databases. Our sample-matched databases identified a large number of unique proteins compared to other metaproteomic studies of vaginal samples, relative to the number of samples analyzed. Shotgun_Sample-Matched database searches identified the most unique human proteins at 1,182, while Hybrid_Sample-Matched identified the most bacterial proteins at 1,418. Although it may be valuable to identify more proteins, it is important that data remains accurate. To verify the accuracy of our optimized database searches, we identified human proteins which past metaproteomic studies had found were differentially abundant depending on BV status. We then tested whether there were similar differences in our results. Because the Hybrid_Sample-Matched databases struck the best balance between human and bacterial data, we analyzed the results of these database searches. The total number of human PSMs we identified in a sample also varied depending on BV status (data not shown), so we normalized the spectral count of each protein to the number of human PSMs identified in that sample, and compared the relative abundance of each protein between samples. Twenty-three human proteins had been identified as significantly differentially abundant in at least one of the studies analyzed[139-143]. In our data, we found significant differences that agreed with past studies for nine of these proteins (Table 6). Two proteins in our data with significantly different abundances did not align with past reports.

Table 5. Comparison of present techniques with past investigations of the vaginal metaproteome in terms of identified human and bacterial proteins, samples analyzed, and database type.

Publication	[a]*N*	Database Type	Human	Bacterial
Dasari, *et al.* 2007[144]	7	SwissProt Human	150	[b]N/A
Shaw, *et al.* 2007[145]	2	IPI Human	685	N/A
Zegels, *et al.* 2009[146]	6	SwissProt Human	339	N/A
Burgener, *et al.* 2011[147]	293	IPI Human	360	N/A
Birse, *et al.* 2015[148]	19	SwissProt Human + SwissProt Bacteria	384	N/A
Muytjens, *et al.* 2017[149]	10	SwissProt Human	1087	N/A
Ferreira, *et al.* 2018[142]	58	Uniprot Human	74	N/A
Starodubtseva, *et al.* 2019[150]	73	SwissProt	675	N/A
Kumar, *et al.* 2021[151]	60	Uniprot Human	1015	N/A
Klatt, *et al.* 2017[152]	688	SwissProt Human + Two-step Uniprot Bacteria	N/A	3334
Cruciani, *et al.* 2013[153]	80	SwissProt	118	13
Arnold, *et al.* 2014[154]	36	Database Not Specified	650	100
Borgdorff, *et al.* 2016[155]	50	SwissProt Human + NCBI Vaginal Lactobacilli + NCBI Vaginal Microbes	549	40
Zevin, *et al.* 2016[141]	10	SwissProt Human + Uniprot Bacteria	434	689
Bradley, *et al.* 2018[143]	16	SwissProt Human + Two-step Uniprot Bacteria	406	106
Farr Zuend, *et al.* 2020[156]	48	SwissProt Human + Two-step Uniprot Bacteria	550	376
Alisoltani, *et al.* 2020[157]	113	Uniprot Human + Uniprot Microbes	1236	1778
Nunn, *et al.* 2020[158]	4	Uniprot Human + Translated Sample Metagenomes	3334	1092
16S_Sample-Matched Databases (this study)	29	Two-step, Sample-matched SwissProt Human + NCBI Bacteria present by 16S Seq	1072	1257
Shotgun_Sample-Matched Databases (this study)	29	Two-step, Sample-matched SwissProt Human +	1182	942

		Translated Sample Metagenomes		
Hybrid_Sample-Matched Databases (this study)	29	Two-step, Sample-matched SwissProt Human + NCBI Bacteria present by 16S Seq + Sample Metagenomes	1068	1418

Comparison of the number of unique proteins identified in studies of the vaginal metaproteome, past investigations and current study.
[a]Number of samples analyzed.
[b]The study did not report the number of this type of proteins identified.

Table 6. Comparison of differentially abundant human proteins in BV identified by past studies with results of Hybrid_Sample-Matched database searches.

Protein	This Study		Past Studies	
	P-value	Higher/Lower in BV	Higher/Lower in BV	Reference
Elafin	>0.1	↓	↓	Stock et al.[139]
Muc5B	<0.05	↑	↑	Borgdorff et al.[140]
Muc5AC	<0.05	↑	↑	Borgdorff et al.[140]
Calprotectin	>0.1	↓	↑	Borgdorff et al.[140]
Complement factor 3	>0.1	↑	↑	Borgdorff et al.[140]
Migration inhibitory factor	>0.1	↑	↑	Borgdorff et al.[140]
Cystatin A	<0.01	↓	↓	Borgdorff et al.[140]
			↓	Ferreira et al.[142]
Lysozyme C	>0.1	↓	↓	Borgdorff et al.[140]
Serine protease inhibitor kazal type 5	<0.1	↓	↓	Borgdorff et al.[140]
Involucrin	<0.01	↓	↑	Zevin et al.[141]
			↓	Ferreira et al.[142]
Cornifin-A	<0.05	↓	↑	Zevin et al.[141]
Cathepsin G	>0.1	↑	↑	Ferreira et al.[142]
Neutrophil elastase	>0.1	↑	↑	Ferreira et al.[142]
Neutrophil defensin 1	>0.1	↑	↑	Ferreira et al.[142]
Leukocyte elastase inhibitor	<0.01	↓	↓	Ferreira et al.[142]
Histone H4	>0.1	↓	↓	Ferreira et al.[142]
SPR3	<0.01	↓	↓	Ferreira et al.[142]
Cornifin-B	<0.01	↓	↓	Ferreira et al.[142]
SPR2A	>0.1	↓	↓	Ferreira et al.[142]

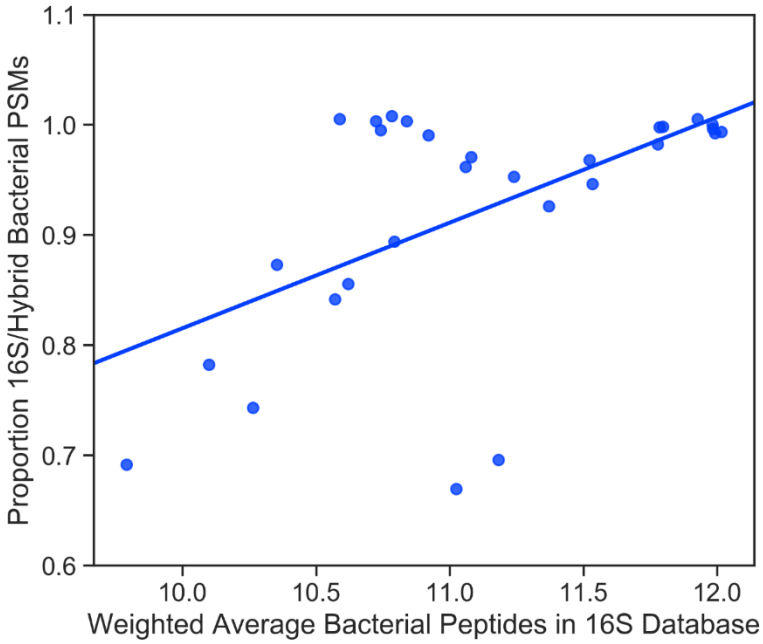
Repetin	>0.1	↑	↓	Bradley et al.[143]
Keratin 6A	>0.1	↑	↓	Bradley et al.[143]
Keratin 16	<0.01	↑	↓	Bradley et al.[143]
Suprabasin	<0.01	↓	↓	Bradley et al.[143]

Human proteins were found to be differentially abundant in BV in past metaproteomic studies compared with results from this investigation. Mann-Whitney U tests were performed on the results of Hybrid_Sample-Matched database searches and significance levels are shown. Each cell in the table specifies whether the protein had a higher (orange up-arrow) or lower (blue down-arrow) abundance in BV+ samples compared to BV− samples.

Shotgun metagenomic sequencing may provide less benefit to metaproteomic analysis of samples with a high abundance of well-sequenced species

Although Hybrid_Sample-Matched database searches identified the most significant bacterial PSMs, in 17 of the 29 samples, the 16S_Sample-Matched databases identified only 5% fewer bacterial PSMs. Performing additional shotgun metagenomic sequencing adds cost and complexity to analysis, so for future investigations, it would be useful to predict whether creating Hybrid_Sample-Matched databases would provide substantially more metaproteomic data. We hypothesized that having more protein sequence data for the bacteria in a sample would correlate with better relative performance of 16S_Sample-Matched databases. To test this hypothesis, we summed the number of tryptic peptide sequences available for the bacterial species present in each sample, weighted to their relative abundance. We then compared this number to the ratio of significant bacterial PSMs identified by the 16S_Sample-Matched database compared to its equivalent Hybrid_Sample-Matched database (Fig. 13). A Spearman's rank-order correlation test found a statistically significant correlation between an increasing number of tryptic peptide sequences available to search and the relative performance of the public sequence-only databases, showing that the benefits of metagenomic sequencing decrease as the amount of sequence information available for organisms increase.

Figure 13. Correlation between increasing public protein sequence data available for species in a sample and performance of 16S_Sample-Matched databases compared to Hybrid_Sample-Matched databases.



Correlation between increasing public protein sequence data available for species in a sample and performance of 16S_Sample-Matched databases compared to Hybrid_Sample-Matched databases. An *in silico* tryptic digest was performed on all publicly available protein sequences for bacteria present in the samples to determine how much protein sequence data was available for searching. The amount of sequence data in each 16S_Sample-Matched database was then calculated by log-transforming the number of tryptic peptides available for each species in the sample, weighted to its relative abundance, and summed across all species. The correlation between this weighted average number of tryptic bacterial peptides present in a 16S_Sample-Matched database is shown against the ratio of significant bacterial PSMs for that sample identified by 16S_Sample-Matched over Hybrid_Sample-Matched database searches. Line of best fit as calculated by linear regression is shown. Spearman's rank-order correlation: $\rho(46) = 0.42$, $P < 0.05$.

We also investigated what organisms had the largest increase in the total number of significant PSMs identified across all samples by utilizing a Hybrid_Sample-Matched database. Some of the largest were Candidatus *Lachnocurva* (29 additional PSMs), *Megasphaera hutchinsoni* (57 additional PSMs), and BVAB2 (378 additional PSMs). All three of these species had very little protein sequence data available at the time of analysis (only one published genome for both Candidatus *Lachnocurva* and *M. hutchinsoni*, and no available genomes for BVAB2), and high relative abundance in at least one sample (72.3%, 10.7%, and 10.4%, respectively). Conversely, *L. crispatus* and *L. iners* both have a large number of published genomes and had

>80% relative abundance in multiple samples, but when using searches of the Hybrid_Sample-Matched databases, we identified 2 fewer to PSMs for *L. crispatus* and 13 fewer total PSMs for *L. iners*.

Discussion

Modern mass spectrometry techniques have improved to the point where they can generate peptide mass spectra much faster than we can analyze and identify them[159]. Therefore, mass spectrometry data analysis techniques require refinement. Although metaproteomic methods have great potential to illuminate microbial physiology and host-microbe interactions in microbial communities, few studies have systematically examined how best to build a protein sequence database in order to speed up analysis and generate the maximum amount of useable data. A database should theoretically include as many proteins as possible to best capture the diversity of a sample, but in practice, increasing statistical stringency drives PSMs below the threshold for significance, reducing the number of analyzable PSMs and increasing analysis time. Many approaches have been proposed to balance these forces including employing a two-step search strategy, *de novo* peptide sequence identification, translating sequenced mRNA from study samples, populating databases with short "metapeptides," and combining public sequences with translated genes from metagenomic sequencing[160-165].

In this study, we tested multiple database types to evaluate the relative performance of different database construction strategies on metaproteomics results from vaginal samples. We found that the number of bacterial protein sequences included in a database has a large effect on the number of human PSMs generated by a search. Larger numbers of bacterial proteins in a database resulted in higher statistical thresholds, driving many human PSMs below the cutoff for significance. This will be a challenge for analyzing samples from many body sites, as the heterogeneity present in the bacterial proteome will often be greater than that of the human proteome. We also found that a database tailored to the vaginal microbial community

(16S_Pooled) generated a substantial number of PSMs with few obvious false positive identifications while a maximally broad database (Global) generated fewer PSMs with lower accuracy, and at a much higher cost per sample searched. This result agrees with prior investigations which show large protein databases underperform more focused databases[120,125-128,166]. We also built 16S_Reference databases to test how limiting the number of bacterial proteins in a database, even from the species known to be present in the sample, affects search results. Surprisingly, these minimal databases still generated approximately 88% as many significant bacterial PSMs as the larger 16S_Sample-Matched databases, though with a much lower percentage in some samples. The 16S_Reference database also generated on average 2% more human PSMs, likely because of the reduced size relative to the 16S_Sample-Matched databases.

Proteins identified as significantly differentially abundant in searches of the Hybrid_Sample-Matched databases largely agreed with past metaproteomic studies of BV. The exceptions were Cornifin A, which our data indicated had a significantly lower abundance in BV, and Keratin 16, which had a significantly higher abundance in BV. Cornifin A is a component of the cornified envelope of keratinocytes, and the trend in our data matched other cornified envelope proteins that were significantly less abundant in BV. Compared to other metaproteomic studies of the vagina, searches of our optimized databases also identified a relatively large number of unique human and bacterial proteins, showing the added value of optimized databases for metaproteomics. An outlier in this comparison was the investigation by Nunn et al., which identified a large number of proteins (3,334 human and 1,092 bacterial) from only four samples[158]. Their study analyzed vaginal mucus collected by Softcup, whereas most published metaproteomic studies of the vagina, including ours, used CVL. Vaginal mucus collected by Softcup may therefore be a richer sample type for future metaproteomic studies.

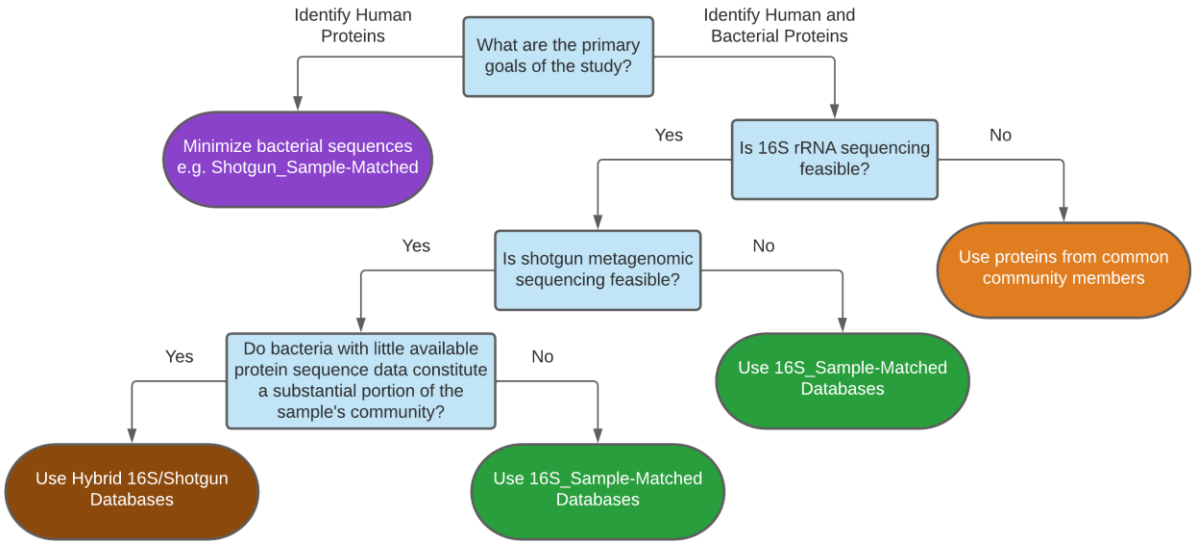
Past studies have generally found databases built using genes from metagenomic sequencing outperform public sequence databases, likely because they better represent the proteomic diversity of the sample[125,127,128,166]. In contrast, our 16S_Sample-Matched databases generated more bacterial PSMs than the Shotgun_Pooled database in 26 of the 29 samples, and the Shotgun_Sample-Matched databases in all but six samples. Some unique features of the vaginal niche may be responsible for these results. Past studies have primarily focused on the gut microbiome where bacterial diversity is often an order of magnitude higher than the vagina[167-169]. There are also many more published genomes available for the species of bacteria common in the vagina compared to those in the gut. Therefore, the relatively large number of publicly available genomes for species of vaginal bacteria may represent deeper sequencing than shotgun metagenomic sequencing of the actual samples. Additionally, because human proteins make up a greater proportion of total proteins in vaginal samples compared to the more bacteria-dense gut, statistical constraints from large numbers of bacterial proteins in a database will have a larger impact on the total number of PSMs generated. These features of the vaginal microbiome manifest in the results of our Hybrid_Sample-Matched databases. These databases combined the deep sequencing represented in public repositories with the heterogeneity identified by shotgun metagenomic sequencing of the samples while remaining small enough to have a minor impact on the number of human PSMs generated.

Our study provides additional evidence that a combined public sequence/metagenomic approach to database construction leads to additional significant bacterial PSMs[129,165], but shotgun metagenomic sequencing adds additional cost and complexity to a metaproteomic study, in some cases without generating more PSMs. We found a correlation between the number of tryptic bacterial peptides available in public repositories and the relative performance of a protein database that only included publicly available sequences. Additionally, we found that PSMs identified by leveraging shotgun metagenomic sequencing in Hybrid_Sample-Matched databases

often came from taxa with relative abundances ranging from ~10–70% in the samples but with little to no publicly available protein sequence data.

The results of our investigation lead to some general suggestions for constructing protein databases for metaproteomic analysis to fit the goals and limitations of a study (Fig. 14). First, if the primary goal of a study is to investigate the human proteome, it is important to note that including more bacterial proteins in the database will drive down the number of human PSMs generated by a search, even when using a two-step search strategy. Including proteins only from the reference genomes of the bacteria present in the sample will likely still provide a substantial amount of data, though at the expense that some bacterial protein diversity will be lost. Second, if bacterial proteins are also of interest in the study, performing 16S rRNA gene sequencing to profile the bacterial community of each sample and create sample-matched databases will generate more bacterial PSMs than a pooled approach, though PSMs from a community database are likely still useful. Third, databases which only include publicly available protein sequences may be adequate for analyzing samples composed of bacteria with many published genomes. However, if species with little available protein sequence data make up a substantial proportion of the community, an analysis will likely be improved by performing metagenomic sequencing.

Figure 14. Decision tree for selection of a metaproteomic database to achieve study goals with available resources.



Created using lucid.app.

While these guidelines are likely to help increase human and bacterial PSMs in future studies, with the additional benefit of reduced analysis times compared to using a more general database, we only tested them on samples from the vaginal microbiome. Database composition may have different effects on metaproteomic studies of other body sites or on purely microbial samples. Additionally, although many search programs are available for metaproteomic studies[119,170], we only used the MS-GF+ search program on our samples. MS-GF+ is primarily designed to search smaller protein databases, which may have negatively impacted the performance of the large Global database. However, MS-GF+ outperforms other frequently used search programs such as Mascot, SEQUEST, and MS-Align+[171,172], hence our selection of this search program for our study.

We did not investigate every variable involved in protein database construction. For example, while the relatively large number of bacterial proteins had an impact on the number of significant human PSMs generated by a database search, including all human proteins in each

database increases database size and may drive down the number of significant bacterial PSMs generated by a search. Future studies could investigate whether filtering human proteins out of a database based on body site or by referencing RNA-sequencing data could lead to more useful data on the microbial metaproteome[113]. Additionally, optimal construction of public sequence databases could be more thoroughly explored, especially regarding what species to include in the database. Our data suggested that databases tailored to the sample based on 16S rRNA gene sequencing data outperform other public sequence databases, and in many cases, databases assembled from shotgun metagenomic sequencing. We tested databases that included all species present in the sample at >0.1% abundance, however, it is possible that databases built with different cutoff thresholds could increase performance. We chose this cutoff because on average 2.7% of significant bacterial PSMs nonexclusively matched at least one taxa with relative abundance between 1% and 0.1% (data not shown). However, because the majority of identifiable spectra will likely come from species at higher abundance, it is possible that focusing on these taxa (e.g., >1% abundance) will increase overall database performance at the expense of functional information for minority community members.

This study provides guidance on database construction for future metaproteomic studies. Our findings support past investigations suggesting that small, focused protein databases have the dual benefit of increasing significant PSMs generated while reducing the amount of computational power required for metaproteomic data analysis. We also show that it is important to consider the specific niche under investigation when building a protein database as microbial diversity and availability of genomic information will impact the performance of different approaches.

Chapter 3: Host and Bacterial Functions in BV Elucidated by Metaproteomics and *in vitro* Investigations

Background

Despite decades of research into BV, the forces which drive shifts in the microbiota and stabilize eubiotic or dysbiotic communities remain poorly understood. BV is influenced by numerous factors. On the host side, behavioral and biological factors alter vaginal physiology, with large effects on the vaginal microbiota. However, investigations into host biological factors have returned largely contradictory results, and the contribution of host biology to BV remains enigmatic. Similarly, while a number of bacterial functions and interactions for major vaginal taxa have been elucidated, the role of most vaginal bacteria in BV is yet unknown.

The wide array of unknowns in BV provides an ideal application for untargeted metaproteomics. This technique not only identifies proteins to reveal the functions of a community, but it also allows those functions to be tied to individual organisms – both the host and different bacterial species. While a number of metaproteomic studies have been performed to investigate BV, these studies generally used large protein sequence databases that were not tailored to each individual sample. As we demonstrated in Chapter 2, such databases do not maximize analyzable data from metaproteomic analysis. Additionally, most of these studies relied on automated function mapping and pathway analysis to derive biological insights from their proteomic data. While these methods can be useful to distill large amounts of data and identify functional differences between sample types[173], they can ignore individual proteins which may not generate a statistical signal, but nonetheless have major biological implications[174]. A non-statistics-driven approach to metaproteomic data analysis can be valuable for identifying functions and interactions between community members, especially when many of those processes are unknown[175].

This chapter describes analysis of metaproteomic data from 29 CVL samples from women with and without BV, both by using statistical methods to identify proteins which are differentially abundant in eubiosis and BV, as well as drawing biological and taxonomic insights from individual protein identifications. We performed this study to test the hypothesis that metaproteomic analysis would reveal functional changes in BV, such as a degraded host barrier and a shift in microbial fermentation away from lactate production, in addition to revealing novel host and bacterial functions relevant to BV. This work was carried out by Elliot Lee in collaboration with Sujatha Srinivasan, Samuel Purvine, Tina Fiedler, Owen Leiser, Sean Proll, Samuel Minot, Danijel Djukovic, Daniel Raftery, Brooke Deatherage Kaiser, and David Fredricks. My contributions to this paper included performing experiments described in Figures 15 – 18 and Tables 7 – 10, as well as analyzing the data and preparing written conclusions.

Results

BV is associated with signs of increased and diversified bacterial metabolism

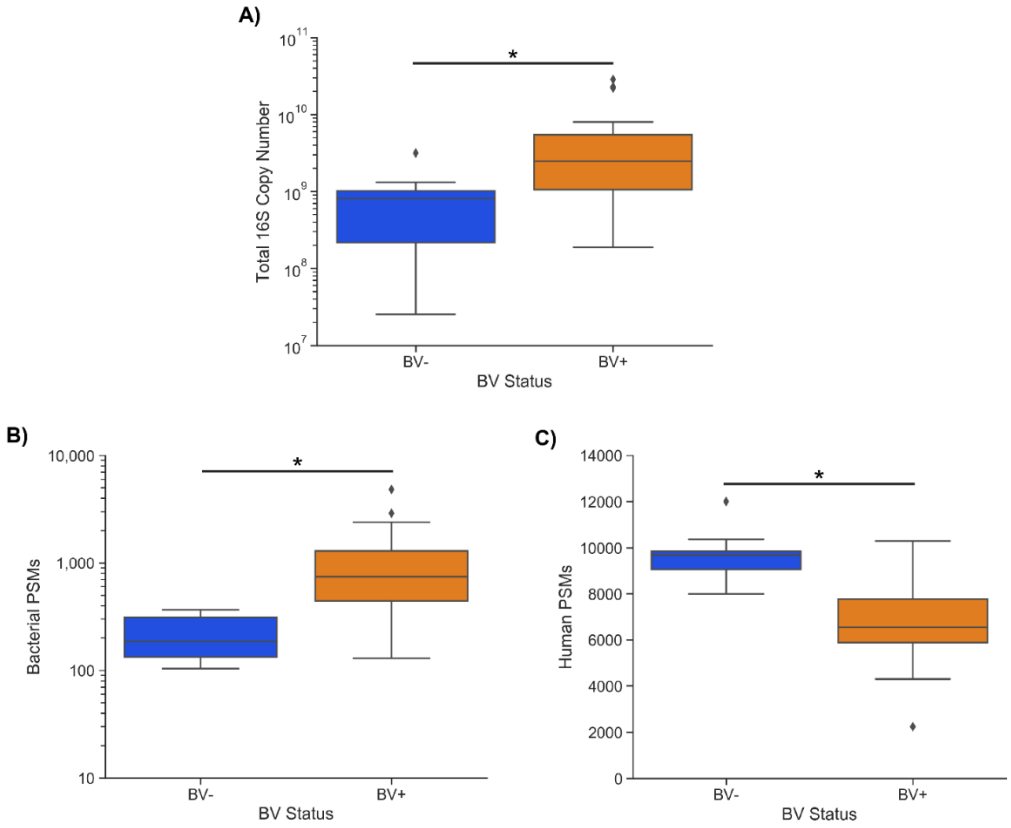
We performed 16S rRNA gene sequencing, broad-range and targeted 16S rRNA qPCR, metagenomic sequencing, and metaproteomics by LC-MS/MS on the same set of 29 CVL samples as described in Chapter 2[25]. Based on the results from our investigations on optimal protein sequence database construction, we analyzed the MS data from each sample using Hybrid_Sample-Matched protein databases and identified sample spectra using MS-GF+, then gathered spectral counts for each identified protein[130]. Because the number of human and bacterial proteins we identified varied widely between samples, we normalized spectral counts for human and bacterial proteins separately in each sample, using the relative abundance of a protein in the sample's bacterial or human proteome for downstream analysis.

To elucidate functional differences in the vaginal proteome during BV, we began our analysis by investigating what proteins were significantly differentially abundant based on a

sample's BV status. For human functions, we determined which individual proteins were significantly differentially abundant based on BV status by Mann-Whitney U Test ($P < 0.05$). For bacterial proteins, we first submitted identified proteins to the eggNOG-Mapper web server[135,136] (v1.0.3) to annotate their functions, then calculated the relative abundance of each functional group in a sample and determined which functions were significantly differentially abundant using the same statistical test. This analysis uncovered multiple observations which suggest the metabolism of vaginal bacteria is not only more diverse in BV, but the BV bacterial biota may be more metabolically active, generally. Bacterial proteins such as RNA polymerase, ATP synthase, and ribosomal proteins had significantly higher relative abundance in the BV+ samples (Table 7). In support of a more metabolically active bacterial community in BV, bacterial load as measured by total 16S copy number was significantly higher in BV+ samples (Fig. 15A). We also identified significantly more bacterial PSMs in the BV+ samples compared to BV- samples (Fig. 15B), which may indicate increased bacterial translation. We also found interesting proteomic signals related to specific nutrients. The carbohydrates stored in vaginal glycogen are a major source of carbon and energy for vaginal bacteria. Carbohydrates can be released from a glycogen molecule by the action of glycoside hydrolase enzymes such as amylase, then transported into bacterial cells by ABC transporters to be metabolized. Curiously, bacterial extracellular amylases and ABC transporters were both more abundant in BV+ samples, but all seven bacterial glycolysis proteins we identified were less abundant in BV (Table 7). Eight human glycolysis proteins were also significantly less abundant in the BV+ samples (Table 8). Additionally, the only glycogen-active human enzyme we identified in the samples was glycogen phosphorylase, which was significantly less abundant in BV+ samples. The lower abundance of glycolysis proteins may indicate other energy sources become a larger fraction of overall bacterial metabolism in BV, which is supported by the fact that bacterial extracellular proteases and enzymes involved in amino acid catabolism were significantly more abundant in the BV+ samples (Table 7). The action of these protein and

amino-acid catabolizing enzymes may be responsible for the fact that we identified significantly fewer human PSMs in BV+ samples (Fig. 15C).

Figure 15. Differences in bacterial load, identified PSMs by BV status.



A) 16S rRNA gene copy number for each sample as measured by broad-range qPCR. B) Number of significant bacterial PSMs identified in each sample. C) Number of significant human PSMs identified in each sample. Stars show comparisons that were significantly different by Mann-Whitney U Test ($P < 0.01$).

Table 7. Select significantly differentially abundant bacterial proteins by BV status.

Protein	Higher/Lower Abundance in BV	Function Category
Fructose-bisphosphate aldolase	↓	Glycolysis
Pyruvate kinase	↓	Glycolysis
Enolase	↓	Glycolysis
Phosphoglycerate kinase	↓	Glycolysis
6-phosphofructokinase	↓	Glycolysis
Glyceraldehyde-3-phosphate dehydrogenase	↓	Glycolysis
Phosphoglycerate mutase	↓	Glycolysis
Phosphoenolpyruvate carboxykinase	↑	Gluconeogenesis
Pyruvate, phosphate dikinase	↑	Gluconeogenesis
Lactic acid dehydrogenase	↓	Lactate Fermentation
Pyruvate formate lyase	↑	Formate Fermentation
Alcohol dehydrogenase	↑	Alcohol Fermentation
Acetate kinase	↑	Acetate Fermentation
Extracellular alpha-amylase	↑	Polysaccharide Metabolism
Glutamate dehydrogenase	↑	Nitrogen Metabolism
Glutamate transporter	↑	Nitrogen Metabolism
Glycine reductase	↑	Amino Acid Metabolism
Glutamine synthase	↑	Amino Acid Metabolism
Ornithine transcarbamoylase	↑	Amino Acid Metabolism
Extracellular Peptidase	↑	Protein Breakdown
RNA Polymerase	↑	General Metabolism
ATP synthase	↑	General Metabolism
ABC transporter	↑	General Metabolism
Ribosome	↑	General Metabolism
Peroxiredoxin	↑	Antioxidant

Significantly differentially abundant bacterial proteins as determined by Mann-Whitney U Test ($P < 0.05$). Proteins with an orange up arrow had significantly higher abundance in BV+ samples, proteins with a blue down arrow had significantly lower abundance. Homologous proteins from different species were identified with eggNOG-Mapper.

Table 8. Select significantly differentially abundant human proteins by BV status.

Protein	Higher/Lower Abundance in BV	Function Category
Ladinin-1	↓	Epithelial Structure
Catenin alpha-2	↓	Epithelial Structure
Involucrin	↓	Epithelial Structure
SPR3	↓	Epithelial Structure
Cornifelin	↓	Epithelial Structure
Protein S100-A10	↓	Epithelial Structure
Tight junction protein ZO-1	↓	Epithelial Structure
Cornulin	↓	Epithelial Structure
Periplakin	↓	Epithelial Structure
Zyxin	↓	Epithelial Structure
Filaggrin	↓	Epithelial Structure
Cadherin-1	↑	Epithelial Structure
Cornifin-A	↓	Epithelial Structure
S100-A11	↓	Epithelial Structure
Sciellin	↓	Epithelial Structure
Vinculin	↓	Epithelial Structure
Transglutaminase 3	↑	Epithelial Repair
Mucin-5B	↑	Mucus
Mucin-5AC	↑	Mucus
Pyruvate kinase PKLR	↓	Glycolysis
Glyceraldehyde-3-phosphate dehydrogenase	↓	Glycolysis
Phosphoglycerate kinase 2	↓	Glycolysis
Alpha-enolase	↓	Glycolysis
ATP-dependent 6-phosphofructokinase, platelet type	↓	Glycolysis
Glucose-6-phosphate isomerase	↓	Glycolysis
ATP-dependent 6-phosphofructokinase, liver type	↓	Glycolysis
Pyruvate kinase PKM	↓	Glycolysis
L-lactate dehydrogenase A chain	↓	Acidification
V-type proton ATPase catalytic subunit A	↑	Acidification
Glycogen phosphorylase, liver form	↓	Glycogen Metabolism

Superoxide dismutase [Cu-Zn]	↑	Antioxidant
Peroxiredoxin-4	↑	Antioxidant
Complement C3	↑	Immune
Leukocyte elastase inhibitor	↓	Protease Inhibitor
Calpastatin	↓	Protease Inhibitor
Cystatin-A	↓	Protease Inhibitor
Serpin B4	↓	Protease Inhibitor
Serpin B5	↓	Protease Inhibitor
Cystatin-B	↓	Protease Inhibitor
Serpin B3	↓	Protease Inhibitor
Serpin I2	↓	Protease inhibitor
Serpin B13	↓	Protease inhibitor
Hemopexin	↑	Heme Sequestration
Heme oxygenase 1	↓	Heme Breakdown

Significantly differentially abundant human proteins as determined by Mann-Whitney U Test ($P < 0.05$). Proteins with an orange up arrow had significantly higher abundance in BV+ samples, proteins with a blue down arrow had significantly lower abundance.

Identified proteins reveal functions of vaginal bacteria

We next examined the taxa associated with bacterial proteins we had identified in the samples to determine what functions different organisms may be performing. We found that most bacterial secreted amylases exclusively matched proteins from *Gardnerella* (Table 9). We only identified secreted *Gardnerella* amylases in samples from BV+ participants, but we also identified a small number of spectra that exclusively matched pullulanases from *Lactobacillus crispatus* in BV- samples, indicating that bacteria play a role in glycogen breakdown in both eubiotic and dysbiotic communities. We also identified a large number of spectra matching SusD/SusE[176] starch-binding domain proteins from multiple species of *Prevotella* and SusC polysaccharide import proteins from both *Prevotella* and *Porphyromonas uenonis*.

Under anaerobic conditions, cofactors consumed during carbohydrate metabolism *via* glycolysis must be regenerated by fermentation. We identified many bacterial fermentative

enzymes in our samples (Table 9) (Fig. 16). Bacteria associated with both health and dysbiosis are known to ferment lactic acid, but surprisingly, we identified more spectra matching pyruvate formate lyase proteins than lactate dehydrogenases across all our samples. Sample spectra matched homologs of pyruvate formate lyase from a wide range of BVAB including Candidatus *Lachnocurva*, *Fannyhessea, Gardnerella, Megasphaera, Peptoniphilus, Prevotella*, and *Veillonella*. Previous metabolomics studies have found elevated levels of succinate, a weak acid produced by phosphoenolpyruvate fermentation, in samples from women with BV[177]. We identified a small number of spectra matching succinate dehydrogenases from *Porphyromonas* and *Prevotella*, indicating organisms in these genera may be primarily responsible for changing concentrations of this metabolite. Curiously, we also identified a substantial number of spectra matching alcohol dehydrogenases. The majority of these spectra exclusively matched *Gardnerella* proteins, with others also matching sequences from Candidatus *Lachnocurva, Mageeibacillus*, and *Sneathia.*

In addition to fermentation enzymes, we also found spectra that matched proteins involved in nitrogen metabolism. Peptides can be a rich source of nitrogen for bacteria[178], and we found spectra matching extracellular proteases from both *Gardnerella* and *Prevotella*. A large number of spectra matched glutamate dehydrogenases, which can play a role in nitrogen assimilation[179,180]. These spectra matched proteins from multiple BVAB including *Aerococcus, Arcanobacterium, Fannyhessea, Gemella, Megasphaera, Mobiluncus, Porphyromonas*, and *Prevotella*. Notably, none of these spectra matched *Gardnerella* proteins, and a BLASTp search of published *Gardnerella* genomes did not identify any glutamate dehydrogenase homologs in this genus. However, we did find spectra that exclusively matched glutamine synthetases and glutamate transporters from *Gardnerella*, which interconvert glutamine/glutamate and transport glutamate across the cell membrane, respectively[181,182]. Diverse BVAB may, therefore, participate in nitrogen cross-feeding with *Gardnerella* through glutamate. Of note, although we identified

many spectra matching pyruvate formate lyase and glutamate dehydrogenase homologs from a wide range of BVAB, we did not find either of these enzymes from *Lactobacillus* spp. A BLASTp search also did not uncover any homologs of these proteins in published genomes of lactobacilli most associated with vaginal health, including *L. crispatus, L. gasseri*, or *L. jensenii.*

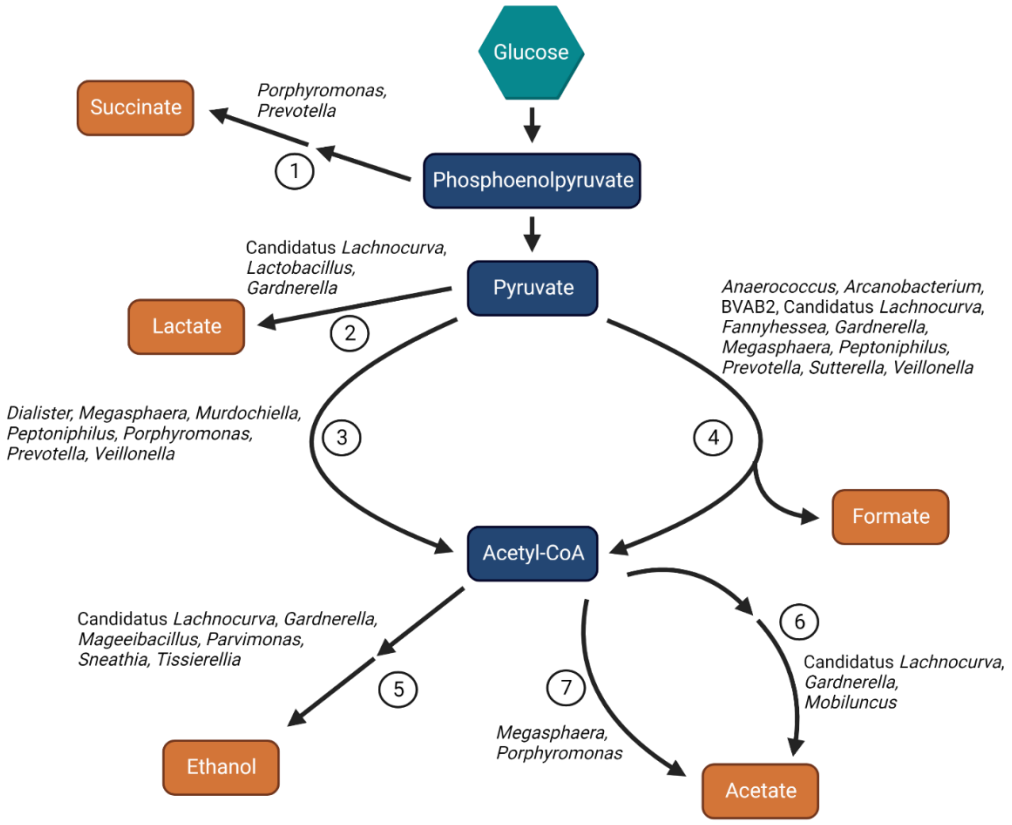
Table 9. Identified bacterial proteins of interest.

Protein	Associated Taxa	BV- Total Spectral Count	BV+ Total Spectral Count
Type I Pullulanase	*Gardnerella*	0	32
Secreted Alpha-Amylase	*Gardnerella*	0	40
Type I Pullulanase	*Lactobacillus crispatus*	6	0
SusC Polysaccharide Import Protein	*Prevotella, Porphyromonas uenonis*	0	161
SusD Polysaccharide Binding Protein	*Prevotella*	0	213
SusE Polysaccharide Binding Protein	*Prevotella*	0	34
Lactate Dehydrogenase	*Lactobacillus*	168	37
Lactate Dehydrogenase	*Gardnerella*	0	49
Lactate Dehydrogenase	Candidatus Lachnocurva	0	8
Pyruvate Formate Lyase	11 BVAB Genera	0	275
Phosphate Acetyltransferase	Candidatus Lachnocurva	0	5
Phosphate Acetyltransferase	*Gardnerella*	0	52
Acetate Kinase	*Gardnerella*	0	32
Acetate Kinase	*Gardnerella, Mobiluncus*	1	41
Acetyl-CoA Hydrolase	*Megasphaera, Porphyromonas*	0	14
Succinate Dehydrogenase	*Porphyromonas, Prevotella*	0	14
Alcohol Dehydrogenase	*Gardnerella*	0	72
Alcohol Dehydrogenase	Candidatus Lachnocurva, *Gardnerella, Mageeibacillus, Parvimonas, Sneathia, Tissierellia*	3	68
Pyruvate:Ferredoxin Oxidoreductase	*Dialister, Megasphaera, Murdochiella, Peptoniphilus, Porphyromonas, Prevotella, Veillonella*	0	80
Vaginolysin	*Gardnerella*	0	301
Extracellular Peptidase	*Gardnerella*	0	15
Extracellular Peptidase	*Prevotella*	0	14
Glutamate Dehydrogenase	*Aerococcus, Arcanobacterium, Fannyhessea, Gemella, Megasphaera, Mobiluncus, Porphyromonas, Prevotella, Streptococcus*	0	167
Glutamate Transporter	*Gardnerella*	0	32
Glutamine Synthetase	*Gardnerella*	0	33
Arginine Deiminase	*Fannyhessea vaginae*	0	6
Arginine Deiminase	*Sneathia vaginalis*	0	8

Ornithine Carbamoyltransferase	*Fannyhessea, Sneathia, Mageeibacillus, Parvimonas*	0	11
Ornithine Decarboxylase	*Dialister micraerophilus*	0	9

Bacterial proteins with potential biological significance identified in CVL samples. All taxa associated with proteins matching the sample spectra are listed, along with total spectral count across all BV- or BV+ samples.

Figure 16. Bacterial fermentation in the vagina.



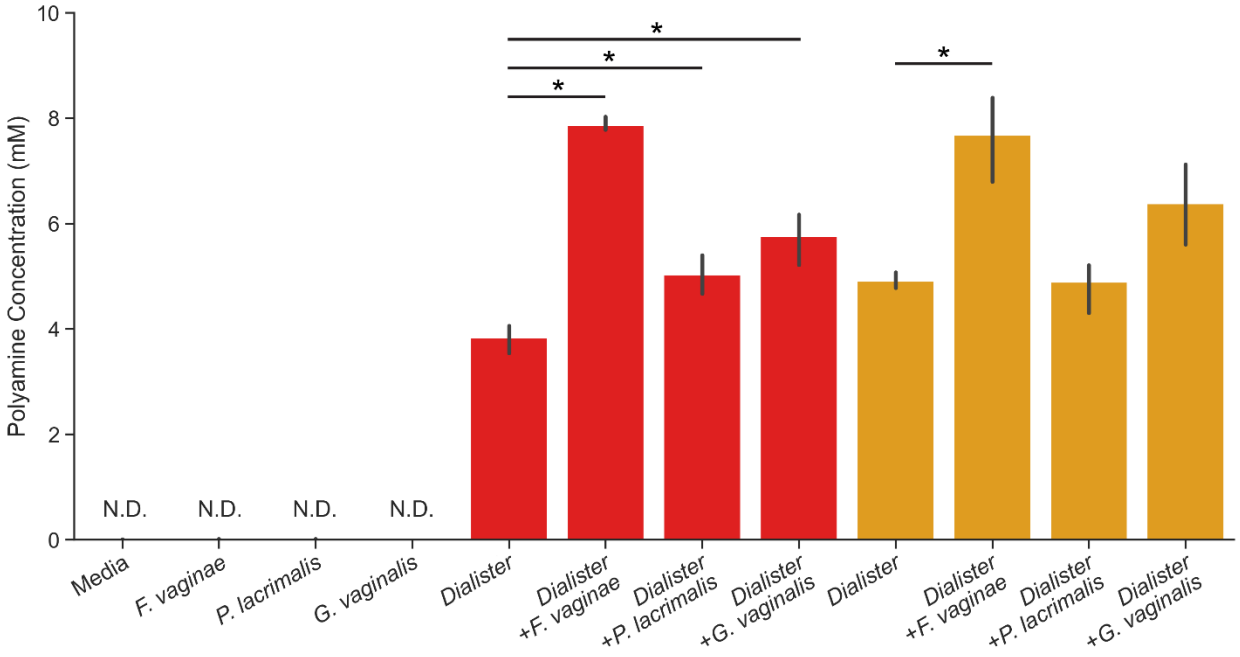
Summary of bacterial fermentation pathways observed in CVL samples. Major intermediates are shown in blue boxes and major fermentation products are shown in orange boxes. Genera associated with observed sample proteins are listed next to each enzymatic step. Numbers group different enzymatic pathways: 1) Malate dehydrogenase, fumarate hydratase, and succinate dehydrogenase. 2) Lactate dehydrogenase. 3) Pyruvate:Ferredoxin oxidoreductase. 4) Pyruvate formate lyase. 5) Acetaldehyde dehydrogenase and alcohol dehydrogenase. 6) Phosphate acetyltransferase and acetate kinase. 7) Acetyl-CoA Hydrolase. Made with BioRender.com.

Dialister micraerophilus cooperates with other BVAB to increase polyamine synthesis

While examining the bacterial proteins in our samples, we noticed the entire biosynthetic pathway for converting arginine into putrescine was present, but with different enzymes associated with different species. This pathway begins with arginine deiminase (ADI) which converts arginine to citrulline. Ornithine carbamoyltransferase (OCT) converts citrulline into ornithine, and finally ornithine decarboxylase (ODC) performs the final catalytic step by converting ornithine into putrescine. We found spectra matching ADI and OCT from both *Fannyhessea vaginae* and *Sneathia vaginalis*, and ODC from *Dialister micraerophilus*, but the complete pathway for putrescine biosynthesis did not appear to be present in any one of these bacterial species (Fig. 18D). BLASTp searches of *F. vaginae* and *S. vaginalis* genomes did not identify any ODC homologs, and conversely, searches of *D. micraerophilus* genomes did not identify any homologs of ADI or OCT. Therefore, we hypothesized that these organisms may cooperate to synthesize putrescine. To test this hypothesis, we grew two isolates of *D. micraerophilus* in mono-culture and co-culture with *F. vaginae* and quantified the concentration of polyamines present in the culture supernatants using an enzymatic assay kit. We also tested mono- and co-cultures with *Gardnerella vaginalis* and *Peptoniphilus lacrimalis*. *G. vaginalis* does not appear to encode any ornithine-producing enzymes, while *P. lacrimalis* encodes a homolog of arginase, an enzyme which directly converts arginine into orninthine[183]. *F. vaginae, G. vaginalis,* and *P. lacrimalis* did not synthesize detectable concentrations of polyamines in mono-culture (Fig. 17), but both isolates of *D. micraerophilus* produced a high concentration of polyamines when growing on their own. In line with our hypothesis, both *D. micraerophilus* isolates produced significantly higher concentrations of polyamines in co-culture with *F. vaginae* (Welch's t-test, $P < 0.05$). The results were less clear for co-cultures with the other two species. While *D. micraerophilus* DSM19965 produced a significantly higher concentration of polyamines in co-culture with both *P. lacrimalis*

and *G. vaginalis*, differences in polyamine concentrations for the *D. micraerophilus* DNF00843 co-cultures with these other bacteria were not statistically significant.

Figure 17. Polyamine production by *D. micraerophilus* in mono- and co-culture with other BVAB.

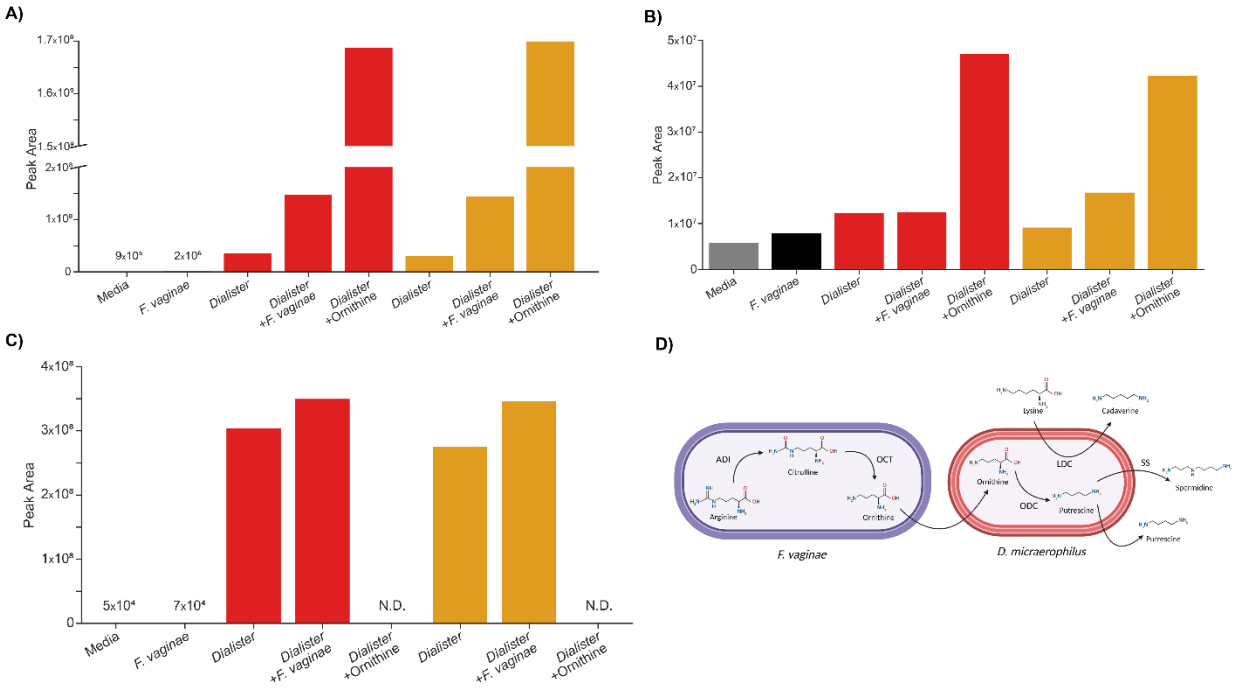


Total polyamine concentrations in supernatants from cultures of *Dialister micraerophilus* strains DSM19965 and DNF00843 grown anaerobically in mono- and co-culture with *Fannyhessea vaginae* DSM15829, *Peptoniphilus lacrimalis* DNF00528, and *Gardnerella vaginalis* ATCC14018 in Brucella H&K media for 72hrs. Cultures with *D. micraerophilus* DSM19965 are shown in red and cultures with *D. micraerophilus* DNF00843 are shown in orange. Bars show standard error of measurements from three separate cultures. N.D.: Measured polyamine concentrations were below the limit of detection for the assay. Stars show statistically significant differences (Welch's T-test, $P < 0.05$).

Although the results of the enzymatic assay demonstrated that *D. micraerophilus* synthesizes a higher concentration of polyamines when in co-culture with *F. vaginae*, it was not clear whether those higher polyamine concentrations were due to increased putrescine synthesis, or increased synthesis of other polyamines such as cadaverine and spermidine. To determine what polyamines *D. micraerophilus* synthesized in these cultures, we performed targeted metabolomics by liquid chromatography tandem mass spectrometry (LC-MS/MS) for putrescine, spermidine, spermine, and cadaverine. We pooled supernatants from all three replicates of the

D. micraerophilus mono-cultures and co-cultures with *F. vaginae* for analysis, and also pooled culture supernatants of both *D. micraerophilus* isolates in media supplemented with 50µg/mL L-ornithine to determine whether this precursor alone could increase putrescine biosynthesis. We were not able to detect spermine in any of the samples, but we did find putrescine in all eight pooled culture supernatants (Fig. 18A). The concentration of putrescine was higher in co-cultures of *D. micraerophilus* with *F. vaginae* than mono-cultures, and much higher when the culture media was supplemented with ornithine. In contrast, spermidine was mostly increased when exogenous ornithine was added to the culture media (Fig. 18B). Interestingly, while the two *D. micraerophilus* isolates synthesized cadaverine both in mono-culture and co-culture with *F. vaginae*, cadaverine was undetectable in cultures supplemented with ornithine (Fig. 18C). These results suggest that on their own, *D. micraerophilus* makes high concentrations of cadaverine and low concentrations of putrescine, and other BVAB like *F. vaginae* can supply these organisms with ornithine to increase putrescine biosynthesis (Fig. 18D).

Figure 18. Syntrophic biosynthesis of putrescine by *D. micraerophilus* in cooperation with *F. vaginae*.



Supernatants from three replicates of the *D. micraerophilus* DSM19965 (red) and DNF00843 (orange) mono-cultures and co-cultures with *F. vaginae*, in addition to cultures supplemented with 50µg/mL L-ornithine, were pooled and analyzed by targeted LC-MS/MS for A) putrescine, B) spermidine, and C) cadaverine. Spermine was not detected in any culture supernatants. Numbers show measured values that were too small to see on the scale. N.D.: Concentrations of the specified polyamine were below the limit of detection. D) Model of polyamine synthesis by *D. micraerophilus* aided by *F. vaginae*. ADI: arginine deiminase, OCT: ornithine carbamoyltransferase, ODC: ornithine decarboxylase, SS: spermidine synthase, LDC: lysine decarboxylase. Created with biorender.com.

Signs of heme and iron sequestration in BV.

Iron is a vital nutrient for many pathogenic bacteria, and is required for the BV-associated bacteria *Gardnerella* to grow[184-186]. In our samples, we identified mass spectra matching iron-dependent pyruvate:ferredoxin oxidoreductase from a wide range of BVAB, implying iron is important for many bacterial taxa associated with dysbiosis (Table 9). We also identified a large number of spectra matching the iron-binding host proteins serotransferrin and lactotransferrin, but the abundances of these proteins were not significantly different based on BV status (Table 10). Similarly, none of the seven hemoglobin subunit proteins we identified were significantly differentially abundant. However, two proteins involved in heme detoxification were differentially

abundant based on sample BV status. In BV+ samples, we found a significantly higher abundance of heme-sequestering hemopexin[187,188] and a significantly lower abundance of heme oxygenase 1, an enzyme which degrades heme into biliverdin and free iron[189] (Table 10).

Table 10. Identified human proteins of interest.

Protein	p-value	BV- Avg Count	BV+ Avg Count	Function Category
Serotransferrin	0.063	41.0	49.8	Iron Binding
Lactotransferrin	0.11	52.7	58.4	Iron Binding
Hemoglobin subunit alpha	0.45	442	242	Iron Binding
Hemoglobin subunit beta	0.40	285	156	Iron Binding
Hemoglobin subunit gamma-1	0.43	18.9	8.55	Iron Binding
Hemoglobin subunit gamma-2	0.43	17.9	8.45	Iron Binding
Hemoglobin subunit delta	0.34	127	60.7	Iron Binding
Hemoglobin subunit epsilon	0.42	13.0	6.40	Iron Binding
Hemoglobin subunit zeta	0.49	4.33	2.10	Iron Binding
Neutrophil elastase	0.42	12.6	16.8	Protease

Identified human proteins are listed with their average spectral count in samples from women without BV (N=9) and samples from women with BV (N=20). For analysis, the spectral count of a protein was divided by the total number of spectra matching human proteins in the sample to calculate a relative abundance for the protein. The relative abundance was then log-2 transformed and transformed relative abundances were compared between BV- and BV+ samples by Mann-Whitney U test. P-values for Mann-Whitney U tests are listed for each identified protein.

Signs of redox stress associated with BV

Early research on the vaginal microbiome found that hydrogen peroxide-producing lactobacilli are associated with health[98,190-193], but subsequent studies have cast doubt on whether the vaginal lumen contains enough oxygen for hydrogen peroxide to have any antimicrobial effect[39,194]. Curiously, although the vaginal epithelium exists in a reduced state as measured by calomel electrode[195], subsequent metabolomic analysis has found metabolite ratios indicative of oxidative stress[177]. In line with these later reports, we found a significantly higher abundance of the host antioxidant enzymes superoxide dismutase[196] and peroxiredoxin-4[197] in samples from BV+ participants (Table 8). Mirroring these differences in the host proteome in BV, we also found a higher relative abundance of bacterial peroxiredoxins in BV+ samples (Table 7).

BV is associated with a disrupted epithelial and mucosal barrier

Among the human proteins we identified in our CVL samples, 16 proteins that play a role in epithelial structure were differentially abundant based on BV status. Of these, 15 had significantly lower abundance in BV+ samples (Table 8). The host protease neutrophil elastase is known to target epithelial proteins and cause tissue damage under inflammatory conditions[198,199], but interestingly, this enzyme was present in both BV+ and BV- samples at similar abundances (Table 10). We did, however, find a significantly lower abundance of multiple host protease inhibitors in BV+ samples, including leukocyte elastase inhibitor (Table 8). While the majority of host proteins had a lower abundance in BV+ samples, the protein cross-linking enzyme transglutaminase 3 (TGase3) had a significantly higher abundance.

In addition to a disrupted epithelial barrier, there was also evidence of disruptions in the mucosal barrier in BV. We found a higher relative abundance of the mucus proteins Mucin-5B and Mucin-5AC in BV+ samples (Table 8). Innate immune proteins such as immunoglobulins and antimicrobial peptides are also present in the mucus barrier and form an additional layer of protection between the host and bacteria[10,11,200]. We found a significantly higher abundance of complement C3 in BV+ samples, one of the central proteins in the complement pathway responsible for formation of the antimicrobial membrane attack complex[201].

Discussion

BV is a complex condition characterized by dramatic changes in the functions performed by the vaginal microbiota. Metaproteomics is therefore a valuable tool both to observe broad proteomic changes in BV, and also to identify functions being performed by specific organisms. Our observations are in line with past metaproteomic studies of BV which found disruptions in the epithelial and mucosal barrier, a decrease in antiproteases, and a diversification of bacterial metabolism away from lactate fermentation[143,144,146,147,153-155,202,203]. Combining our metaproteomic

data with qPCR data of total 16S copy numbers, we also found evidence that overall bacterial metabolism may be increased in BV. In agreement with past studies, women with BV had significantly higher 16S rRNA gene copy numbers than those without BV[204], and we also found a higher relative abundance of bacterial proteins involved in general metabolism and energy production such as RNA polymerase, ATP synthase, and ribosomal proteins.

Increased bacterial metabolism in BV may in part be driven by increased carbohydrate utilization. Although we identified pullulanase enzymes from commensal *L. crispatus*, bacterial secreted amylases were significantly more abundant in BV+ samples, suggesting BVAB, especially *Gardnerella,* more actively degrade vaginal glycogen than commensal lactobacilli. *Gardnerella* were also the only group of BVAB that matched bacterial pullulanases in our samples, suggesting another major functional role for these organisms in the BV microbiota could be releasing carbohydrates from vaginal glycogen for all BVAB to consume. *Prevotella* spp. and *Porphyromonas uenonis* also encode pullulanase, but we did not identify any spectra matching these proteins. We did, however, identify a large number of spectra matching starch utilization system (Sus)[205] family proteins from these organisms. We identified SusC proteins, a porin involved in transporting polysaccharides across the cell membrane, from both *Prevotella* and *Porphyromonas*. We also identified a large number of spectra matching SusE and SusF starch-binding proteins from multiple species of *Prevotella.* These cell surface-anchored proteins bind branched polysaccharides such as starch and glycogen, making it easier for surface attached amylases to access them, and increasing the likelihood that released carbohydrates will be captured by the cell[206,207]. Although it is unclear whether *Prevotella* and *P. uenonis* participate in glycogen breakdown *in vivo* since we did not identify any of their glycogen-degrading enzymes among the peptide spectra, these results suggest that these organisms consume polysaccharides, and in the case of *Prevotella*, bind glycogen to increase their chances of scavenging carbohydrates.

Despite our data suggesting carbohydrate metabolism and competition is increased in BV, the literature is mixed on whether BV is associated with lower concentrations of vaginal carbohydrates. One past study did not find a significant difference in vaginal glycogen concentrations between women with a *Lactobacillus*- or non-*Lactobacillus*-dominated microbiota, but did find a positive correlation between increasing concentrations of glycogen and lactic acid[208]. The low environmental pH produced by high lactic acid concentrations may therefore suppress bacterial metabolism, leaving a higher concentration of intact glycogen. Notably, although past studies have found human salivary amylase present in vaginal fluid by ELISA assay[209] and metaproteomics[158], we did not identify any spectra matching human amylase enzymes in our CVL samples. It can be difficult to detect low-abundance proteins by mass spectrometry, and differences in peptide fragmentation or detection can lead to proteins being differentially detected. Still, the fact we identified roughly 10x as many host PSMs per sample as bacterial PSMs, and that we found more than 70 spectra matching bacterial proteins with predicted glycosidic activity, raises the interesting possibility that there are substantially more bacterial amylases present in the vagina than human amylases. The only host glycogen-active enzyme we did identify was glycogen phosphorylase, which had a significantly lower abundance in BV+ samples. This enzyme catalyzes the removal of glucose-1-phosphate from the outermost chain of glucose residues in a molecule of glycogen[210], although it is unclear whether the enzyme catalyzes the reaction in the catabolic or anabolic direction in the vaginal lumen.

Using taxonomic information associated with our metaproteomic data, we were able to tie specific bacterial taxa to different functions. Notably, we found evidence of secreted amylases from both *L. crispatus* and *Gardnerella*. Past studies have stated that commensal lactobacilli cannot directly metabolize glycogen and are reliant on the activity of human α-amylase to release fermentable sugars from the polysaccharide[209,211,212], but more recent reports have found substantial evidence that commensal lactobacilli encode pullulanase enzymes that can remove

carbohydrates from glycogen and express these enzymes *in vivo*[158,213,214]. Additionally, a growing body of evidence suggests that a wide range of commensal lactobacilli and BVAB encode secreted amylases that allow them to utilize carbohydrates stored in vaginal glycogen[215-217]. As acidic vaginal pH resulting from the fermentation of glycogen is a major barrier to pathogen colonization of the vagina, utilization of glycogen by commensal lactobacilli and BVAB warrants further investigation.

Past metabolomics studies have identified changes in metabolite concentrations in BV including lower concentrations of lactate and higher concentrations of succinate and acetate[177,218,219]. Our study provides additional proteomic data supporting these observations, as bacterial lactate dehydrogenases were significantly less abundant in BV+ samples, and although succinate dehydrogenases were not significantly more abundant in BV, we only detected homologs of this enzyme in BV+ samples, which matched proteins from *Prevotella* and *Porphyromonas*. The metabolomics studies referenced above did not report on differences in other fermentation products, but fermentative enzymes were among the most commonly identified bacterial proteins in our samples. We identified spectra matching acetate- and ethanol-producing enzymes from common BVAB including Candidatus *Lachnocurva*, *Gardnerella, Megasphaera, Mageeibacillus,* and *Sneathia*, among others. Succinate, acetate, and ethanol all reduce the pH of their solvent much less than lactate[220-222], so redirecting metabolic carbon flux away from lactate to other fermentation products may contribute to the increased pH observed in BV. Fermentation of these organic acids may also serve an immunomodulatory role for BVAB. One past report found that lactic acid has no effect on chemotaxis by human leukocytes, but succinate strongly inhibits leukocyte motility[223]. Thus, succinate fermentation by *Prevotella* and *Porphyromonas* could help inhibit immune responses to the BV microbiota. We also identified homologs of pyruvate formate lyase from a wide range of BVAB, which was unexpected, as formate has a similar acidity to lactate[224,225]. It is unclear, however, in which direction these pyruvate formate

lyases are performing their enzymatic reaction. Although some BVAB may be converting pyruvate into acetyl-CoA and formate as part of fermentation, others may be running the reaction in reverse; using formate as a source of carbon for anabolic reactions. Further research is required to determine the role of pyruvate formate lyase in the BV bacterial biota.

Similar to potential cross-feeding of carbon *via* formate, we also found signs of nitrogen cross-feeding *via* glutamate. Glutamate dehydrogenase facilitates nitrogen assimilation by fixing ammonia into glutamate and is known to play a role in the pathogenesis of other bacteria[179,226]. We identified spectra matching glutamate dehydrogenases from nine genera of BVAB, with the notable exception of *Gardnerella*. We did identify glutamine synthetases from *Gardnerella,* though, which interconverts ammonia and glutamate into the amino acid glutamine[227]. Whether *Gardnerella* uses this enzyme for anabolism or catabolism may depend on available nutrients, but past studies have found *Prevotella* promotes *Gardnerella* growth by feeding them ammonia[104]. Therefore, other BVAB may support the growth of *Gardnerella* by feeding these organisms both glutamate and ammonia to use for biosynthesis.

Fannyhessea vaginae is also known to form synergistic interactions with BVAB including *Gardnerella*[102,103]*.* In this study, we discovered a new syntrophic relationship between *Fannyhessea vaginae* and *Dialister micraerophilus*. *D. micraerophilus* is known to encode genes involved in spermine and spermidine production[228], and multiple studies have associated this organism with malodor and increased levels of the foul-smelling polyamines putrescine and cadaverine[25,218,219,229]. In our CVL samples, we identified spectra matching the two enzymes necessary to convert arginine to ornithine in both *Sneathia vaginalis* and *Fannyhessea vaginae*, but only identified ornithine decarboxylase, which synthesizes putrescine from ornithine, from *D. micraerophilus*. These metaproteomic findings mirrored our genomic analysis, which failed to identify a complete biosynthetic pathway for putrescine in any one of these species. When we grew *D. micraerophilus* in monoculture, the bacteria synthesized extremely high concentrations

of polyamines – approximately 100-fold higher than in healthy saliva[230]. And in co-culture with *F. vaginae*, both isolates of *D. micraerophilus* produced significantly higher concentrations of polyamines. Not only do these data establish a syntrophic relationship between *F. vaginae* and *D. micraerophilus* for the production of polyamines, they also suggest *D. micraerophilus* could be responsible for a disproportionately large fraction of the foul-smelling polyamines that are characteristic of BV. We performed targeted metabolomics analysis by LC-MS/MS to determine what species of polyamines *D. micraerophilus* produces in mono- and co-culture with *F. vaginae*, in addition to mono-culture with excess ornithine. *D. micraerophilus* produced cadaverine and putrescine in mono- and co-culture, though it made a higher concentration of putrescine in co-culture with *F. vaginae.* Interestingly, growing *D. micraerophilus* with exogenous 50µg/mL L-ornithine led to a substantial increase in the synthesis of putrescine and spermidine, and the apparent abrogation of cadaverine synthesis. Thus, although cadaverine and putrescine are the products of different amino acid catabolic pathways, the regulation of these pathways in *D. micraerophilus* appears to be connected, with an excess of putrescine precursors leading to a shutdown of the cadaverine pathway. Although a higher concentration of foul-smelling polyamines is a well-established characteristic of BV, there has been little discussion of what biological function these compounds may serve the bacteria that synthesize them. The initial enzymatic steps of amino acid catabolism generate ATP and produce nitrogen that can be used to build biomass[231], but this does not explain why *D. micraerophilus* takes the extra step to convert the end products of amino acid breakdown into polyamines. The alkaline nature of these compounds may be one explanation[232]. In response to acidic conditions, *E. coli* has been shown to secrete cadaverine and putrescine which increases their environment's pH[233]. Cadaverine, putrescine, and spermidine produced by *D. micraerophilus* may serve a similar pH buffering function. Polyamines also have immunomodulatory functions, with high concentrations of these molecules suppressing inflammatory functions of various immune cells including lymphocytes, neutrophils, and macrophages[234]. Additionally, putrescine is known to inhibit the function of TGase3 by

competing for the enzyme's active site with target proteins[235]. Transglutaminases help stabilize healing wounds in epithelial tissue by forming crosslinks between extracellular matrix proteins, but putrescine can inhibit wound healing by reducing the mechanical strength of the damaged tissue[236,237]. Putrescine may therefore inhibit wound healing in the vaginal epithelium, making it easier for BVAB to access nutrients in vaginal tissue.

Disruption of the epithelial barrier is one of the most consistent observations in metaproteomic studies of BV[140,143,154]. In this study, we identified 15 host proteins that contribute to epithelial structure which had a lower abundance in BV. Many of these epithelial proteins were among the most commonly identified host proteins in our samples, and their reduction in BV may have contributed to the lower number of human PSMs we identified in samples from BV+ participants. Because these host epithelial proteins had a lower abundance in BV, they are also likely one of the main substrates for BVAB peptidases and source of amino acids for these bacteria. Although the relative abundance of the host protease neutrophil elastase was very similar between BV+ and BV- samples, host proteases also likely contribute to disruption of the epithelial barrier since multiple host protease inhibitors had significantly lower abundance in BV. It is unclear whether the BV microbiota influences the host to reduce expression of leukocyte elastase inhibitor, or whether the action of bacterial proteases degrade this protein, but future therapeutic approaches to treat BV may be improved by identifying and disrupting these signaling pathways, or supplementing treatment with antiproteases such as leukocyte elastase inhibitor to make the vaginal lumen less proteolytic.

There is conflicting evidence for the presence of oxidative stress in BV. Vaginal fluid has anti-oxidant properties[39,194] and a study measuring redox potential of the vaginal epithelium found conditions in BV were more reducing than oxidative[195]. However, a metabolomic study of CVL found that the ratio of reduced glutathione to oxidized glutathione in BV was indicative of oxidative stress[177]. In our data there was a higher abundance of both host and bacterial peroxidases in BV+

samples. This was surprising, as oxygen concentrations in the vagina are generally low, though past studies have found transient increases in oxygen during arousal or perturbations such as tampon insertion[4,238,239]. Free heme is one potential source of oxidative stress, as heme can react with oxygen to produce reactive oxygen species[240-243]. In our data, we identified the CD59-dependent *Gardnerella* cytotoxin vaginolysin, which can target red blood cells and release heme into the vaginal lumen[95]. We observed a higher abundance of the heme-sequestering host protein hemopexin in BV+ samples, which may also represent a host response to higher concentrations of free heme. Although iron sequestration is an important part of the innate immune system[184] and multiple studies have investigated the potential of iron-binding lactotransferrin to treat BV[81,82], we did not find significantly different abundances of either lactotransferrin or serotransferrin depending on BV status. Future studies of the role of heme and heme-sequestering proteins in BV may be valuable.

This study demonstrates the value of analyzing metaproteomic data from a high level to identify broad functional differences between microbial communities, and also parsing individual protein identifications to uncover important functions performed by different organisms. We identified signs of increased, diversified bacterial metabolism in BV, along with potential host responses to a damaged epithelial barrier and increased concentrations of free heme. We also discovered a new, syntrophic interaction between *D. micraerophilus* and *F. vaginae* to increase putrescine biosynthesis. Our metaproteomic data indicated additional, undescribed synergistic interactions between BVAB, including carbon cycling *via* the activity of pyruvate formate lyase, and cross feeding of glutamate to *Gardnerella*. Investigating these potential interactions could lead to new treatments to disrupt the mutually beneficial interactions between BVAB that stabilize the BV microbiota. Future metaproteomics studies analyzing more samples or using samples with higher protein concentrations, such as Softcup samples of cervicovaginal mucus, are likely to uncover even more functions.

Chapter 4: Glycogen Metabolism by Vaginal Bacteria

Background

Glycogen is the primary vehicle for carbohydrate storage in animals[244]. A homodimer of glycogenin proteins forms the core of a glycogen molecule, to which glucose residues are attached[245]. Additional glucose monomers are then added in chains by α-1,4 glycosidic bonds, forming helical structures that are generally 10 – 12 residues in length[246]. To maximize the storage capacity of glycogen, additional chains are attached using branching α-1,6 glycosidic bonds. Although the average size of glycogen molecules differs between body sites, a single molecule can include up to 55,000 glucose residues before steric hindrance prevents additional synthesis.

Extracellular glycogen is too large to transport across cellular membranes and must be broken down before it can be imported and metabolized[247]. But despite being composed of a single species of carbohydrate attached by only two types of glycosidic bonds, glycogen breakdown is deceptively complex. Glycogen can be catabolized by three types of enzymes: α-glucosidases, α-amylases, and pullulanases (Fig. 19). α-glucosidases break the outermost α-1,4 glycosidic bond on a chain, releasing glucose monomers[248]. α-amylases, in contrast, break interior α-1,4 glycosidic bonds. Different amylases prefer substrates of different lengths, and can often work on multiple sizes of polysaccharide, releasing glucose, maltose, or other small glucose polymers. Finally, pullulanases are active against α-1,6 glycosidic bonds, debranching entire chains from a molecule of glycogen. Adding to this complexity, many of these enzymes have overlapping activities, and it is difficult to predict substrate specificity by homology alone[249]. Thus, these enzymes are classed together in the glycoside hydrolase 13 family of carbohydrate-active enzymes[250].

Figure 19. Glycogen breakdown by GH13 family enzymes.

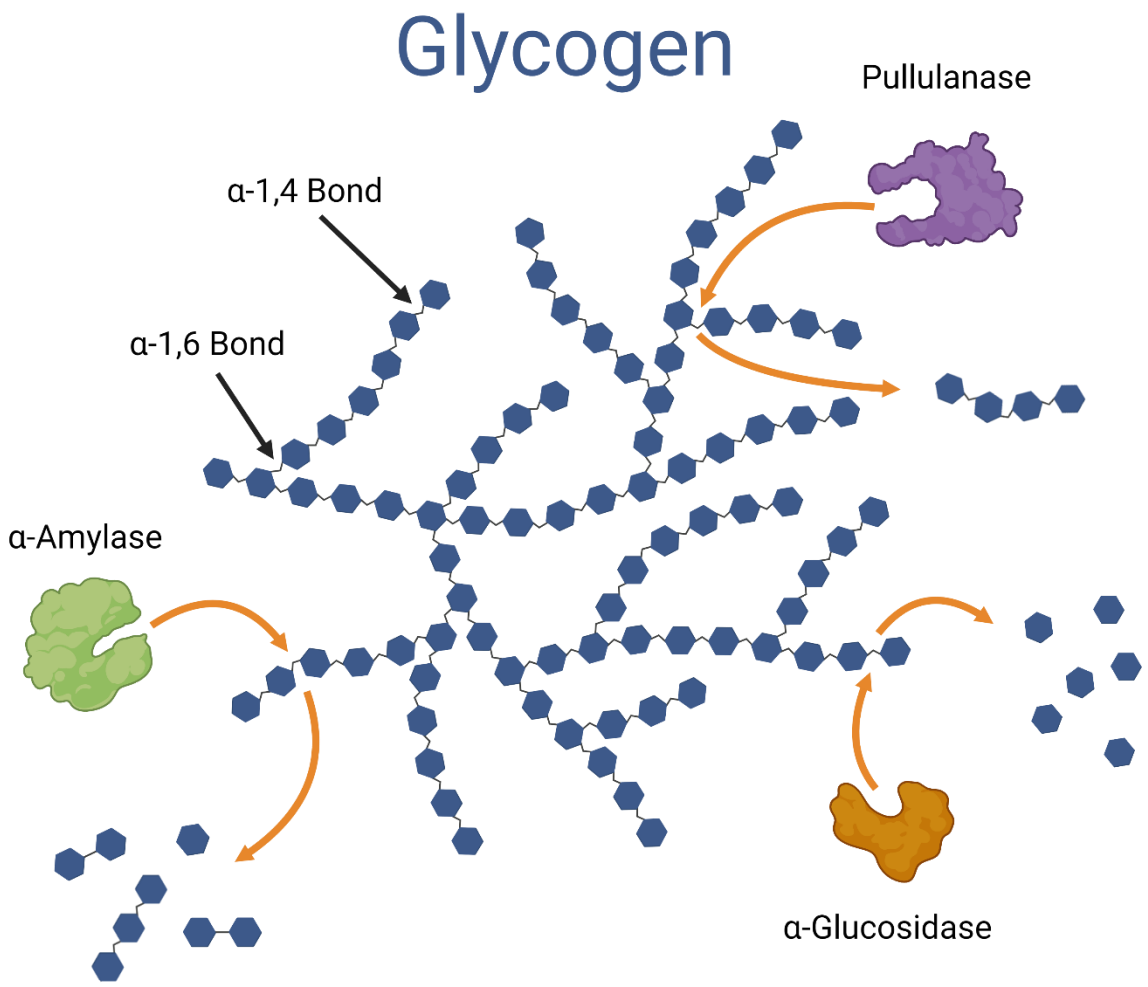


Diagram of glycosidic bonds in glycogen broken by glycoside hydrolase enzymes, and the carbohydrates they release when breaking down a glycogen molecule. Made with BioRender.com.

Although reliably quantifying glycogen can be difficult, estimates for glycogen concentrations in human vaginal fluid range from 0.1 – 32 µg/mL[158,251]. Early studies found few vaginal *Lactobacillus* isolates that could grow on glycogen, so commensal lactobacilli were assumed to be incapable of metabolizing the polysaccharide directly[211,212]. Thus, glycogen metabolism by commensal vaginal bacteria has been understudied. Curiously, although a study from 1979 found 100% of the 79 *Gardnerella* isolates they tested could grow on glycogen[252], glycogen metabolism by BVAB has similarly received little attention. Instead, most research on

the breakdown of vaginal glycogen has focused on human enzymes. Past studies have used enzyme-linked immunosorbent assays (ELISA) to prove that human salivary amylase is present in vaginal fluid, and that vaginal lactobacilli can grow on the breakdown products of this enzyme working on glycogen[209]. Other studies have found that host amylase is sensitive to pH, with reduced enzymatic activity as the environment becomes more acidic[253]. This result suggests that, in addition to its antibacterial effects, lactic acid may also regulate bacterial growth by influencing the rate of carbohydrate release from vaginal glycogen.

Despite the long-standing dogma, recent studies using metatranscriptomics and metaproteomics have demonstrated that commensal lactobacilli encode enzymes with glycoside hydrolase activity and express these enzymes *in vivo*[158,208,214,254,255]. Various BVAB also encode glycoside hydrolases[215-217], and our data discussed in Chapter 3 demonstrates they also express them in the vagina. These findings have major implications for BV, as they imply vaginal bacteria are much more proactive in metabolizing vaginal glycogen than previously thought. Competition for carbohydrates among vaginal bacteria would directly affect the vaginal microbiota by determining what organisms can access energy-rich nutrients. It would also affect vaginal pH, as BVAB primarily ferment carbohydrates into the relatively weak acids acetate and succinate, rather than lactate[177].

Little is known about glycogen breakdown by vaginal bacteria, or how the ability to metabolize this polysaccharide influences community dynamics of the vaginal microbiota. This chapter details a search for glycoside hydrolase enzymes in vaginal bacteria, and experiments to elucidate how these proteins impact bacterial fitness. We performed this study to test the hypothesis that diverse commensal lactobacilli and BVAB can metabolize glycogen, and glycogen-metabolizing bacteria will have a competitive advantage when glycogen is the primary carbohydrate in their environment. This work was carried out by Elliot Lee in collaboration with Sujatha Srinivasan, Tina Fiedler, Samuel Minot, and David Fredricks. My contributions to this

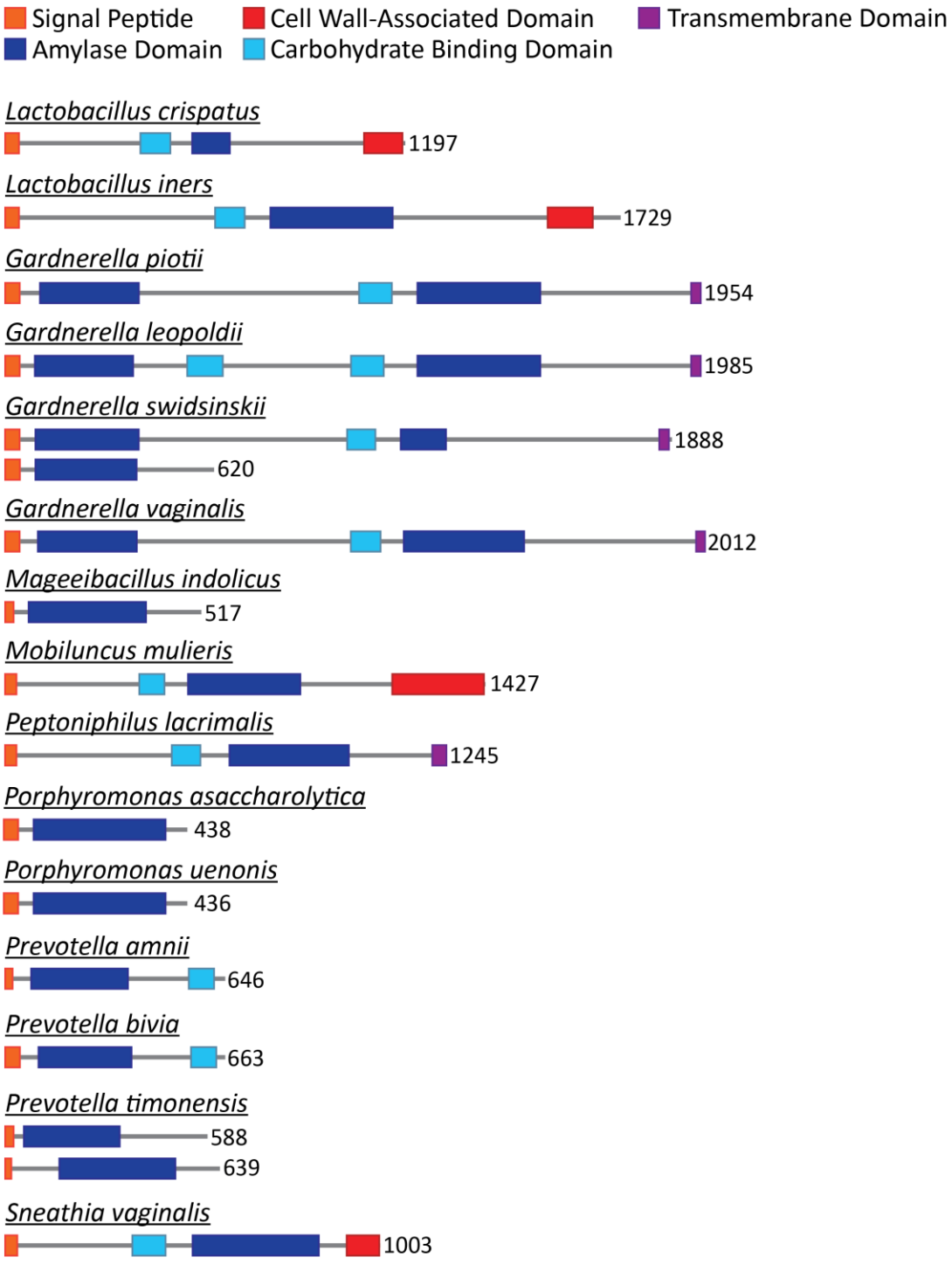
paper included performing experiments described in Figures 20 – 28 as well as analyzing the data and preparing written conclusions.

Results

Diverse vaginal bacteria encode enzymes with predicted glycosidase activity

We began by screening the genomes of common vaginal bacteria for secreted proteins with glycoside hydrolase activity. We used the dbCAN[256,257] server to identify genes with glycosidase domains and the Phobius[258] server to predict signal peptides. Using this approach, we identified secreted glycosidases in the genomes of 15 species of vaginal bacteria, including 13 species of BVAB (Fig. 20). Notably, while we identified such enzymes in the genomes of many *L. crispatus* and *L. iners* isolates, we could not find any evidence of these enzymes in other common vaginal lactobacilli including *L. gasseri, L. jensenii, L. johnsonii, L. mulieris,* or *L. vaginalis.*

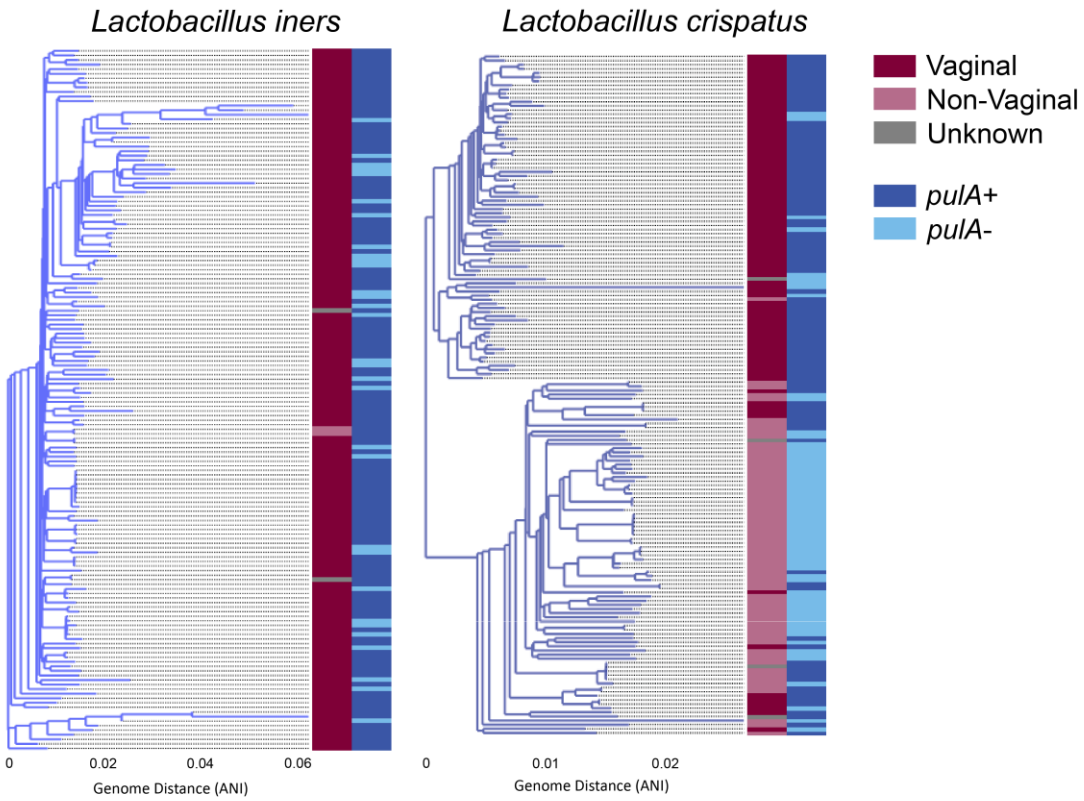
Figure 20. Secreted proteins from vaginal bacteria with predicted glycoside hydrolase domains.



Representative glycoside hydrolases with signal peptides from species of vaginal bacteria. The length of the protein in amino acids is shown next to each. Domain prediction performed by HMMER[259] homology analysis.

It was notable that *L. crispatus* and *L. iners* appeared to be the only vaginal lactobacilli encoding secreted glycosidases, so we analyzed the distribution of these genes in the two species (Fig. 21). Both bacteria possessed a single, large pullulanase enzyme encoded by a *pulA* gene. 79% (122/154) of *L. iners* isolates we screened were *pulA*+ while 61% (99/162) of *L. crispatus* isolates were *pulA*+. However, when we clustered *L. crispatus* isolates by average nucleotide identity (ANI), we noticed they generally grouped depending on whether the strain had been isolated from the human vagina or another source (human gut, fowl crop, etc.). Interestingly, vaginal *L. crispatus* isolates were significantly more likely to encode a *pulA* gene than non-vaginal isolates (Z test for population proportions, 78/89 vaginal isolates *pulA*+, 18/69 non-vaginal isolates *pulA*+, $P < 0.001$).

Figure 21. Distribution of *pulA* genes in *L. iners* and *L. crispatus*.



Distribution of *pulA* genes in *L. iners* and *L. crispatus*. Strains of the species were grouped by average nucleotide identity (ANI). Isolation source is shown for isolates of both species. Strains where isolation source could not be found are marked in gray.

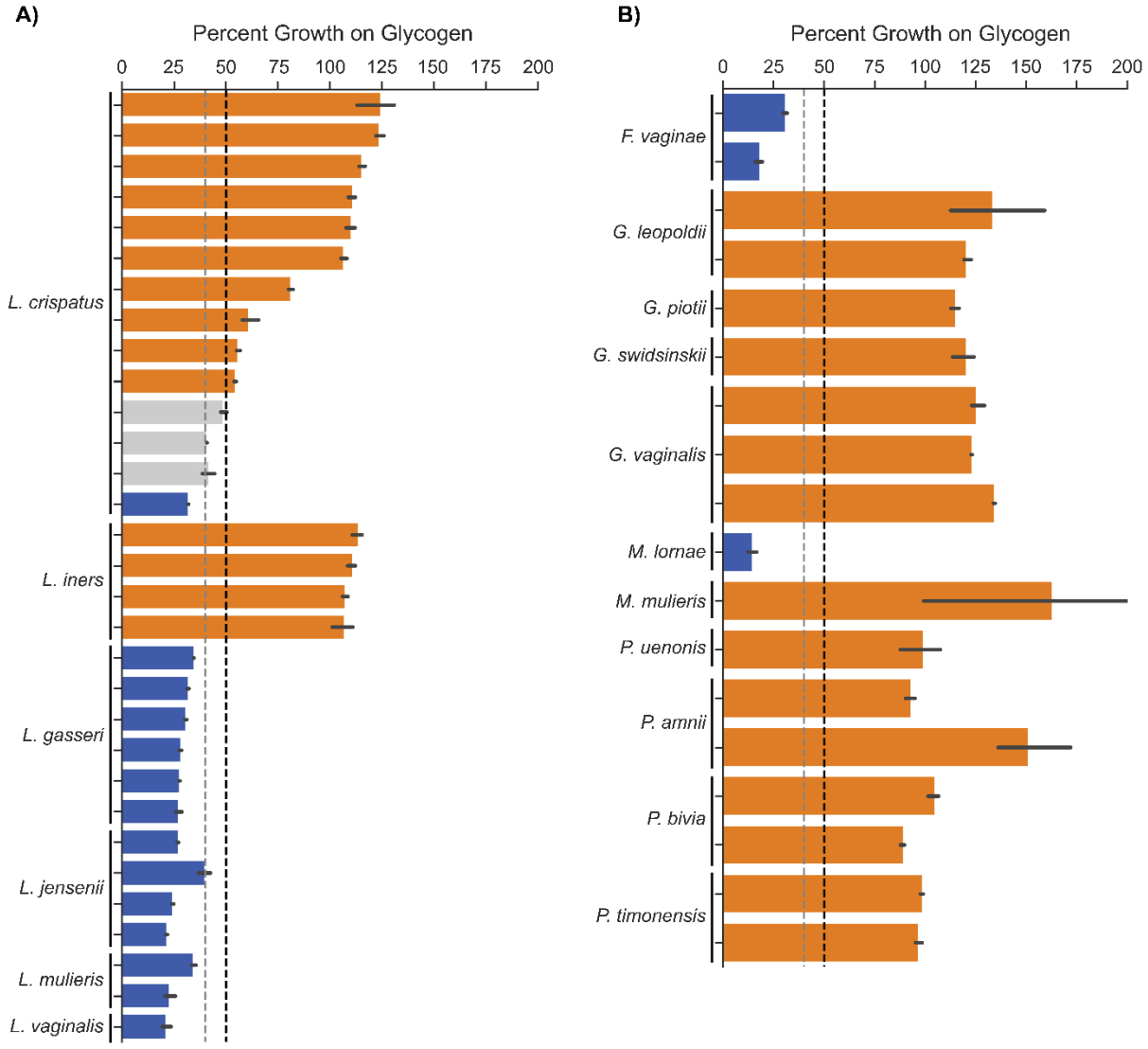
Diverse vaginal bacteria are capable of metabolizing glycogen

To verify the results of our genomic analysis, we cultured strains of vaginal bacteria in media with equal concentrations of either glucose or bovine liver glycogen. We observed a diversity of growth phenotypes, so based on our genomic analysis and observations of glycogen breakdown detailed in the next section, we defined a glycogen-metabolizer as an isolate which achieved >50% of the growth on glycogen as glucose, as measured by OD600. We defined a non-glycogen-metabolizer as a strain which grew to <40% of the density on glycogen as glucose. We labelled isolates that fell between these cutoffs as having an indeterminate phenotype. In line with our genomic analysis, *L. crispatus* and *L. iners* were the only vaginal lactobacilli that

metabolized glycogen (Fig. 22A). While all four *L. iners* isolates we tested had equivalent growth on glucose and glycogen, we found a range of growth phenotypes in *L. crispatus*. Most *L. crispatus* strains grew well on glycogen, but a few grew much slower on the polysaccharide and reached a relatively low percent growth compared to their growth on glucose. Finally, we found four strains of *L. crispatus* which did not meet our definition of a glycogen-metabolizing strain. The non-glycogen-metabolizing strain, *L. crispatus* 125-2-CHN, is a vaginal isolate without a *pulA* gene, supporting the hypothesis that a *pulA* gene enables lactobacilli to catabolize glycogen. In addition to *L. crispatus*, we also tested isolates of *L. gasseri*, *L. jensenii, L. mulieris,* and *L. vaginalis*. In agreement with our genomic analysis, all the isolates of these species we tested were non-glycogen-metabolizers.

We also cultured an assortment of BVAB on both glucose and bovine liver glycogen. We were not able to identify secreted glycosidases in genomes of *Fannyhessea vaginae* or *Megasphaera lornae*, and isolates of these species also did not show substantial growth on glycogen as compared to glucose (Fig. 22B). However, we identified numerous isolates of BVAB which were able to metabolize glycogen including *Mobiluncus mulieris, Porphyromonas uenonis*, *Prevotella amnii, Prevotella bivia, Prevotella timonensis*, and all four named species of *Gardnerella*.

Figure 22. Ability of vaginal bacteria to metabolize glycogen.

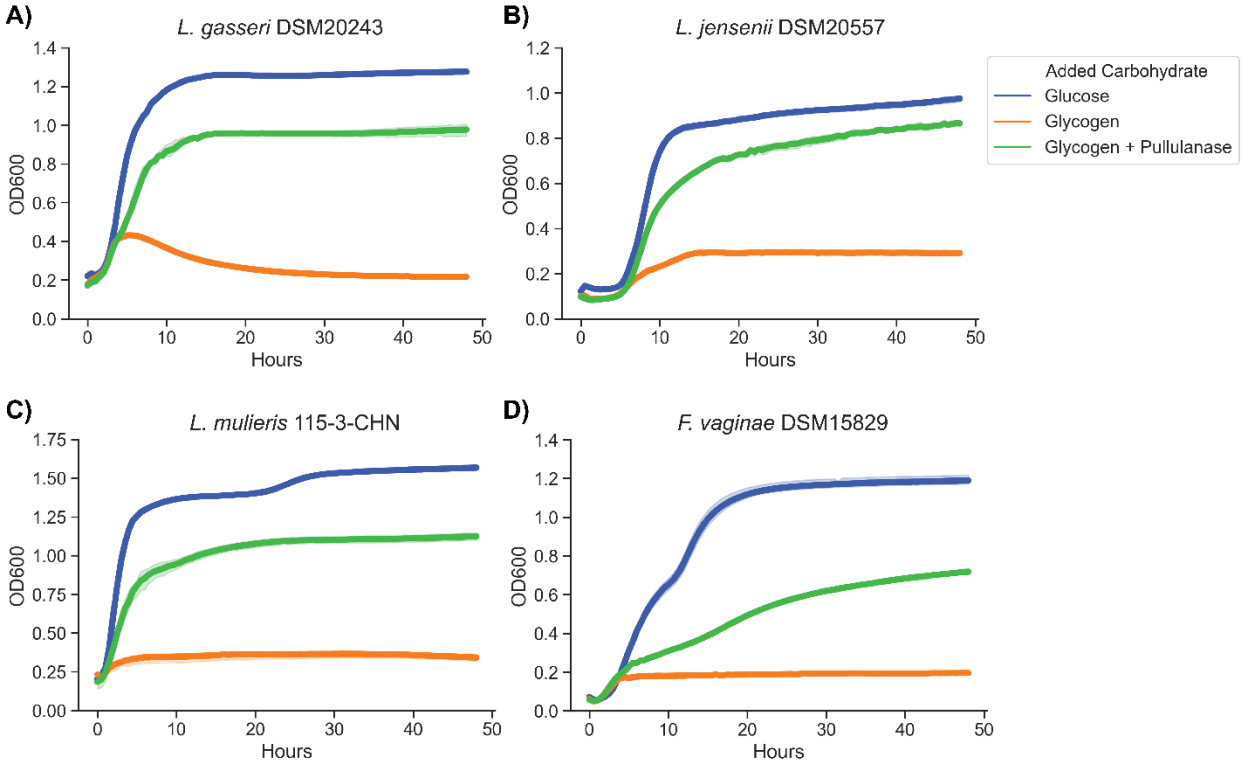


Percent growth of vaginal bacteria on bovine liver glycogen compared to their growth on an equivalent concentration of glucose. Growth was measured by OD600 after 48 hours of growth in anaerobic conditions at 37°C. Each bar represents a different isolate of the bacterial species. The black dotted line shows the 50% cutoff to identify an isolate as a glycogen-metabolizer. The gray dotted line shows the 40% cutoff to identify an isolate or a non-glycogen-metabolizer. Glycogen-metabolizers have orange bars, non-glycogen-metabolizers have blue bars. Isolates with an indeterminate phenotype that fall between these cutoffs have light gray bars. The black lines on each bar show standard error of three separate cultures.

To test whether non-glycogen-metabolizers failed to grow on glycogen due to an absence of extracellular glycoside hydrolases, we also cultured these organisms on bovine liver glycogen in the presence of an exogenous pullulanase enzyme. In the presence of pullulanase, we were

able to partially recover the growth of *L. gasseri, L. jensenii, L. mulieris,* and *F. vaginae* on glycogen (Fig. 23A-D).

Figure 23. Growth of non-glycogen-metabolizing bacteria in the presence of exogenous pullulanase.



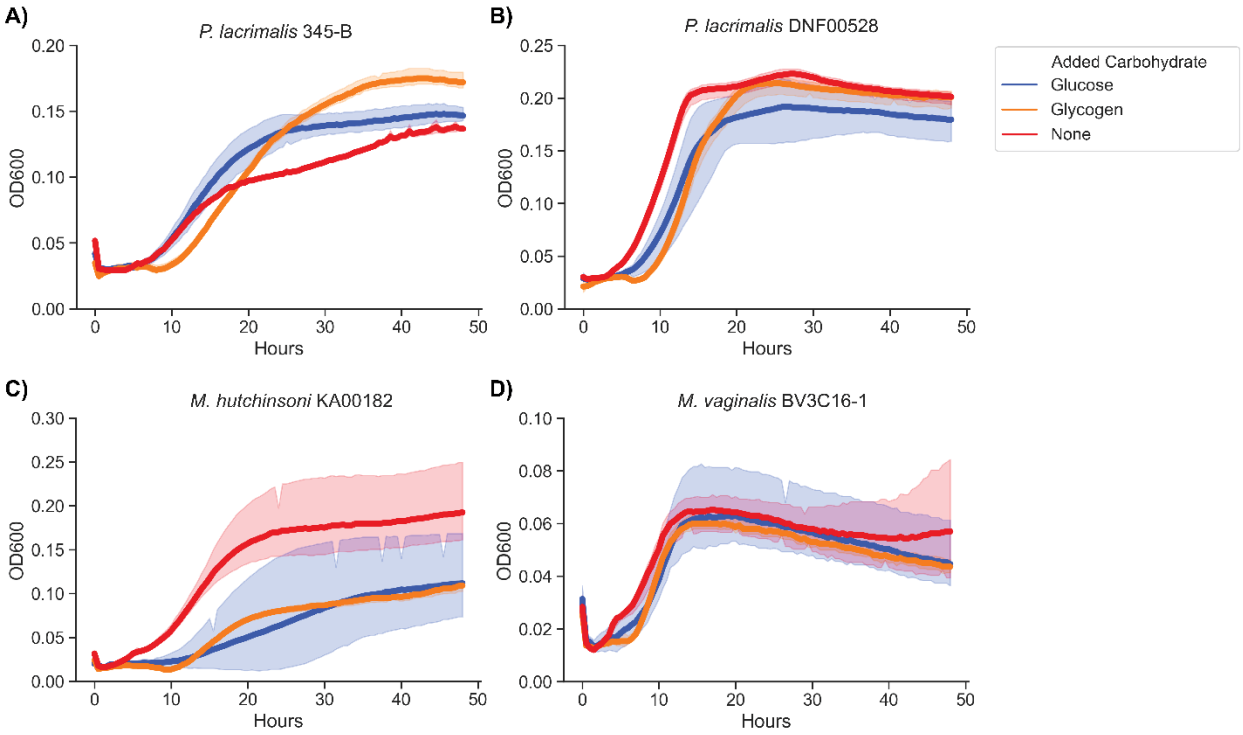
Growth of A) *Lactobacillus gasseri* DSM20243, B) *L. jensenii* DSM20557, C) *L. mulieris* 115-3-CHN, and D) *Fannyhessea vaginae* DSM15829 under anaerobic conditions at 37C in PYG-mod-YG broth. Bacteria were grown in broth with 1.5% glucose, 1.5% bovine liver glycogen, 1.5% bovine liver glycogen plus 10U/mL *Bacillus lichenformis* pullulanase. Light colored areas show the 95% confidence interval at each timepoint for three biological replicates.

Differential patterns of carbohydrate utilization by BVAB

Although *Peptoniphilus lacrimalis* is a species of BVAB which encodes a secreted glycosidase, the two isolates of this organism we tested grew to similar densities whether their media contained glucose, glycogen, or no added carbohydrates (Fig. 24A, 24B), indicating that in PYG-mod-YG media, these organisms may use different sources of nutrients besides carbohydrates. We observed similar growth patterns in *Megasphaera hutchinsoni* and

Megasphaera vaginalis, two species of BVAB which do not appear to encode any secreted glycoside hydrolases (Fig. 24C, 24D).

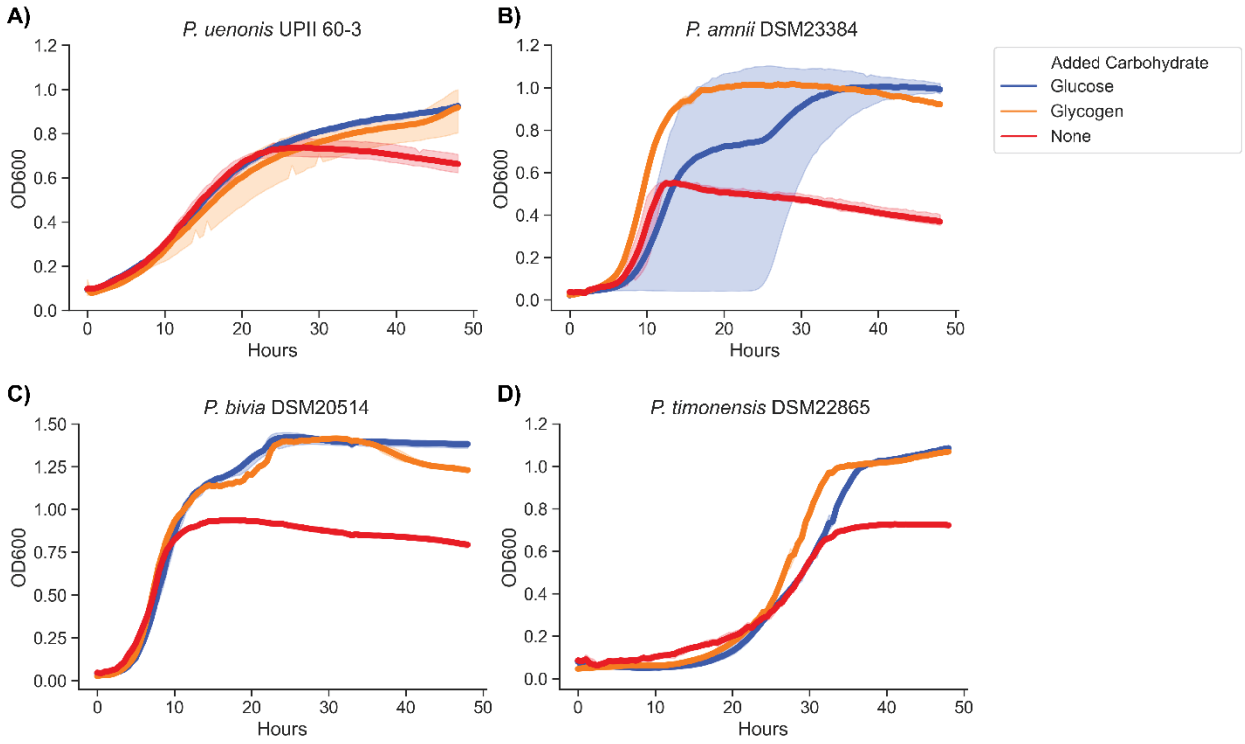
Figure 24. BVAB which do not metabolize carbohydrates in PYG-mod-YG media.



Growth of A) *Peptoniphilus lacrimalis* 345-B, B) *Peptoniphilus lacrimalis* DNF00528, C) *Megasphaera hutchinsoni* KA00182, and D) *Megasphaera vaginalis* BV3C16-1 under anaerobic conditions at 37C in PYG-mod-YG broth. Bacteria were grown in broth with 1.5% glucose, 1.5% bovine liver glycogen, or no added carbohydrates. Light colored areas show the 95% confidence interval at each timepoint for three biological replicates.

Analyzing the growth curves of other BVAB also uncovered interesting patterns. Although *P. uenonis* UPII 60-3 grew to a higher density on both glucose and glycogen compared to their growth in media with no added carbohydrates, initially, they grew nearly identically without added carbohydrates (Fig. 25A). All three species of *Prevotella* we tested displayed similar growth patterns (Fig. 25B-D). These growth dynamics indicate these bacteria may initially utilize non-carbohydrate sources of nutrients for growth, before switching to carbohydrate metabolism after their preferred nutrient sources have been depleted.

Figure 25. Glycogen-metabolizing BVAB which preferentially use non-carbohydrate nutrients.



Growth of A) *Porphyromonas uenonis* UPII 60-3, B) *Prevotella amnii* DSM23384, C) *Prevotella bivia* DSM20514, and D) *Prevotella timonensis* DSM22865 under anaerobic conditions at 37C in PYG-mod-YG broth. Bacteria were grown in broth with 1.5% glucose, 1.5% bovine liver glycogen, or no added carbohydrates. Light colored areas show the 95% confidence interval at each timepoint for three biological replicates.

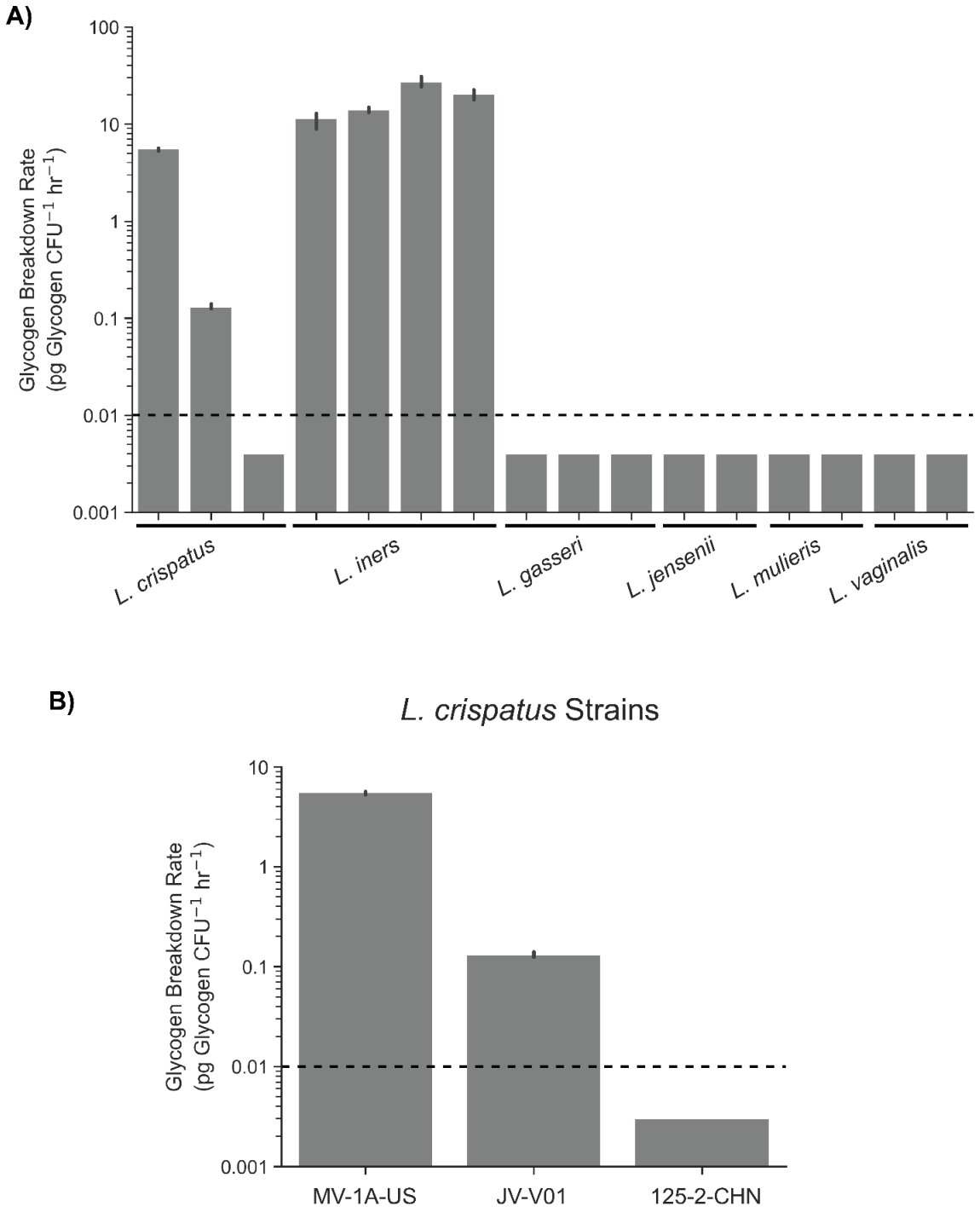
Glycogen breakdown by the enzymes of vaginal bacteria

Our culture experiments showed that diverse vaginal bacteria can utilize glycogen, though *L. crispatus* and *L. iners* may be the only commensal lactobacilli that are glycogen-metabolizers. Next, we sought to compare the glycogen catabolizing activity of the surface-attached and fully secreted glycosidase enzymes made by these bacteria. We cultured isolates of vaginal bacteria on their preferred agar media, then used a sterile swab to suspend the bacteria and their secreted enzymes in Maximum Recovery Diluent[260] (MRD) – an isotonic medium formulated to keep cells alive without providing them nutrients to grow. We incubated the cell suspension with 1mg/mL bovine liver glycogen anaerobically at 37°C for 15 minutes to 24 hours, pelleted the cells, then

quantified the amount of glycogen remaining in the supernatant with a Lugol's Iodine solution containing $CaCl_2$[261]. We also measured the density of live cells in the original bacterial suspension by performing colony forming unit (CFU) counting, which allowed us to quantify the rate of glycogen breakdown in pg glycogen degraded CFU^{-1} hr^{-1}.

We began by quantifying the rate of glycogen breakdown for the enzymes of commensal lactobacilli in MRD at pH 6.5. In line with our culture data, strains of *L. crispatus* and *L. iners* both had detectable rates of glycogen breakdown, but we could not observe any glycogen breakdown from strains of *L. gasseri, L. jensenii, L. mulieris* or *L. vaginalis* (Fig. 26A). We found additional concordance between results from the breakdown assay and culturing experiments in *L. crispatus*. *L. crispatus* MV-1A-US is a vaginal isolate that grew as well on glycogen as glucose in our culturing experiments, while *L. crispatus* JV-V01 barely passed the 50% relative growth threshold to identify it as a glycogen-metabolizer. In the glycogen breakdown assay, JV-V01 degraded glycogen approximately 10-fold slower than MV-1A-US (Fig. 26B). The pullulanase enzymes of these two isolates have identical amino acid sequences, so these differences may be due to differential expression. *L. crispatus* 125-2-CHN is a vaginal isolate that lacks a functional *pulA* gene and did not metabolize glycogen. In the breakdown assay, we did not observe detectable glycogen breakdown by the enzymes of *L. crispatus* 125-2-CHN (Fig. 26B).

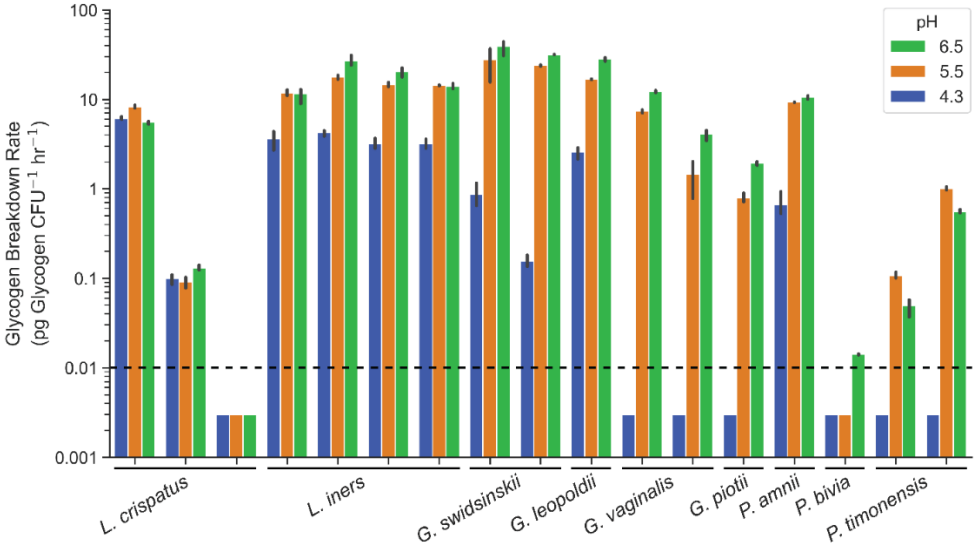
Figure 26. Glycogen breakdown by commensal lactobacilli.



Rate of glycogen breakdown by the enzymes of vaginal lactobacilli at pH 6.5. Each bar is a separate bacterial isolate. Black lines show standard error of three biological replicates. Black dotted line shows limit of detection for the assay. A) Glycogen breakdown rates of different species of commensal lactobacilli. B) Glycogen breakdown rates of vaginal isolates of *L. crispatus*.

We also measured the rate of glycogen breakdown for isolates from various species of BVAB. Because acidic pH is an important barrier to pathogen colonization of the vagina, we also hypothesized that acidic conditions would inhibit the activity of BVAB glycoside hydrolases, but not those of commensal lactobacilli. Therefore, we also measured the rate of glycogen breakdown at pH 4.3 (representing a eubiotic pH), and pH 5.5 (representing a normal pH during BV) by acidifying the solution with concentrated lactic acid. We observed glycogen breakdown from isolates of all four named species of *Gardnerella* as well as *P. amnii, P. bivia,* and *P. timonensis* (Fig. 27). In line with our hypothesis, acidic pH had little effect on the amylolytic enzymes of *L. crispatus*, as both glycogen-metabolizing strains MV-1A-US and JV-V01 had nearly identical rates of glycogen breakdown at all three pHs that we tested. In contrast, all four isolates of *L. iners* that we tested had a lower rate of glycogen breakdown at pH 4.3 compared to pH 5.5. Similarly, all the isolates of BVAB we tested had a significant drop in their rate of glycogen breakdown at pH 4.3, with rates for some species dropping below the limit of detection for our assay.

Figure 27. Effect of acidic pH on the glycoside hydrolase enzymes of vaginal bacteria.



Rate of glycogen breakdown by the enzymes of vaginal bacteria at pH 6.5 (blue), pH 5.5 (orange), and pH 4.3 (green). Each set of bars is a separate bacterial isolate. Black lines show standard error of three biological replicates. Black dotted line shows limit of detection for the assay.

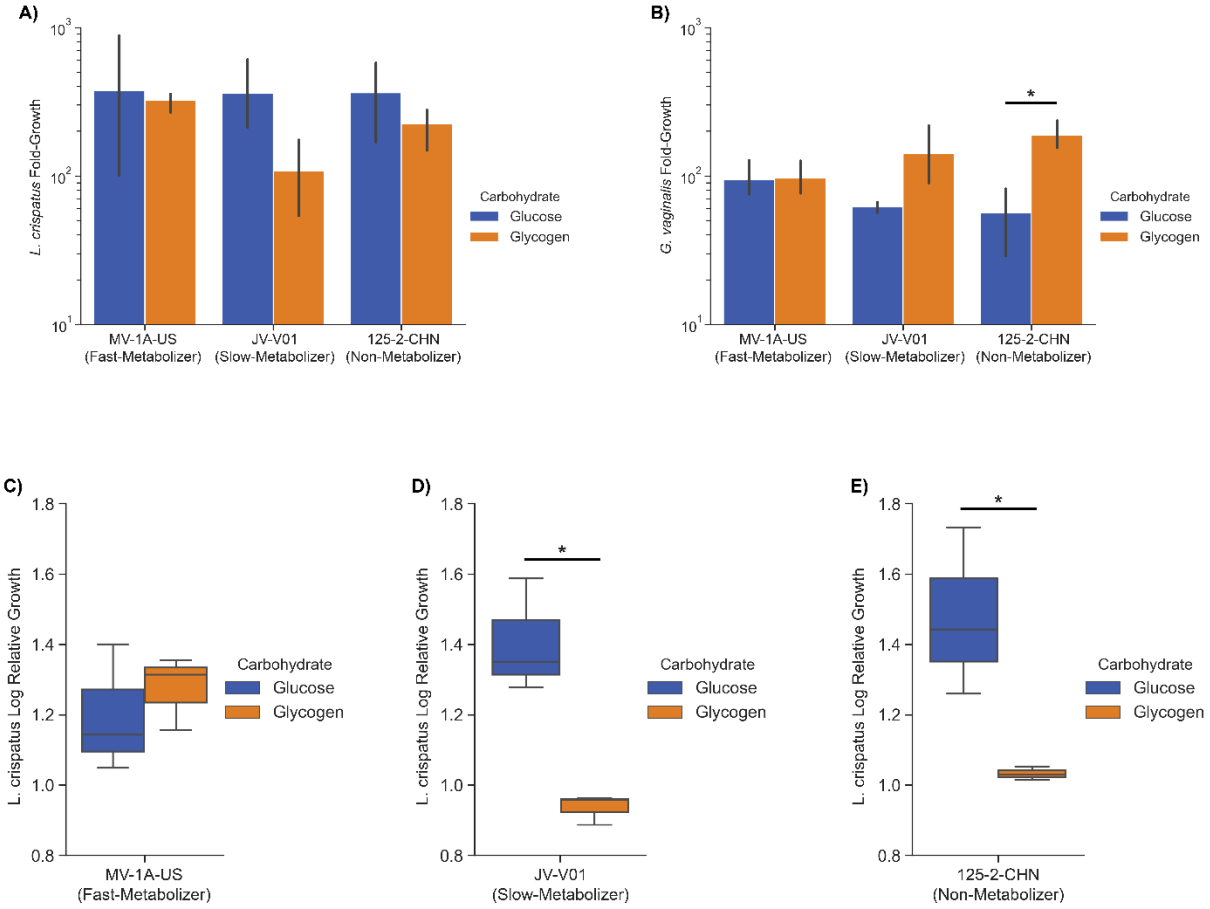
L. crispatus isolates with a defect in glycogen metabolism have a competitive disadvantage against *G. vaginalis* when growing on glycogen

In the above experiments, we observed large inter-strain differences in glycogen metabolism and breakdown between isolates of *L. crispatus.* Because *L. crispatus* must compete against diverse glycogen-metabolizing bacteria to maintain their dominance in the vaginal microbiota, we wanted to understand how glycogen metabolism can impact the relative fitness of *L. crispatus* against other organisms. We chose three isolates of *L. crispatus* to test: fast-metabolizing MV-1A-US, slow-metabolizing JV-V01, and non-metabolizing 125-2-CHN. We co-cultured each of these isolates with glycogen-metabolizing *G. vaginalis* ATCC14018 in PYG-mod media with either 0.5% glucose or 0.5% bovine liver glycogen. After 48 hours co-culture in anaerobic conditions, we extracted genomic DNA from the cultures, then used species-specific qPCR to measure the growth of each *L. crispatus* isolate and *G. vaginalis* on both types of carbohydrate.

We measured fold-growth of both organisms in the cultures by dividing the number of cells of each species present after 48 hours by the number inoculated into each culture. Growth of all three *L. crispatus* isolates was not significantly different between co-cultures on glucose and glycogen (Student's T-test, $P > 0.05$) (Fig. 28A), however, *G. vaginalis* grew significantly better on glycogen than on glucose in co-culture with non-metabolizing *L. crispatus* 125-2-CHN (Student's T-test, $P < 0.05$) (Fig. 28B). We then compared the relative fitness of the *L. crispatus* isolates against *G. vaginalis* by first log-transforming the fold-growth values for both species to correct for stochasticity of the measurements. We then divided the transformed fold-growth of *L. crispatus* by the transformed fold-growth of *G. vaginalis* to calculate the relative growth of *L. crispatus* against their competitor in each co-culture. Fast-metabolizing *L. crispatus* MV-1A-US performed similarly against *G. vaginalis* on both glucose and glycogen (Fig. 28C). Both slow-

metabolizing and non-metabolizing strains, however, performed significantly worse against *G. vaginalis* on glycogen compared to glucose (Student's T-test, $P < 0.05$) (Fig. 28D, 28E).

Figure 28. Relative fitness of fast-, slow-, and non-glycogen-metabolizing *L. crispatus* against other vaginal bacteria.



Isolates of *L. crispatus* were co-cultured with *G. vaginalis* ATCC14018 in PYD-mod media containing either 0.5% glucose or 0.5% bovine liver glycogen. The number of cells of each species in a sample was determined by measuring the concentration of 16S gene copies by species-specific qPCR, then normalizing by the number of 16S gene copies present in the genome of each isolate. Fold-growth was calculated by dividing the density of cells of either *L. crispatus* or *G. vaginalis* present in the culture after 48 hours anaerobic incubation at 37°C by their initial density. Fold-growth of A) *L. crispatus* and B) *G. vaginalis* in co-cultures of *G. vaginalis* with either fast-, slow-, or non-glycogen-metabolizing strains of *L. crispatus*. Lines show standard error of three separate cultures. Fold-growths of the organisms were log (base 10) transformed, then relative growth was calculated for C) fast-, D) slow-, and E) non-glycogen-metabolizing strains of *L. crispatus* against *G. vaginalis* to determine the relative fitness of the lactobacilli against *G. vaginalis* in co-culture. Box plots show the minimum, maximum, and medians of three separate cultures. Stars show significant differences as determined by Student's T-test ($P < 0.05$).

Discussion

Studies published in the 1960's found that a very small proportion of lactobacilli isolated from the human vagina were able to metabolize glycogen[211,212]. Based on these results, a dogma arose that host amylases are primarily responsible for releasing fermentable carbohydrates from vaginal glycogen. Although subsequent studies have shown that some isolates of *L. crispatus* are able to metabolize glycogen directly[214], the false notion that commensal lactobacilli cannot directly utilize glycogen has persisted, appearing in publications as recently as 2018[262]. Despite vaginal glycogen being an abundant source of energy for vaginal bacteria, and likely one of the primary sources of carbon for the lactic acid which inhibits pathogen colonization, relatively little research has been performed on the amylolytic enzymes of the vaginal microbiota. Here, we show that extracellular glycoside hydrolases are present in a wide range of vaginal bacteria. Curiously though, of the multiple species of commensal lactobacilli commonly found in the vagina, these enzymes appear to be unique to *L. crispatus* and *L. iners*. These also happen to be the two species of *Lactobacillus* that most commonly dominate the eubiotic vaginal microbiota[25], so the ability to metabolize glycogen may contribute to their dominance. Although they appear to be non-glycogen-metabolizing species, *L. gasseri* and *L. jensenii* also sometimes dominate the eubiotic vaginal community. In these instances, it is unclear whether rare strains of these species are present that can metabolize glycogen themselves, or whether they rely on the activity of glycoside hydrolases made by other organisms to secure carbohydrates. In support of the latter model, we found that isolates of both *L. gasseri* and *L. jensenii* grew to a relatively high density on glycogen in the presence of exogenous pullulanase.

While analyzing the distribution of *pulA* genes in *L. crispatus* and *L. iners*, an interesting pattern emerged. *pulA* genes were present uniformly across isolates of *L. iners*, but not only did *L. crispatus* strains tend to cluster based on their isolation source, vaginal isolates were also significantly more likely to encode a *pulA* gene than non-vaginal isolates. This pattern suggests

that vaginal *L. crispatus* may be under selective pressure to maintain their *pulA* genes. Our competition experiments with *L. crispatus* isolates support this hypothesis. We found that *L. crispatus* isolates with low glycosidase activity (*L. crispatus* JV-V01) and no detectable glycosidase activity (*L. crispatus* 125-2-CHN) both had a competitive disadvantage against glycogen-metabolizing *G. vaginalis* when they were growing on glycogen, while an isolate with relatively high glycosidase activity (*L. crispatus* MV-1A-US) had similar fitness against *G. vaginalis* whether the bacteria were growing in the presence of glucose or glycogen.

The differences in glycogen metabolism phenotypes between *L. crispatus* isolates were striking. We observed greater than 10-fold differences in the rates of glycogen breakdown between the glycogen-metabolizing *L. crispatus* strains MV-1A-US and JV-V01, which likely contributed to the reduced fitness of the latter isolate against other bacteria. According to the publicly available genomes for these isolates, both MV-1A-US and JV-V01 have pullulanase enzymes with identical amino acid sequences, so these differences are likely a matter of differential protein expression. However, additional experiments are required to determine why *L. crispatus* isolates have such a diversity of glycogen-metabolizing phenotypes. It is also unclear how these differences affect dynamics of the vaginal microbiota. Although 16S rRNA gene sequencing may show that *L. crispatus* constitutes >90% of a community, strains with diverse glycogen metabolism phenotypes could coexist in the microbiota. Cheater/cooperator dynamics may be present in such a community, as strains that do not express a large pullulanase enzyme benefit from the carbohydrates released by those that do. Such dynamics could contribute to community instability, but *L. crispatus*-dominated communities tend to be the most resilient to compositional changes, so these inter-strain dynamics warrant further investigation.

In addition to *L. crispatus* and *L. iners*, many species of BVAB encode secreted enzymes with glycoside hydrolase activity. In fact, a relatively small number BVAB do not possess such enzymes. *Fannyhessea vaginae, Dialister micraerophilus, Eggerthella*-like spp., and

Megasphaera spp. were some of the few BVAB without any such enzymes in their published genomes. Among BVAB with secreted glycoside hydrolases, we observed growth on glycogen from *G. vaginalis, G. piotii, G. leopoldii, G. swidsinskii, M. mulieris, P. uenonis, P. amnii, P. bivia,* and *P. timonensis*. The only organism which encoded a secreted glycoside hydrolase and did not show substantial growth on glycogen was *P. lacrimalis*, although these organisms did not appear to utilize carbohydrates under the culture conditions we tested. These results demonstrate that many BVAB are just as capable of utilizing glycogen for growth as commensal lactobacilli, and for many species, much more so.

Observing the growth of various vaginal bacteria in the presence of different nutrients also revealed an interesting partitioning of preferred nutrient sources. While some organisms such as *Gardnerella* did not begin growing unless carbohydrates were present in their media, others such as *Prevotella* initially grew just as well without added carbohydrates. This suggests that while some BVAB preferentially use carbohydrates as an energy source, others may initially use different nutrients such as peptides or fatty acids. This separation of preferred nutrients could reduce competition between BVAB, promoting the synergistic interactions that are known to occur between them.

A previous study on glycoside hydrolases cloned out of various vaginal bacteria found large differences in the reaction rates of enzymes from different organisms, as well as differences in pH optima[263]. We found a diverse set of glycoside hydrolase enzymes in our genomic analysis, including some with cell surface-associated domains, and others which appeared to be fully secreted. Additionally, some organisms encoded multiple secreted glycoside hydrolases. Therefore, we sought to compare the function of all the enzymes secreted by different organisms, across a range of physiological pHs. We developed an *in vitro* assay where bacterial cultures were scraped from solid media and suspended with a known concentration of glycogen, then the rate of glycogen breakdown for the isolate was quantified. Using this assay, we found 1,000-fold

differences in the rate that individual CFUs of different bacteria degrade glycogen. *Gardnerella* spp. and *L. iners* tended to degrade glycogen relatively fast, while there were large differences in the rates of glycogen breakdown between different *Prevotella* species. BVAB tended to degrade glycogen most efficiently at pH 6.5, with a large drop in breakdown rate at pH 4.3. Notably, acidic pH had no effect on glycogen breakdown by *L. crispatus*, as both glycogen-metabolizing isolates we tested had the same rate of glycogen breakdown across all three pHs. Notably, this assay tests glycogen breakdown rates of enzymes secreted by bacteria while they are growing on solid media, prior to being suspended in MRD. We grew all our isolates on rich NYCIII or Brucella H&K agar plates which both contain relatively high concentrations of glucose. *L. crispatus* downregulates expression of its pullulanase enzymes when glucose is present[214], so other vaginal bacteria may similarly regulate their glycosidases based on glucose availability. However, we still detected glycosidase activity from numerous isolates of vaginal bacteria, indicating these organisms do not have their glycoside hydrolase enzymes under strict regulation, but are likely prone to express them in a range of conditions.

The results of our *in vitro* glycogen breakdown assay have implications for the dynamics of the vaginal microbiota. BV is especially likely to develop following disruptions to vaginal pH, such as those caused by menses or unprotected penile-vaginal sex[264-266]. Such disruptions may increase the activity of BVAB glycoside hydrolases, allowing these organisms to scavenge more carbohydrates from vaginal glycogen, monopolizing this vital nutrient away from commensal lactobacilli. In BV, vaginal pH is typically maintained at 5.0. Even at this more acidic pH, BVAB had a high rate of glycogen breakdown. Thus, glycogen metabolism by BVAB may help prevent lactobacilli from re-establishing themselves in the community by competing with these commensals for energy-rich carbohydrates. Conversely, at low pH, the acid-tolerant enzymes of *L. crispatus* may allow them to metabolize glycogen while the glycoside hydrolases of BVAB are impaired. Notably, the glycoside hydrolases secreted by *L. iners* behaved more similarly to those

made by BVAB, showing reduced efficiency at pH 4.3. This phenotype could also contribute to the reduced stability of communities dominated by *L. iners*.

Acidic pH is one of the primary barriers to pathogen colonization of the vagina. Since vaginal glycogen is likely the main source of carbohydrates which commensal lactobacilli ferment into lactic acid, glycogen metabolism warrants additional study, including host regulation of glycogen synthesis. While studies have shown glycogen deposition in the vaginal epithelium increases with the onset of puberty[267,268], relatively little is known about the regulation of glycogen synthesis in vaginal epithelial cells. Gaining a better understanding of how host processes control glycogen deposition in the vagina could help clinicians maintain an optimal environment for colonization by commensal lactobacilli. Additionally, elucidating the effect of glycogen metabolism on competition between vaginal bacteria could lead to insights into the development of BV. The dominant paradigm on BV holds that commensal lactobacilli primarily consume carbohydrates while BVAB metabolize peptides and sialic acids from host tissue and mucus. In this study, we found that diverse BVAB are also capable of metabolizing glycogen. Additionally, some BVAB such as *F. vaginae* and *Gardnerella* spp. require carbohydrates for robust growth, so competition for this energy-rich resource likely plays into the dynamics between commensal lactobacilli and BVAB. The fact that *L. crispatus* and *L. iners* appear to be the only glycogen-metabolizing vaginal lactobacilli also raises interesting questions about inter-*Lactobacillus* dynamics. Do non-glycogen-metabolizing species of *Lactobacillus* perform some beneficial function, or do they merely compete for carbohydrate resources freed up by glycogen-metabolizing species? Elucidating these interactions could lead to interventions to stabilize eubiotic vaginal communities.

While we were able to identify a wide range of new glycogen-metabolizing vaginal bacteria in this study, we did not perform a comprehensive census of these organisms. *Sneathia vaginalis* encodes a secreted glycoside hydrolase, but we did not test its ability to metabolize or break down glycogen. Similarly, we found amylolytic enzymes in *P. lacrimalis,* but we were not able to confirm

glycogen metabolism in this organism under the culture conditions we tested. Since *P. lacrimalis* did not appear to utilize carbohydrates in our experiments, culturing them in media with a lower concentration of peptides may promote expression of their amylolytic enzymes and metabolism of glycogen. Additionally, while our competition experiments between *L. crispatus* and other vaginal bacteria on glycogen were suggestive, the lack of a genetic system in any of these organisms prevents us from drawing definitive conclusions about the fitness effects of *pulA* genes. Future studies using knock-out mutants of these organisms in more natural, *in vivo* conditions, can more rigorously test how the ability to metabolize glycogen impacts competition between vaginal bacteria.

In this study, we find that glycogen metabolism is widespread among BVAB, and unique among commensal bacteria to *L. crispatus* and *L. iners*. Fitness of *L. crispatus* isolates may also correlate with their ability to efficiently metabolize glycogen. These results have implications for the microbiology of BV, as well as treatment of incident and recurrent BV, in addition to selection of optimal vaginal probiotics.

Chapter 5: Design of a Novel Genetic System for *Gardnerella*

Background

Gardnerella are a nearly ubiquitous genus of vaginal bacteria, present in 70% of women without BV and 97% of women with BV[25,54]. Initially classed under the single species *Gardnerella vaginalis*, this diverse genus has recently been split into the species *G. vaginalis, G. leopoldii, G. piotii,* and *G. swidsinskii*, along with a number of unnamed genomospecies[269]. Due to their prevalence and the diverse functions they perform including protein digestion by proteases, glycoprotein breakdown by sialidases, host cell cytotoxicity by vaginolysin, and glycogen metabolism by pullulanases, *Gardnerella* have long been thought to play a pivotal role in the development of BV. However, the lack of genetic tools to manipulate these organisms have frustrated efforts to perform mechanistic studies into the role of different virulence factors in *Gardnerella* and their contribution to BV.

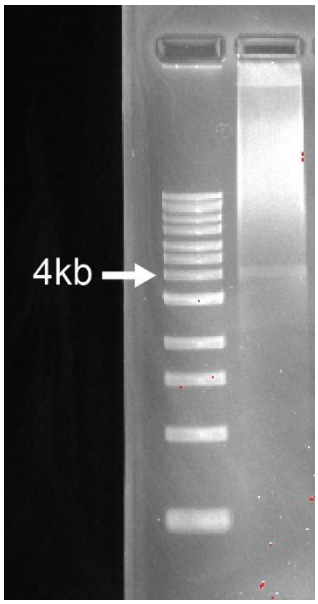
This chapter describes the design and construction of a novel DNA vector to test our hypothesis that a native *Gardnerella* plasmid can serve as the backbone for a vector to express exogenous DNA in these organisms. This work was carried out by Elliot Lee in collaboration with Christopher Johnston, Dakota Jones, Elsa McMahon, Sujatha Srinivasan, Tina Fiedler, Samuel Minot, and David Fredricks. My contributions to this paper included performing experiments described in Figures 29 – 30 and Tables 11 – 12, as well as analyzing the data and preparing written conclusions.

Results

Design of p1199S-*tetM*

While performing single-molecule real-time genome sequencing (SMRTseq)[270] on a collection of *Gardnerella* isolates, we identified a putative native plasmid in *Gardnerella* DNF01199S. We performed a miniprep plasmid isolation on a culture of this strain and purified a ~4kb plasmid that was the same size as the plasmid identified in the SMRTseq data (Fig. 29). We then sequenced the purified plasmid, which we named "p1199S," and confirmed its sequence against the original SMRTseq data.

Figure 29. Plasmid p1199S.



The 4,028bp plasmid p1199S isolated from a *Gardnerella* culture by Miniprep, linearized with the restriction enzyme XhoI, and visualized on a 1% agarose gel. Smearing in lane is indicative of chromosomal DNA from the *Gardnerella* DNF01199S strain.

To adapt this seemingly cryptic plasmid for use as an *E. coli-Gardnerella* sshuttule vector, we sought to identify an antibiotic resistance cassette that could be incorporated and used for selection. *Gardnerella* have high inter-strain variability in their minimum inhibitory concentration (MIC) of tetracycline[271]. We noticed that our sequenced isolates varied in the presence or absence

of a *tetM* gene, which we hypothesized would confer tetracycline resistance to these strains. To test this hypothesis, we grew 25 *Gardnerella* isolates on NYCIII agar plates containing 16µg/mL tetracycline, using *Lactobacillus reuteri* CF48-3A as a tetracycline-resistant control[272,273]. We found a perfect correlation between the presence of a putative *tetM* gene in these *Gardnerella* isolates and ability to grow on this concentration of tetracycline (Table 11).

Table 11. Tetracycline resistance of *Gardnerella* isolates.

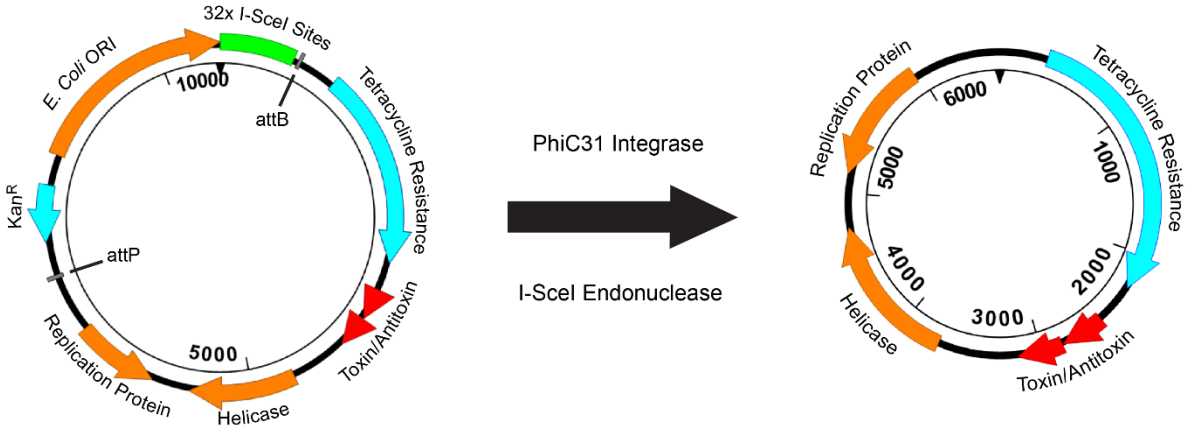
Strain	Species	*tetM*	Tetracycline MIC
ATCC14018	*G. vaginalis*	–	<16µg/mL
DNF00109	*G. vaginalis*	+	>16µg/mL
KA00390	*G. vaginalis*	–	<16µg/mL
DNF01204	*G. vaginalis*	–	<16µg/mL
DNF00038	*G. vaginalis*	–	<16µg/mL
DNF00354	*G. vaginalis*	–	<16µg/mL
DNF00622A	*G. vaginalis*	–	<16µg/mL
DNF01149	*G. vaginalis*	–	<16µg/mL
DNF00571G	*G. piotii*	–	<16µg/mL
DNF00257	*G. piotii*	–	<16µg/mL
DNF01205	*G. leopoldii*	–	<16µg/mL
DNF01195	*G. leopoldii*	+	>16µg/mL
DNF00550	*G. leopoldii*	–	<16µg/mL
DNF01151	*G. swidsinskii*	–	<16µg/mL
DNF01198P	*G. swidsinskii*	+	>16µg/mL
DNF01189	*G. swidsinskii*	–	<16µg/mL
KA00127	*G. swidsinskii*	+	>16µg/mL
DNF01199S	*G.* genomospecies 3	+	>16µg/mL
DNF01144	*G.* genomospecies 3	–	<16µg/mL
DNF00536	*G.* genomospecies 7	–	<16µg/mL
KA00747	*G.* genomospecies 7	+	>16µg/mL
DNF00172	*G.* genomospecies 8	–	<16µg/mL
KA00288	*G.* genomospecies 8	–	<16µg/mL
DNF01141	*G.* genomospecies 8	–	<16µg/mL
KA00255	*G.* genomospecies 13	–	<16µg/mL

tetM genotypes of *Gardnerella* isolates and their phenotypes growing on NYCIII agar plates with 16µg/mL tetracycline. Presence/absence of a *tetM* gene in the isolate is shown by blue/orange shading respectively. An isolate's MIC of tetracycline is denoted by blue (<16µg/mL) or orange (>16µg/mL).

Based on these results, we decided to use *Gardnerella's tetM* gene as a selection marker for transformed bacteria, and cloned this gene out of strain DNF01199S. Next, to simplify generating recombinant DNA to transform into *Gardnerella,* we used Gibson assembly to combine p1199S, *tetM*, and the pMC *E. coli* shuttle vector into a single DNA construct[270]. The pMC backbone allows the parental plasmid to be amplified to high copy number in *E. coli* strain

ZYCY10P3S2T[274], then subsequently excised by addition of arabinose. Arabinose activates expression of φc31 integrase in the *E. coli*, which joins the attP and attB recognition sites in the plasmid, splitting it into separate pMC and p1199S-*tetM* molecules (Fig 30). Arabinose simultaneously activates expression of the I-SceI restriction enzyme, which degrades pMC, leaving p1199S-*tetM* as the only circular DNA molecule in a plasmid prep of these bacteria. Using this system, we were able to purify large quantities of p1199S-*tetM* for transformation experiments.

Figure 30. Induction of the *E. coli-Gardnerella* shuttle vector to synthesize *Gardnerella* plasmid p1199S-*tetM*.



Map of the *E. coli* shuttle vector pMC-p1199S-*tetM* (left) and the excised *Gardnerella* plasmid p1199S-*tetM* (right). Induction with arabinose activates φc31 integrase and I-SceI endonuclease in *E. coli*, which unites the attP and attB sites in the shuttle vector, releasing the *Gardnerella* plasmid and simultaneously degrading the *E. coli* backbone. Nucleotide position is shown on the inner ring of the plasmids. The *E. coli* backbone contains a kanamycin resistance cassette (Kan^R), an *E. coli* origin of replication, and 32 consecutive I-SceI recognition sites. ORFs on the *Gardnerella* plasmid are depicted as arrows. They include a replication initiation protein, helicase, and putative toxin/antitoxin system.

Electrotransformation of *G. vaginalis* ATCC14018

We decided to first target *G. vaginalis* ATCC14018, since it is the type strain of *G. vaginalis,* it is susceptible to tetracycline, and it encodes multiple proteins of interest including vaginolysin, sialidase, and pullulanase. However, a bacterium's DNA restriction/modification (RM)

systems can be a major barrier to introducing exogenous DNA[275]. Based on genome and methylome data of *G. vaginalis* ATCC14018 gathered by SMRTseq and analyzed using the REBASE database[276], we determined this isolate contained one RM system which targets the unmethylated $^{5'}$GGCC$^{3'}$ DNA sequence motif. Therefore, we first tested our recombinant vector by transforming *G. vaginalis* ATCC14018 with 1µg, 5µg, and 10µg p1199S-*tetM* DNA, in addition to 1µg p1199S-*tetM* DNA that had been pre-treated with HaeIII methyltransferase, which methylates DNA at $^{5'}$GGCC$^{3'}$ sites to form $^{5'}$GG^{m5}CC$^{3'}$. Cells were transformed by electroporation and plated onto both NYCIII and NYCIII + 16µg/mL tetracycline agar plates to confirm viability and screen for transformants. All three transformations with unmethylated DNA failed to produce transformants, but cells transformed with HaeIII-treated DNA produced a small number of transformant colonies (Table 12). We were able to confirm successful transformation of these colonies by culturing the bacteria in NYCIII media containing 16µg/mL, re-isolating the p1199S-*tetM* plasmid by Miniprep, and confirming its sequence against our original construct.

Table 12. Conditions and results of electrotransformation experiments with p1199S-*tetM* in *G. vaginalis* ATCC14018.

Quantity DNA (µg)	HaeIII MTase Treated	OD600 of Cells When Competence Induced	Number Transformant Colonies	Transformation Efficiency (Transformants per µg DNA)
1	No	1.3	0	0
5	No	1.3	0	0
10	No	1.3	0	0
1	Yes	1.3	3	15
1	Yes	0.28	3	15
1	Yes	0.38	29	145
1	Yes	0.68	14	70
1	Yes	0.91	18	90
1	Yes	1.00	3	15
1	Yes	1.15	0	0

All transformations were performed by electroporation in 0.1cm cuvettes with the following electrical parameters: 25µF, 200Ω, and 2,000V.

The cells transformed in the above experiment were harvested and made competent when they were in the late log-phase of their growth. We hypothesized that cells in mid log-phase would

be more receptive to plasmid DNA, so we made competent cells from cultures at a range of points across their growth curve and performed an electrotransformation with HaeIII-treated p1199S-*tetM* DNA. In this experiment, cells in early log-phase growth had the highest transformation efficiency, reaching 145 transformant colonies per µg of DNA (Table 12).

Discussion

Genetic manipulation of microorganisms has been a paradigm-shifting technology for studying the biological roles of genes. The lack of such genetic tools to manipulate *Gardnerella* has impaired the study of these organisms, and prevented mechanistic investigations into key genes. In this work, we described the first recombinant vector to introduce exogenous DNA into *Gardnerella*. By conferring tetracycline resistance into the tetracycline-susceptible strain *G. vaginalis* ATCC14018, we demonstrated that this tool self-replicates in the bacteria and can be used to express new genes in these bacteria.

While we succeeded in transforming our p1199S-*tetM* vector into the target bacteria, transformation efficiency remains low. Low efficiencies are acceptable when the primary goal is to create a single transformant capable of expressing an exogenous gene, but higher transformation efficiency is required for more complex manipulations such as homologous recombination or creation of transposon-sequencing libraries[277,278]. Although we increased transformation efficiency roughly 10-fold by harvesting cells at early log-phase growth for competence, additional optimizations of our transformation protocol will be necessary. We may be able to achieve higher efficiency by altering the parameters of electrotransformation, or changing our procedure for making competent cells, i.e. by growing cells with excess NaCl prior to harvesting them to weaken their cell walls[279].

A functional DNA vector in *G. vaginalis* opens many exciting new possibilities to study these organisms. The bacteria could be transformed with a plasmid encoding an anaerobic

fluorescent protein such as Y-FAST[280] to observe multi-species biofilm formation in real time. A similar system could also be used to study expression of different proteins by placing the fluorescent protein under control of the same promoters as different proteins of interest. This system can also be leveraged to create gene knock-out mutants of *Gardnerella* for true mechanistic studies of gene function[281]. Currently, we have only tested this vector in a single strain of *G. vaginalis*. There are substantial phenotypic and genotypic differences between species of *Gardnerella*, i.e. the high-activity sialidases NanH2 and NanH3 that are unique to *G. piotii*, so expanding this vector to other *Gardnerella* species could help elucidate differences in this genus.

Chapter 6: Future Directions

Summary of Work

BV is a complex condition that is likely caused by a confluence of diverse host and bacterial factors. Gaining a better understanding of the etiology of BV and the functions of bacteria in eubiosis and dysbiosis will be critical to improving BV treatment and reducing recurrence. This dissertation describes novel metaproteomic methods to perform untargeted analysis of host and bacterial functions in the vagina, as well as a novel genetic tool to interrogate the function of the nearly ubiquitous BVAB *Gardnerella*. Application of optimized data analysis methods to metaproteomic data from CVL samples uncovered previously unknown functions of the host and specific BVAB which may contribute to BV, in addition to a new syntrophic relationship between *F. vaginae* and *D. micraerophilus* to increase putrescine biosynthesis. Thorough examination of the identified proteins also revealed pullulanase enzymes made by both *L. crispatus* and *Gardnerella*, which led to the discovery that glycogen metabolism is unique to *L. crispatus* and *L. iners* among commensal lactobacilli, but widespread among common species of BVAB. These results have broad implications for dynamics of the vaginal microbiota, as *L. crispatus* isolates with defects in glycogen metabolism have a competitive disadvantage against glycogen-metabolizing *G. vaginalis.* Despite these findings, there are numerous unanswered questions to explore regarding host and bacterial functions in the vagina, and their relationship to BV.

Final Questions

How do host processes contribute to health and BV?

One of the long-standing enigmas of BV is that, despite the various pathogenic functions carious out by BVAB, BV is not associated with classic signs of inflammation such as pain, redness, or leukocyte infiltration. Thus, the host contribution to BV remains unclear. The

metaproteomic data discussed in this dissertation align with past studies that BV is associated with dysregulation of host proteases, particularly leukocyte elastase inhibitor. Uncontrolled activity of elastase may contribute to tissue damage, and release peptides which can be metabolized by BVAB. However, few studies address regulation of host antiproteases in the vagina, so it is unknown how signaling by the host or BVAB may lead to decreased expression of antiproteases. Elucidation of these mechanisms could lead to supplemental therapies which make the vaginal lumen less proteolytic, and therefore less suitable for BVAB. Our data also uncovered novel differences in various heme-detoxifying enzymes. These may represent a host response to heme released by red blood cells lysed by BVAB toxins. Further investigations into the role of heme as a source of iron for BVAB, and host responses to clear this compound, could lead to interventions to prevent BVAB from accessing this vital nutrient.

What is the role of BVAB-specific enzymes in the BV community?

Our metaproteomic analysis of CVL samples uncovered a number of enzymes that were highly expressed by multiple BVAB, but apparently absent from the genomes of commensal lactobacilli. Among these were pyruvate formate lyase and glutamate dehydrogenase. Although there are no published studies investigating formate in relation to bacterial vaginosis, pyruvate formate lyase was the most commonly identified bacterial fermentation enzyme in our data. Therefore, this enzyme likely performs an important metabolic function for BVAB, but whether it is catabolic, anabolic, or facilitates cross-feeding of carbon between organisms remains unclear. Additionally, we observed glutamate dehydrogenase homologs from diverse BVAB, with the notable exception of *Gardnerella*. While diverse *Gardnerella* functions have been hypothesized or shown to benefit other BVAB, with the exception of ammonia cross-feeding from *Prevotella,* few cross-feeding interactions from other BVAB are known to benefit *Gardnerella*. Thus, the potential for other BVAB to benefit *Gardnerella* by secreting glutamate represents a potential target for disrupting synergistic interactions in the BV microbiota.

What role do polyamines play in BV?

Malodor is a characteristic symptom of BV, and the most common reason patients seek treatment. While it is not definitive evidence, the fact we only identified peptide spectra matching putrescine-producing ornithine decarboxylases from *D. micraerophilus* suggests these organisms could be one of the main bacterial sources of this foul-smelling chemical among BVAB. It also appears that *F. vaginae* and *D. micraerophilus* can work together to increase biosynthesis of putrescine. While polyamines like putrescine are a feature of other anaerobic infections such as diabetic wounds, it is unclear what biological function they serve for bacteria. Our understanding of BV would be furthered by determining whether bacteria benefit from synthesizing these compounds by buffering the pH of their environment, facilitating nitrogen cross-feeding, modulating the host immune response, or some other function.

How does inter-strain diversity of *L. crispatus* influence the vaginal microbiota?

One of the most surprising results described in this dissertation was the clear delineation between *L. crispatus* isolated from the human vagina, and those isolated from other animals and/or body sites. A vaginal community dominated by *L. crispatus* is the least likely to shift to BV, so this species has long been considered the optimal commensal species of *Lactobacillus*. Different isolates of *L. crispatus* displayed a wide array of glycogen-metabolizing phenotypes, raising the possibility that there are other characteristics for which *L. crispatus* contain a large degree of heterogeneity. Such a situation would complicate the homogenous view of an *L. crispatus*-dominated community, as diverse strains of this species could co-exist, performing complementary functions in a similar manner to BVAB. While currently a remote possibility, this model could in part explain the failure of probiotic therapies using mono-cultures of *L. crispatus* to reduce the incidence of recurrent BV. Further investigations into the inter-strain diversity of *L.*

crispatus could therefore lead to probiotics containing a complementary consortia of *L. crispatus* isolates, which may prove more resilient than individual strains alone.

How does competition for vaginal glycogen influence the composition of the bacterial community?

In order to grow, heterotrophic bacteria must harness energy from the breakdown of chemicals. Many organisms, including the majority of vaginal bacteria, use carbohydrates as a major source of energy. Carbohydrate fermentation is especially important for the composition of the vaginal community, since the end-products of fermentation, whether weak or strong acids, influence what species can survive in the vagina. Thus, our discovery that many species of vaginal bacteria can directly metabolize glycogen has significant implications for competition between commensal lactobacilli and BVAB. While our genomic analysis and competition experiments suggest glycogen-degrading pullulanases confer a competitive advantage to *L. crispatus* when glycogen is the primary species of carbohydrate present in the environment, additional studies are necessary to fully understand how glycogen metabolism influences competition between vaginal bacteria. How does glycogen metabolism affect competition between glycogen-metabolizing and non-glycogen-metabolizing species of *Lactobacillus*? How does it affect competition between the two glycogen-metabolizing species *L. crispatus* and *L. iners*? And what factors determine whether commensal lactobacilli or BVAB are the primary recipients of carbohydrates released from vaginal glycogen by glycoside hydrolase enzymes? Answering these questions may lead to a better understanding of how the vaginal microbiota shifts from eubiosis to BV.

Chapter 7: Materials and Methods

Study population and sample collection

The parent study enrolled 242 women from the Public Health, Seattle and King County Sexually Transmitted Diseases Clinic (STD clinic) between September 2006 and June 2010[25]. The study was approved by the Institutional Review Board at the Fred Hutchinson Cancer Research Center and all study participants provided written informed consent. Vaginal fluid samples were collected for molecular characterization, Gram-staining, pH, saline microscopy and potassium hydroxide preparation. BV was diagnosed using Amsel clinical criteria and confirmed by Gran-stain using the Nugent method. BV diagnosis was the same by both Nugent score and Amsel criteria for all 29 participants whose samples were analyzed by mass spectrometry. Cervicovaginal lavage (CVL) was collected by instilling 10 mL sterile saline into the vagina using a needleless syringe, and walls of the vagina were washed to remove any adherent cells. After ~1 minute, the lavage fluid was aspirated and stored at -80oC. CVL was used for proteomic analyses. A random cross-sectional case-control set of 20 samples from women with BV and 10 samples from women without BV were selected for proteomics. As there is greater bacterial heterogeneity in BV and we used a 2:1 case to control ratio for sample selection to have better representation of women with BV. One sample could not be analyzed leaving 29.

Swab preparation and DNA extraction from vaginal swab samples

Vaginal swabs collected from study participants were frozen at -80°C until processing. Swabs were prepared for extraction by placing swab tip in a tube with 500 µL filtered 0.9% saline (100K MWCO). Swabs were vortexed 1-2 min and then removed. Vaginal fluid from the swab tip was centrifuged at 14,000 rpm for 10 min at 4°C to pellet cells. Genomic DNA was extracted from pellets using the BiOstic Bacteremia DNA Isolation Kit (Qiagen, Germantown, MD, USA). DNA was eluted in 150 µL buffer. DNA extraction controls (blank swab without contact with human

mucosa) were included for every 15 samples to assess contamination from extraction reagents or collection swabs. DNA was stored at -80°C until sequencing.

Broad-range PCR and sequencing for microbiota characterization

Total bacterial DNA concentrations (16S rRNA gene copies) were measured using a qPCR assay targeting the V3-V4 region of the 16S rRNA gene[25]. Samples were evaluated for the presence of PCR inhibitors using a qPCR assay targeting a segment of exogenously added jellyfish DNA and inhibition was defined as a delay in the threshold of >2 cycles compared to no-template controls[282]. Relative abundances of bacterial taxa sequence reads were measured using broad-range PCR targeting the V3-V4 region of the 16S rRNA gene and sequencing on the Illumina MiSeq instrument (Illumina San Diego, CA)[283]. Raw sequence reads were demultiplexed using the Illumina MiSeq's onboard software. Demultiplexed reads were processed using *barcodcop* v0.4.1 (Hoffman NG. barcodecop. 2019. https://github.com/nhoffman/barcodecop) to enforce barcode quality using default setting and to ensure exact barcode matches to the forward and reverse reads. The *DADA2* package version 1.6.0 was used for quality filtering, read trimming, error correction and dereplication, paired-end assembly and chimera removal resulting in a list of unique sequence variants[284]. Sequence variants were classified using the phylogenetic placement tool *pplacer*[285] and a curated set of vaginal bacteria[25]. Sequence reads are available from the NCBI Short Read Archive (Accession number PRJNA881379).

Metagenomics library preparation and sequencing

Genomic DNA (gDNA) from vaginal samples was quantified via Qubit® fluorometer and Quant-iT™ dsDNA Assay Kit, high sensitivity (Life Technologies-Invitrogen, Carlsbad, CA, USA). Sequencing controls included a bacterial mock community (ATCC MSA-1003, Manassas, VA, USA), and a bacterial isolate *Fannyhessea vaginae* DNF00720. Sequencing libraries were prepared from 250pg gDNA with a quarter reaction workflow using the Nextera XT Library Prep

Kit (Illumina, San Diego, CA, USA). Libraries were pooled by volume and post-amplification cleanup was performed with 0.8X Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA). The library pool size distribution was validated using the Agilent High Sensitivity D5000 ScreenTape run on an Agilent 4200 TapeStation (Agilent Technologies, Inc., Santa Clara, CA, USA). Additional library QC and cluster optimization was performed using Life Technologies-Invitrogen Qubit® 2.0 Fluorometer (Life Technologies-Invitrogen, Carlsbad, CA, USA). Sequencing was performed on a NovaSeq 6000 S1-300 flowcell (Illumina, San Diego, CA, USA). Image analysis and base calling were performed on board the NovaSeq 6000 with Real Time Analysis v3.4.4 software. Generation of Fastq files was performed with Illumina's bcl2fastq 2.20 Conversion software. The *geneshot* workflow was used to process Fastq files[286]. This workflow fed data into metaSPAdes for assembly and Prokka for annotation of bacterial genes[287,288].

Proteomic sample preparation

CVL samples were reduced, denatured, and digested to peptides for mass spectrometry analysis. First, 105 mg of solid urea was added to 250 µL of CVL to yield a final concentration of 7 M. Dithiothreitol was then added to a final concentration of 5 mM and the sample was incubated for 30 minutes at 60°C with gentle shaking (300 rpm) in a thermomixer. Following incubation, 2.25 mL of 50 mM ammonium bicarbonate was added to each tube. USB brand trypsin was resuspended to 1 µg/µL in acetic acid, and 5 µL was added to each sample. Samples were incubated at 37°C with gentle shaking (300 rpm) overnight. Digested peptide samples were centrifuged for 5 minutes at 12,000 x g to pellet any remaining solid debris and the supernatant was subjected to solid phase extraction (SPE). SPE cartridges (Phenomenex; Strata C18-T (55 µm, 140 Å), 100 mg / 1 mL, cat # 8B-S004-EAK) were loaded into the vacuum manifold and samples were processed according to the manufacturer's recommendation. Briefly, cartridges were conditioned with 1 mL methanol and washed with 1 mL 0.1% trifluoroacetic acid (TFA) in water. The sample was added to the cartridge, followed by a wash with 1 mL 5% acetonitrile/95%

0.1% TFA in water. Finally, the sample was eluted with 1 mL 80% acetonitrile and 20% 0.1% TFA in water. Samples were concentrated to near dryness using a SpeedVac and resuspended in 30 µL 0.1% TFA water. A BCA assay (Pierce) was performed to determine peptide concentration. Samples were diluted to 0.1 µg/µL and stored at -80°C until MS analysis.

LC-MS/MS analysis

A Waters nano-Acquity dual pumping UPLC system (Milford, MA) was configured for on-line trapping of a 5 µL injection at 5 µL/min with reverse-flow elution onto the analytical column at 300 nL/min. Columns were packed in-house using 360 µm o.d. fused silica (Polymicro Technologies Inc., Phoenix, AZ) with 2-mm sol-gel frits for media retention and contained Jupiter C18 media (Phenomenex, Torrence, CA) in 5 µm particle size for the trapping column (150 µm i.d. x 4 cm long), with 3 µm particle size for the analytical column (75 µm i.d. x 70 cm long). Mobile phases consisted of (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile with the following gradient profile (min, %B): 0, 1; 2, 8; 20, 12; 75, 30; 97, 45; 100, 95; 110, 95; 115, 1; 150, 1. MS analysis was performed using a Velos Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) outfitted with a custom electrospray ionization (ESI) interface. Electrospray emitters were custom made by chemically etching 150 µm o.d. × 20 µm i.d. fused silica[289]. The heated capillary temperature and spray voltage were 350°C and 2.3 kV, respectively. Data was acquired for 100 minutes after a 15-minute delay from when the gradient started. Orbitrap spectra (AGC 1 × 106) were collected from 400 to 2000 m/z at a resolution of 60 k followed by data-dependent ion trap MS/MS (collision energy 35%, AGC 1 × 10^4) of the 10 most abundant ions. A dynamic exclusion time of 45 seconds was used to discriminate against previously analyzed ions using a − 0.55- to 1.55-Da mass window. Each sample was analyzed in two separate replicates.

Database construction for identification of peptide sequences

Seven different protein sequence databases were built for identification of peptides. All databases included human protein sequences from SwissProt (release 2019_02) and 16 common contaminants including human keratins that could be introduced during sample processing and trypsins that could be left over from sample preparation. The Global database consisted of all bacterial, fungal, and *Trichomonas vaginalis* sequences available on NCBI RefSeq (RefSeq release 97), downloaded June 19th, 2020. A separate 16S_Sample-Matched database was built for each sample and included all RefSeq sequences for bacterial taxa present in the sample at >0.1% abundance according to 16S rRNA gene sequencing, except BVAB2 for which there were no available genomes at time of publication. The 16S_Pooled database was constructed using all bacterial protein sequences used to build the 16S_Sample-Matched databases, as well as all RefSeq protein sequences for *T. vaginalis, Chlamydia trachomatis, Neisseria gonorrhoeae*, and the common vaginal fungi *Alternia alternata, Candida albicans, Candida glabrata, Candida tropicalis, Pichia kudravzevii*, and *Saccharomyces cerevisiae* (RefSeq release 99) downloaded June 19th, 2020. For the 16S_Pooled and 16S_Sample-Matched databases, sequences for individual microbes were downloaded from RefSeq. The 16S_Reference databases also only included RefSeq proteins for bacterial taxa present in the sample at >0.1% abundance, but only included proteins translated from the genome of the reference strain for that species, as listed on the NCBI website[137]. After shotgun metagenomic sequencing was performed on each sample, a sample-matched Shotgun_Sample-Matched database was built using only the translated bacterial proteins identified in that sample. A single Shotgun_Pooled database was also built by pooling together the translated bacterial proteins from all 29 samples. Sample-matched Hybrid_Sample-Matched databases were built using all the proteins from a sample's 16S_Sample-Matched and Shotgun_Sample-Matched databases. All nonhuman FASTA sequences were normalized before being incorporated into databases with the

SpeciesSeqPrepper.py program. *Gardnerella* species were delineated using the same groups described in Vaneechoutte, et al. (2019)[269]. Recently published *Gardnerella* genomes were submitted to the DSMZ genome-to-genome distance calculator[290] and a cutoff of 70% similarity was used to determine new species.

Database construction for comparison of additional strains versus additional species

A baseline database was built using human SwissProt sequences (release 2019_02) and 16 common contaminants, as well as protein sequences from the strains of 10 common vaginal bacteria that provided the most PSMs in a search of the 16S_Sample-Matched databases. These strains were *Gardnerella swidsinskii* GS10234, *Gardnerella leopoldii* UMB0912, *Gardnerella vaginalis* UMB0411, *Gardnerella vaginalis* DNF01149, *Gardnerella piotii* UGhent 18.01*, Lactobacillus iners* SPIN 1401G, *Megasphaera lornae, Prevotella timonensis* DSM 22865, Candidatus *Lachnocurva*, and *L. crispatus* JV-V01. Databases were built by adding protein sequences from 20, 40, 60, 80, or 100 randomly chosen genomes. For the Additional Strains databases, these genomes were chosen from 103 *L. crispatus genomes* available in RefSeq release 97. For the Additional Species databases, these genomes were chosen from 103 random bacterial species of different genera assembled to the scaffold level as part of the Human Microbiome Project (RefSeq release 99). Three separate databases were built and tested for each size.

Database searching

Approximately half of bacterial peptides identified in searches of the first replicates were not identified in the second replicates and vice-versa, so both replicates for each sample were combined for further proteomic analysis. Peptide identification was performed with MS-GF+ (v2019.01.22). Isotope error range was -1 to 2, maximum modifications per peptide was set to 3,

and peptides were only considered if they were at least partially tryptic. The Nextflow workflow manager (v19.07.0) was used to parallelize and automate the data analysis pipeline.

A two-step database search method was used to maximize data from each sample[160]. Briefly, after an initial database search was completed, every protein matched as part of that peptide search was recorded, regardless of statistical significance. A subset protein database was then constructed using only the sequences of the proteins identified in the initial search. A second search was performed using these subset databases, and the results of this search were used for downstream analysis. A decoy database was created and searched concurrently to calculate false discovery rate (FDR) and *q* values.

Sample selection for comparison of Global against other database types.

Because the Global database was 100x larger and more expensive to search against compared to the next largest database (16S_Pooled), a subset of samples was selected for comparison of the Global, 16S_Pooled, 16S_Sample-Matched, Shotgun_Pooled, Shotgun_Sample-Matched, and Hybrid_Sample-Matched databases. In our dataset, 31% of samples were BV-, so 2 BV- samples and 4 BV+ samples were chosen at random from the groups. Spectra in these samples were then searched against all seven database types as described above and the results were used to compare the database types.

Determination of cost and computing requirements for searching databases

The Nextflow computational pipeline was run with the “-with-report” flag so an .html report would be generated for each execution. The number of CPU hours required to run each search was then totaled from this document, and cost to run each search was calculated based on Amazon Web Services published cost per hour for a c6g.4xlarge spot instance (as of March 2022).

Proteomic data analysis

Search results were analyzed in JupyterLab v1.1.4 (v1.1.4) using Python (v3.6.4). False discovery rate was limited by only including peptide-spectrum matches (PSMs) with a Q-value <0.01. Spectra were classified in a hierarchical manner. If the spectra matched any decoy proteins, it was flagged as a decoy identification and discarded. If the spectra matched any contaminant proteins, it was flagged as a contaminant identification and similarly discarded. Otherwise, if the spectrum matched any eukaryotic proteins, it was categorized preferentially as human, then fungal, then *Trichomonas* if it only matched *Trichonomas* proteins. Finally, the spectrum was characterized as bacterial if it only matched bacterial proteins. Peptides were included in further analysis if they met one of the following conditions: A) MS-GF+ assigned the PSM a spectral probability value <1E-15. B) The peptide was one of two unique peptides that matched to the same protein. To calculate the relative abundance of certain human proteins, the total number of PSMs attributed to a protein was divided by the total number of human PSMs. These data were then log transformed (base 2) for statistical analysis. Statistical tests were performed using the scipy (v1.3.1) “stats” package. Mann-Whitney U tests were used to assess differences in relative abundance of specific proteins across samples.

Comparison of database performance – PSMs generated

For each sample, the number of significant PSMs was determined for both human and bacterial spectra as described above. Comparisons were made between databases using the scipy (v1.3.1) “stats” package to perform Wilcoxon Signed-Rank tests on each combination of databases. A database was considered to generate significantly more human or bacterial PSMs than the other if the test reported $p < 0.01$. Relative identification rates for human or bacterial PSMs was determined by first calculating the average number of significant PSMs of the given type identified in the sample when it was searched against the 16S_Pooled, 16S_Sample-

Matched, Shotgun_Pooled, Shotgun_Sample-Matched, and Hybrid_Sample-Matched databases. The number of significant PSMs found in that sample by a single database search was then divided by the average significant PSMs for the sample across all databases. This proportion was the relative identification rate.

Functional analysis

Functional annotations for human and bacterial proteins were separately gathered by querying identified protein sequences against the eggNOG-Mapper web server (v1.0.3)[135,136]. For PSMs that matched multiple proteins, only the first protein match was queried. Each GO number as assigned by eggNOG-Mapper was then given a spectral count in each sample by totaling the number of PSMs associated with its protein(s). The spectral count of each functional annotation was then divided by the total number of human or bacterial PSMs in the sample, then $\log_2$ transformed and tested for statistical significance as described above.

Differential protein abundance analysis.

To calculate the relative abundance of human proteins, the total number of PSMs attributed to a protein in a sample was divided by the total number of human PSMs in that sample. Due to their heterogeneity, bacterial proteins were grouped functionally by annotating them with the eggNOG-Mapper web server (v1.0.3)[135,136]. For bacterial PSMs that matched multiple proteins, only the first protein match was queried. Proteins were then grouped based on the “Preferred_name” field assigned to each by eggNOG-Mapper, and relative abundance of each group in a sample determined by dividing the number of PSMs matching that group by the total number of bacterial PSMs in that sample. Protein relative abundances were then log transformed (base 2) for statistical analysis. Statistical tests were performed using the scipy (v1.3.1) “stats” package. Mann-Whitney U tests were used to assess differences in relative abundance of specific proteins across samples.

Taxonomic analysis of proteomic data

Taxonomic assignment of peptides was performed using the taxonomic information attached to protein sequences in each database. All potential protein hits in the database were identified and all species encoding those proteins were noted.

Quantification of unique proteins

The number of unique human and bacterial proteins identified by each database search was determined by first applying the filtering criteria described above to identify the valid peptides identified across all 29 samples. Each non-redundant protein in the database was then iterated through and counted as identified if it met one of the following conditions: A) More than one non-identical peptide had been identified across all samples that matched the protein. B) A peptide matching the protein had been identified more than once across all samples. C) A peptide matching the protein had been identified with spectral probability value <1E-15. After a protein had been counted as identified, all peptides matching it were removed from the pool of peptides being considered so other proteins matching these peptides would not also be counted.

Calculation of weighted average bacterial peptides in 16S_Sample-Matched databases and correlation with relative performance of 16S_Sample-Matched and Hybrid_Sample-Matched databases.

A protein sequence database is expected to generate more PSMs if it has more protein sequence data for the species present in a sample, especially those at high relative abundance which are likely to contribute a larger share of peptides to the proteome detectable by mass spectrometry. Therefore, to determine completeness of a database populated with publicly available protein sequences, a weighted mean of the number of tryptic bacterial peptides available for a search program to compare sample mass spectra against was calculated. An *in silico* tryptic digest of publicly available bacterial protein sequences was performed and the number of unique tryptic

peptides longer than 5 amino acids summed[291]. Then, the following equation was used to calculate the weighted mean number of tryptic peptides for bacteria in each 16S_Sample-Matched database, where P_i is the number of tryptic peptides available for species *i* and A_i is the relative abundance of species *i* in the sample:

$$\sum_{i=1}^{n} \ln(1 + G_i) \times A_i$$

The relative performance of the 16S_Sample-Matched and Hybrid_Sample-Matched databases was then calculated for each sample by dividing the number of significant bacterial PSMs identified by the 16S_Sample-Matched database by the number of significant bacterial PSMs identified by the Hybrid_Sample-Matched database for that sample. The scipy (v1.3.1) "stats" package was then used to determine the Spearman's Rank-Order correlation between these values. Code in the "Taxa Abundance Correlation.ipynb" notebook was used to determine how the number of PSMs identified for different taxa changed between searches of 16S_Sample-Matched and Hybrid_Sample-Matched databases.

Software used for initial database construction

Protein sequences from individual microbial species were downloaded from RefSeq using the DownloadFromNCBIFTPtxt.py program. Prior to database assembly, all nonhuman protein sequences were normalized with SpeciesSeqPrepper.py, then sequences from separate strains were combined into a single file with StrainCombiner.py. When a separate database tailored to each individual sample was required, a CSV file indicating which species should go into each sample's database was constructed, and code in the "Tailored DB Building.ipynb" notebook was used to assemble each 16S_Sample-Matched database. For the database, LargeDatabasePrepper.py was used to generate searchable database files approximately 300MB in size. Code in the "Community DB Building.ipynb" notebook was used to build the 16S_Pooled

database. Shotgun metagenomic sequencing databases were built by taking protein data from translated bacterial open reading frames and assembling sample-matched databases for Shotgun_Sample-Matched, or pooling together all translated bacterial protein sequences for Shotgun_Pooled. Software used to build these databases is in the “Metagenomic DB Building.ipynb” Jupyter Lab notebook file. Hybrid_Sample-Matched databases were built by adding new proteins from a sample’s Shotgun_Sample-Matched database to all the sequences already in its 16S_Sample-Matched database and collapsing together any proteins with identical amino acid sequences. Software used to build these databases is in the “Metagenomic Hybrid_Sample-Matched DB Building.ipynb” Jupyter Lab notebook file. 16S_Reference databases were built by identifying the reference strain for each species used to build the 16S_Sample-Matched databases. For most strains, the reference strain listed on the NCBI website was used. The only species without a listed NCBI reference strain were those for which only a single genome was available, and in these cases, the one publicly available genome was used as the reference strain. The reference strains were then placed in a CSV file, and taxa where the only available genome was to be used as the reference strain, the reference strain in the CSV file was listed as “A.” Software used to build the 16S_Reference databases is in the “Reference Genome DB Building.ipynb” Jupyter Lab notebook file.

Software used for two-step database searches

MSGF_MultipleDBs.nf and MSGF_TailoredDBs.nf were used to stage files to Amazon Web Services (AWS) S3, queue the searches to run on AWS EC2, and convert the search results to TSV format. For searches using a separate database for each sample, the runTailoredDB.sh script was used to initiate a search, while the runMultiDB.sh script was used when each sample was searched against a single database. Software used to create the subset 16S_Sample-Matched databases is located in the “Tailored DB Building.ipynb” Jupyter Lab notebook file. Software used to create the subset databases for Shotgun_Pooled, Shotgun_Sample-Matched,

and Hybrid_Sample-Matched databases are in their respective Jupyter Lab notebook files, listed above. For the larger Global, 16S_Pooled, and Shotgun_Pooled, multiple elliot_utils.py functions were used to generate the subset databases. First, the “condenseHugeDBResults” function was used to condense the results of each sample search into a single file by comparing hits on a single spectrum and keeping the one with the highest MSGF score. “getHitsInResults” was used to identify all proteins hit in every sample regardless of significance, and “refineHugeDatabase” was used to build the subset database with those protein sequences. Results of the 16S_Pooled, and Shotgun_Pooled databases were condensed into a single file with “condenseHugeDBResults” before analysis.

Contaminant protein sequences

Proteins such as keratins and trypsins are commonly introduced into samples by processing procedures and from the researchers themselves. Therefore, the following contaminant proteins were included in all metaproteomic databases as a minimal set of peptides that should be excluded from downstream analysis while simultaneously attempting to preserve true identifications. These proteins were: *Sus scrofa* trypsin precursor (sp|P00761|TRYP_PIG), Promega trypsin artifact 1 (Trypa1), Promega trypsin artifact 2 (Trypa2), Promega trypsin artifact 3 (Trypa3), Promega trypsin artifact 4 (Trypa4), Promega trypsin artifact 5 (Trypa5), Trypsin artifact 6 (Trypa6), *Bos taurus* trypsinogen (sp|P00760|TRYP_BOVIN), *Bos taurus* chymotrypsinogen A (CTRA_BOVIN), *Bos taurus* chymotrisinogen B (CTRB_BOVIN), *Homo sapiens* serum albumin precursor (sp|P02768|ALBU_HUMAN), *Bos taurus* serum albumin precursor (sp|P02769|ALBU_BOVIN), *Homo sapiens* keratin type II cytoskeletal 1 (K2C1_HUMAN), *Homo sapiens* keratin type II cytoskeletal 2 (K22E_HUMAN), *Homo sapiens* keratin type I cytoskeletal 9 (K1C9_HUMAN), and *Homo sapiens* keratin type I cytoskeletal 10 (K1C10_HUMAN).

Calculation of diversity statistics from DNA sequencing data

Diversity statistics were calculated for sequencing data generated by 16S rRNA gene sequencing and shotgun metagenomic sequencing using Phyloseq[292]. Code used to perform these calculations is in the "bvr01_proteomics_rarefaction.R" file. Reads from 16S rRNA gene sequencing and reads identified as bacterial DNA by MetaPhlAn2[293] were used as the raw data. Alpha diversity was calculated by Shannon diversity and Beta diversity between samples was visualized on multidimensional scaling plots using Bray-Curtis distance. Rarefaction curves for each sample and DNA sequencing method were generated using the "rarecurve" function in Phyloseq.

Quantification of polyamine concentrations in bacterial culture supernatants

The bacterial isolates used in this experiment were *Dialister micraerophilus* DSM19965, *Dialister micraerophilus* DNF00843, *Fannyhessea vaginae* DSM15829, *Peptoniphilus lacrimalis* DNF00528, and *Gardnerella vaginalis* ATCC14018. The bacteria were first grown in pure culture by recovering them from frozen stocks on Brucella H&K agar plates (Hardy Diagnostics, Santa Maria CA), followed by two subcultures into 2mL Brucella H&K media. At each step, bacteria were grown in an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C for 72hrs and purity of the cultures was confirmed between each subculture by Gram stain. These pure cultures of actively growing bacteria were then used to inoculate the experimental cultures. For mono-cultures, 200µL of bacterial culture was inoculated into 1.8mL Brucella H&K media. For co-cultures, 100µL of each bacterial isolate was inoculated into 1.8mL Brucella H&K media. A mono-culture of each isolate was grown, and a co-culture was set up so both *D. micraerophilus* isolates grew in co-culture with the *F. vaginae, P. lacrimalis*, and *G. vaginalis* isolates. Three replicates of each culture was made in separate tubes. Experimental cultures were grown anaerobically at 37°C for 72hrs. Following this incubation,

cultures were briefly vortexed to break up biofilms that had formed on the bottom of culture tubes, then 1mL of each culture was removed and centrifuged at 10,000x*g* for 10 minutes to pellet cells. The supernatant was removed, and stored at -20°C until polyamine quantification was performed.

Polyamine quantification was performed with a Fluorometric Total Polyamine Assay Kit (Sigma-Aldrich, St. Louis, MO). To clean up samples prior to analysis, culture supernatants were thawed and 4µL Sample Clean-Up Mix was added to 200µL of supernatant. The mixture was incubated at RT for 30 minutes, then transferred to 10kDa Cut-Off Corning Spin-X UF Concentrators (Corning Inc., Corning, NY). Cleaned-up samples were diluted 1:100, 1:500, and 1:1000 in ultrapure distilled water (Invitrogen, Waltham, MA), then these diluted samples were analyzed according to kit specifications. Fluorescence of samples was read on a Biotek FLx800 fluorescence microplate reader (Agilent, Santa Clara, CA). Differences in polyamine concentration between samples was calculated by performing Welch's T-test to compare. Differences were considered significant if $P < 0.05$.

Targeted polyamine analysis by LC-MS

In addition to the *D. micraerophilus* cultures described above, additional cultures with exogenous ornithine were grown. L-ornithine (Sigma-Aldrich, St. Louis, MO) was dissolved in water to a concentration of 500µg/mL, then filter-sterilized by passing it through a 0.2µm sterile nylon filter (Thermo Fisher Scientific, Waltham, MA). 100µL of pure culture of *D. micraerophilus* DSM19965 or *D. micraerophilus* DNF00843, and 200µL of 500µg/mL L-ornithine were then added to 1.7mL Brucella H&K liquid media, bringing the final concentration of L-ornithine in the culture to 50µg/mL. Three replicates of these cultures for both *D. micraerophilus* isolates were set up in separate tubes. The bacteria were then cultured anaerobically at 37°C for 72hrs, before culture supernatants were collected as described above.

Prior to preparation for LC-MS/MS analysis, culture supernatants were thawed, 50µL of each were pooled, briefly vortexed to mix, then stored at -20°C until sample preparation was performed.

Four basic polyamines (putrescine, cadaverine, spermidine, and spermine) for targeted LC-MS analysis were extracted using a protein precipitation method. Samples were first homogenized in 200µL purified deionized water at 4°C, and then 800µL of methanol were added. Samples were then vortexed, stored for 30min at -20°C, sonicated in an ice bath for 10 minutes, centrifuged for 15min at 14,000RPM and 4°C, and then 600µL of supernatant was collected from each sample. Lastly, recovered supernatants were dried on a SpeedVac and reconstituted in 0.5mL of LC-matching solvent for LC-MS acquisition. Liquid chromatography was performed on X-Bridge Amide, 2.1 x 150mm, 2.5µm columns (Waters, Milford, MA) with Nexera LC-20AD XR pumps (Shimadzu, Kyoto, Japan) and a PAL HTC-XT auto-sampler (CTC Analytics, Zwingen, Switzerland) set at 4°C. Separation mode was set to HILIC. Solvent A was 10mM ammonium acetate in 95% water, 2% MeOH, 3% acetonitrile, 0.2% acetic acid. Solvent B was 10mM ammonium acetate in 5% water, 2% MeOH, 93% acetonitrile, 0.2% acetic acid. MS was performed on an API 6500+ QQQ-MS (SCIEX, Toronto, Canada) with ionization mode set to positive and data was acquired using AB Sciex Analyst (v1.7.2).

Screening bacterial genomes for secreted glycoside hydrolases

Publicly available bacterial genomes were submitted to the dbCAN3 server[257] to identify genes with GH13 family domains. Amino acid sequences of candidate genes were subsequently submitted to the Phobius server[258] to predict signal peptides in the proteins. Identified glycoside hydrolases with signal peptides were then submitted to the HMMER[259] web server for domain visualization. When this process failed to identify any secreted glycoside hydrolase genes in a species, a broader screen was performed using a species-specific BLASTp search on the NCBI BLAST server[294], with a secreted glycoside hydrolase from a related organism as a query

sequence. For lactobacilli, the *L. crispatus* pullulanase QLK32088.1 was used as a query sequence. For BVAB, the *G. vaginalis* pullulanase BAQ33673.1 was used as a query sequence.

Phylogenetic analysis of *pulA* genes in lactobacilli

The genes in genomes map (gig-map) workflow (S Minot, 2023, https://github.com/FredHutch/gig-map) was used to screen genomes of *L. iners* and *L. crispatus* for *pulA* enzymes. *pulA* genes were first deduplicated, leaving the proteins WP_006735465.1, MCT7826185.1, WP_240403716.1, and WP_240404214.1 as query sequences for *L. iners* and QYA52828.1, QYA51816.1, QYA52413.1, QYA52414.1, QLK32088.1, and UAY50421.1 as query sequences for *L. crispatus*. Genomes for *L. iners* (GenBank Release 255) and *L. crispatus* (GenBank Release 253) were downloaded, then query genes were aligned to the genomes and visualized with gig-map. Isolation source for strains of these bacteria was determined according to the metadata associated with each genome's BioSample page. Difference in the proportion of vaginal vs. non-vaginal *L. crispatus* isolates encoding a *pulA* gene was determined by one-proportion Z test.

Screening vaginal bacteria for glycogen metabolism

Initial growth of bacterial isolates was performed in an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C for 48hrs at each subculture. Bacteria were recovered from frozen stocks on agar plates, then subcultured twice in liquid media prior to testing. Purity of cultures was verified by Gram stain between each subculture. Solid and liquid culture media differed by isolate as organisms differed in the types of media they could grow (Table 13). For testing growth on different carbohydrates, a stock solution of NYCIII and PYG-mod-YG[295] media were made at 90% normal volume, and without adding glucose. Stock carbohydrate solutions were also made by dissolving glucose (Thermo Fisher, Waltham, MA) or bovine liver glycogen (Sigma-Aldrich, St. Louis, MO) in water to a concentration

of 50mg/mL (for NYCIII media) or 150mg/mL (for PYG-mod-YG media). Carbohydrate stock solutions and water were then filter-sterilized by passing them through a 0.2µm sterile nylon filter (Thermo Fisher, Waltham, MA). On the day of testing, separate formulations of media were prepared by adding filter-sterilized stock carbohydrate solution or water to 90% volume media at a ratio of 1:10, to produce media with 0.5% w/v (for NYCIII media) or 1.5% w/v (for PYG-mod-YG) glucose/glycogen, or no added carbohydrates. 180µL of prepared media were added to wells of a sterile polystyrene 96 well plate (Corning Inc, Kennebunk, MA) then 20µL of bacterial culture were added to the wells. The plates were sealed, transferred to a PLAS Labs Model 830 glove box (PLAS Labs, Lansing, MI), then grown in 5% CO_2, 5% H_2, 90% N_2, and <2% O_2 at 37°C for 48hrs in a Biotek Epoch 2 microplate reader (Agilent, Santa Clara, CA). OD600 of the cultures was measured every 30 minutes. Growth data was subsequently analyzed and visualized in JupyterLab v1.1.4 (v1.1.4) using Python (v3.6.4) and Seaborn (v0.9.0).

Table 13. Bacterial strains and growth media for glycogen metabolism screening.

Species	Strain	Agar Media	Liquid Media
L. crispatus	ATCC33197	NYCIII	NYCIII
L. crispatus	MV-1A-US	NYCIII	NYCIII
L. crispatus	JV-V01	NYCIII	NYCIII
L. crispatus	125-2-CHN	NYCIII	NYCIII
L. crispatus	DNF00082	NYCIII	NYCIII
L. crispatus	DNF00163	NYCIII	NYCIII
L. crispatus	DNF00378	NYCIII	NYCIII
L. crispatus	DNF00458	NYCIII	NYCIII
L. crispatus	DNF00772	NYCIII	NYCIII
L. crispatus	DNF00807	NYCIII	NYCIII
L. crispatus	KA00050	NYCIII	NYCIII
L. crispatus	KA00233	NYCIII	NYCIII
L. crispatus	KA00379	NYCIII	NYCIII
L. iners	DSM13335	NYCIII	NYCIII
L. iners	LEAF2052A-d	NYCIII	NYCIII
L. iners	UPII 60-B	NYCIII	NYCIII
L. iners	UPII 143-D	NYCIII	NYCIII
L. gasser	DSM20243	NYCIII	NYCIII
L. gasseri	JV-V03	NYCIII	NYCIII
L. gasseri	MV-V22	NYCIII	NYCIII
L. gasseri	EX336960VC01	NYCIII	NYCIII
L. gasseri	224-1	NYCIII	NYCIII
L. gasseri	SJ-9E-US	NYCIII	NYCIII
L. gasseri	SV-16A-US	NYCIII	NYCIII
L. jensenii	DSM20557	NYCIII	NYCIII

L. jensenii	208-1	NYCIII	NYCIII
L. jensenii	269-3	NYCIII	NYCIII
L. mulieris	115-3-CHN	NYCIII	NYCIII
L. mulieris	JV-V16	NYCIII	NYCIII
F. vaginae	DSM15829	NYCIII	NYCIII
F. vaginae	DNF00180	NYCIII	NYCIII
G. piotii	JCP8066	NYCIII	NYCIII
G. leopoldii	CCUG72425	NYCIII	NYCIII
G. leopoldii	DNF01205	NYCIII	NYCIII
G. swidsinskii	DNF001180P	NYCIII	NYCIII
G. vaginalis	ATCC14018	NYCIII	NYCIII
G. vaginalis	ATCC14019	NYCIII	NYCIII
G. vaginalis	ATCC49145	NYCIII	NYCIII
P. lacrimalis	345-B	NYCIII	NYCIII
P. lacrimalis	DNF00528	NYCIII	NYCIII
L. vaginalis	EX336960VC11	Brucella H&K	PYG-mod-YG
M. hutchinsoni	KA00182	Brucella H&K	PYG-mod-YG
M. lornae	UPII 199-6	Brucella H&K	PYG-mod-YG
M. vaginalis	BV3C16-1	Brucella H&K	PYG-mod-YG
M. mulieris	DSM2710	Brucella H&K	PYG-mod-YG
M. mulieris	28-1	Brucella H&K	PYG-mod-YG
P. amnii	DSM23384	Brucella H&K	PYG-mod-YG
P. amnii	21A-A	Brucella H&K	PYG-mod-YG
P. bivia	DSM20514	Brucella H&K	PYG-mod-YG
P. bivia	DNF00650	Brucella H&K	PYG-mod-YG
P. timonensis	DSM22865	Brucella H&K	PYG-mod-YG
P. timonensis	5C-B1	Brucella H&K	PYG-mod-YG

Solid agar media and liquid broth media used to grow isolates of vaginal bacteria to screen for ability to metabolize glycogen.

To test growth on glycogen in the presence of exogenous pullulanase, non-glycogen-degrading bacteria were grown, and media were prepared, as described above. An additional preparation of media was made by adding glycogen stock solution at a 1:10 ratio, and *Bacillus lichenformis* pullulanase M2 (Megazyme, Bray, Ireland) at a 1:1000 ratio to 90% volume media without carbohydrates, making glycogen media with 1U/mL of pullulanase enzyme. Microplate cultures were set up, grown, and monitored as described above.

In vitro quantification of glycogen breakdown rates

Bacterial isolates were cultured anaerobically at described above. Agar media used to cultivate bacteria are shown in Table 14. Bacteria were initially recovered from frozen stocks on agar plates, then subsequently subcultured two additional times an solid media, incubating them for

48hrs at each subculture, and verifying their purity by Gram stain. Bacterial growth from the final subculture was suspended in maximum recovery diluent (Sigma-Aldrich, St. Louis, MO) using sterile cotton tipped swabs (Puritan, Guilford, ME). 500µL bacterial suspensions were added to 500µL of 2mg/mL bovine liver glycogen (Sigma-Aldrich, St. Louis, MO) dissolved in MRD, bringing the final concentration to 1mg/mL glycogen. For experiments at pH 4.3 and 5.5, 5M DL-Lactic acid (Sigma-Aldrich, St. Louis, MO) was added to the bacterial suspension to acidify it to the target pH. MColorpHast 4.0-7.0 pH strips (Millipore Sigma, Burlington, MA) were used to confirm suspension pH. Each time the assay was run, 500µL empty MRD was used as a negative control, and 500µL 167µg/mL *Bacillus lichenformis* amylase (Sigma-Aldrich, St. Louis, MO) was used as a positive control for glycogen breakdown. Bacterial suspensions were incubated in the same anaerobic chamber at 37°C for 15 minutes to 24 hours, depending on the activity of the isolate's glycoside hydrolases. During the incubation, plates to count colony forming units (CFUs) in the initial bacterial suspension were set up by diluting the original suspension to 1E4, 5E5, 1E5, and 5E5, then spotting 10 separate 10µL spots of each dilution onto half of the isolate's preferred agar media. These plates were grown anaerobically at 37°C as described above before colonies were counted and CFUs/mL calculated for the initial bacterial suspension. Following the isolate's incubation with glycogen, the suspensions were centrifuged at 10,000x*g* for 10min to pellet cells and cellular debris, then 27µL of supernatant were added to wells of a sterile polystyrene 96 well plate (Corning Inc, Kennebunk, MA). 173µL of color development solution[261] consisting of 18mL 30% CaCl2 solution (Sigma-Aldrich, St. Louis, MO) mixed with 145µL of 5% Lugol's iodine (LabChem, Zelienople, PA) was added to each sample in the well. The absorbance of each well at 460nm was read using a Biotek Epoch 2 microplate reader (Agilent, Santa Clara, CA) and compared against a standard curve of glycogen dissolved in MRD to calculate the final glycogen concentration of each sample supernatant. Concentrations were subtracted from the concentration of the negative control to find the amount of glycogen degraded during the incubation, then this number was divided by the CFU/mL of the initial suspension and the

incubation time to determine the rate of glycogen breakdown for the isolate. Three separate suspensions were made at each pH to make three biological replicates for each isolate, and four technical replicates were measured for each sample, and averaged for analysis.

Table 14. Solid media used to grow bacterial strains for *in vitro* glycogen breakdown testing.

Species	Strain	Agar Media
L. crispatus	MV-1A-US	Brucella H&K
L. crispatus	JV-V01	Brucella H&K
L. crispatus	125-2-CHN	Brucella H&K
L. iners	DSM13335	NYCIII
L. iners	UPII 143-D	NYCIII
L. iners	LEAF 2052A-d	NYCIII
L. iners	SPIN 10541	NYCIII
L. gasseri	DSM20243	Brucella H&K
L. gasseri	SJ-9E-US	Brucella H&K
L. gasseri	224-1	Brucella H&K
L. jensenii	DSM20557	Brucella H&K
L. jensenii	208-1	Brucella H&K
L. mulieris	115-3-CHN	Brucella H&K
L. mulieris	JV-V16	Brucella H&K
L. vaginalis	DNF00112	Brucella H&K
L. vaginalis	EX336960VC11	Brucella H&K
G. swidsinskii	CCUG72429	NYCIII
G. swidsinskii	DNF00747	NYCIII
G. leopoldii	CCUG72425	NYCIII
G. vaginalis	ATCC49145	NYCIII
G. vaginalis	ATCC14019	NYCIII
G. piotii	JCP8066	NYCIII
P. amnii	DSM23384	Brucella H&K
P. bivia	DSM20514	Brucella H&K
P. timonensis	DSM22865	Brucella H&K
P. timonensis	5C-B1	Brucella H&K

Solid agar media used to grow isolates of vaginal bacteria to test glycogen breakdown rates of their enzymes. The same media was used for initial cultivation and CFU quantification.

Competition experiments

Bacteria were grown in an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C. *L. crispatus* MV-1A-US, *L. crispatus* JV-V01, *L. crispatus* 125-2-CHN, and *G. vaginalis* ATCC14018 were initially recovered from frozen stocks on NYCIII agar plates by incubating them anaerobically at 37°C for 48hrs. The bacteria

were then subcultured twice into 2mL PYG-mod liquid media for 24hrs, verifying culture purity at each subculture by Gram stain. To get the isolates into roughly the same phase of growth, 1mL of *G. vaginalis* culture was subcultured into 4mL PYG-mod media, and 250µL of the *L. crispatus* cultures were subcultured into 5mL PYG-mod media. These cultures were incubated anaerobically at 37°C for 2hrs. Following this incubation, ODs of all cultures were in the range of 0.26 – 0.44. Cultures were normalized to OD 0.26. 1mL of OD-normalized culture was centrifuged at 10,000x*g* for 10min, the supernatant removed, and pellet stored at -20°C for later analysis. The OD-normalized cultures were then used to inoculate experimental tubes. PYG-mod media with 5mg/mL of either glucose or glycogen was prepared by adding 50mg/mL filter-sterilized glucose or bovine liver glycogen (prepared as described above) to 90% volume PYG-mod media without glucose at a 1:10 ratio. 100µL of OD-normalized *G. vaginalis* and 100µL of OD-normalized *L. crispatus* were then inoculated into 1.8mL of PYG-mod media. Three separate co-culture tubes were set up for each of the three combinations of *L. crispatus* strains and *G. vaginalis*, on both glucose and glycogen. Co-culture tubes were incubated anaerobically at 37°C for 48hrs. Following the incubation, culture tubes were briefly vortexed to resuspend bacteria, 1mL of culture was removed, centrifuged at 10,000x*g* for 10min, the supernatant was removed, and the pellets were stored at -20°C until DNA extraction was performed.

To avoid overloading extraction columns, bacterial pellets were first diluted 1:1000 in PBS (Corning, Manassas, VA), then bacterial gDNA was extracted from diluted pellets using QIAamp BiOstic Bacteremia DNA Kit (Qiagen, Hilden, Germany). Species-specific qPCR to quantify concentration of 16S gene copies was performed for *L. crispatus* and *G. vaginalis* on the extracted gDNA as described previously[296]. qPCR was performed on a QuantStudio 6 Flex Real-Time PCR system (Thermo Fisher, Waltham, MA) and analyzed with QuantStudio Real-Time PCR Software (v1.3). The concentration of 16S copies was normalized by the number of 16S gene copies in the genomes of each bacterial isolate to calculate cells per mL of culture. Bacterial fold-growth was

calculated by dividing the number of *L. crispatus* or *G. vaginalis* cells in the cultures after 48hrs by the number of cells inoculated into the cultures, as determined by qPCR of OD-normalized inoculum cultures. Log relative growth was calculated first by taking the log (base 10) of both the *L. crispatus* and *G. vaginalis* fold-growth in each culture. The log-transformed fold-growth of *L. crispatus* was then divided by the log-transformed fold-growth of *G. vaginalis* to determine the performance of the *L. crispatus* isolates in each culture compared to *G. vaginalis*. Statistical tests were performed using the scipy (v1.3.1) "stats" package. Fold-growth and log relative growth were compared by Student's T test, with significant differences inferred if $P < 0.05$.

SMRTseq of *Gardnerella* genomes

Single molecule real-time sequencing (SMRT-Seq) was performed on a Sequel-I instrument (Pacific Biosciences, Menlo Park, CA) as described previously[297]. Genome and methylome data were submitted to REBASE[276] to identify nucleotide sequences recognized by the restriction enzymes encoded by each isolate.

p1199S isolation and sequencing

Gardnerella strain DNF01199S was grown in NYCIII liquid media an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C. A QIAamp DNA Miniprep kit (Qiagen, Hilden, Germany) was used to isolate p1199S plasmid DNA from a 5mL culture of the bacteria. Isolated DNA was digested with the restriction enzyme XhoI (New England Biolabs, Ipswich, MA), and visualized by gel electrophoresis to confirm presence of the plasmid. Whole plasmid sequencing of p1199S was performed by Plasmidsaurus (Plasmidsaurus, Eugene, OR).

Gardnerella tetracycline resistance testing

Gardnerella isolates were grown in an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C. Bacteria were recovered from frozen stocks on NYCIII agar plates, then subsequently subcultured twice into 2mL NYCIII liquid media. Purity of cultures was verified at each subculture by Gram stain. The final cultures were then normalized to 0.5 McFarland turbidity using fresh NYCIII media, then 25µL of the standardized cultures were spotted onto an NYCIII agar plate, and an NYCIII agar plate containing 16µg/mL tetracycline. 25µL of *L. reuteri* CF48-3A was spotted onto each tetracycline plate as a tetracycline-resistant control[273]. Resistance of isolates was determined according to Clinical and Laboratory Standards Institute antimicrobial susceptibility testing protocols[298].

pMC-p1199S-*tetM* construction and minicircle preparation

p1199S and *tetM* were both cloned from *Gardnerella* p1199S gDNA, then these fragments were assembled with pMC by Gibson assembly[299]. The parental construct was transformed into *E. coli* ZYCY10P3S2T (System Biosciences, Palo Alto, CA) by heat shocking according to the SBI Minicircle DNA Technology user manual. Transformant colonies were selected for by growing on agar plates containing 50µg/mL kanamycin. A successfully transformed colony was chosen and stocked, then induction of minicircles was performed according to the SBI Minicircle DNA Technology user manual by growing the bacteria in media containing arabinose. p1199S-*tetM* minicircle DNA was then purified from the induced *E. coli* using a Qiagen Plasmid Midi Kit (Qiagen, Hilden, Germany). The sequence of purified plasmid was confirmed against the construct sequence by whole-plasmid sequencing performed by Plasmidsaurus (Plasmidsaurus, Eugene, OR). “GGCC” sites on the plasmid were protected by treating them with HaeIII MTase (New England Biolabs, Ipswich, MA) by incubating 1µg of the DNA with 1µL of the enzyme at 37°C for 4hrs in the presence of 160µM S-Adenosyl methionine (New England Biolabs, Ipswich, MA).

Following enzyme treatment, DNA was purified with Monarch PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, MA).

Preparation of competent *G. vaginalis*

Gardnerella strain ATCC14018 was grown in an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C. Cells were recovered from frozen stocks on NYCIII agar plates, then subcultured into NYCIII liquid media. 500µL of culture were then subcultured into 5mL NYCIII media and grown overnight. 4mL of this culture were then inoculated into 40mL of modified de Man-Rogosa-Sharpe media[300] supplemented with 5g/L glucose, 0.4g/L L-cysteine HCl, and 100mL/L filter-sterilized horse serum. The bacteria were grown in these media anaerobically at 37°C until they reached the target OD, then they were harvested by centrifugation at 4°C at 4,052x*g* for 10min, washed two times with buffer consisting of 0.2625g/L citric acid and 0.5M maltose at pH 5.8, then resuspended in 200µL of the same buffer, mixed with 200µL 80% glycerol, and stored at -80°C in 50µL aliquots.

G. vaginalis electrotransformation and verification

Competent cells were mixed on ice with plasmid DNA, then transferred into pre-chilled 0.1cm gap Gene Pulser/MicroPulser Electroporation Cuvettes (Bio-Rad, Hercules, CA). Electroporation was then performed on a Gene Pulser Xcell (Bio-Rad, Hercules, CA) with 25µF, 200Ω, and 2,000V. Immediately following electroporation, 950µL of pre-warmed NYCIII media was added to the cells and they were transferred to an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 to recover at 37°C for 1 hour. 100µL of transformed cells were then plated on NYCIII agar plates and NYCIII agar plates containing 16µg/mL tetracycline. The plates were incubated anaerobically at 37°C for 48hrs, then transformed colonies on the tetracycline plates were counted to determine transformation efficiency. After the first successful transformation, transformant colonies were subsequently

grown in 5mL NYCIII media with 16µg/mL tetracycline, and p1199S-*tetM* was recovered from them using a QIAamp DNA Miniprep kit (Qiagen, Hilden, Germany). The sequence of the purified plasmid was confirmed by sequencing performed by Plasmidsaurus (Plasmidsaurus, Eugene, OR).

Data availability

Database files, other protein sequence files, and metagenomic sequencing data are available at https://doi.org/10.6084/m9.figshare.20164277.v1. Code files and metaproteomic search results are available at https://doi.org/10.5281/zenodo.7749517.

Chapter 8: Works Cited

1 Siddique, S. A. Vaginal anatomy and physiology. *Urogynecology* **9**, 263-272 (2003).

2 Tettamanti Boshier, F. A. *et al.* Complementing 16S rRNA gene amplicon sequencing with total bacterial load to infer absolute species concentrations in the vaginal microbiome. *Msystems* **5**, e00777-00719 (2020).

3 Eckhart, L., Lippens, S., Tschachler, E. & Declercq, W. Cell death by cornification. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1833**, 3471-3480 (2013).

4 Wagner, G. & Levin, R. Oxygen tension of the vaginal surface during sexual stimulation in the human. *Fertility and sterility* **30**, 50-53 (1978).

5 Asscher, A., Turner, C. & De Boer, C. H. Cornification of the human vaginal epithelium. *Journal of anatomy* **90**, 547 (1956).

6 Blaskewicz, C. D., Pudney, J. & Anderson, D. J. Structure and function of intercellular junctions in human cervical and vaginal mucosal epithelia. *Biology of reproduction* **85**, 97-104 (2011).

7 Patton, D. L. *et al.* Epithelial cell layer thickness and immune cell populations in the normal human vagina at different stages of the menstrual cycle. *American journal of obstetrics and gynecology* **183**, 967-973 (2000).

8 Averette, H., Frost, P. & Weinstein, G. Autoradiographic Analysis of Cell Proliferation Kinetics in Human Genital Tissues. *Obstetrics & Gynecology* **31**, 580 (1968).

9 Adnane, M., Meade, K. G. & O'Farrelly, C. Cervico-vaginal mucus (CVM)–an accessible source of immunologically informative biomolecules. *Veterinary research communications* **42**, 255-263 (2018).

10 Yarbrough, V. L., Winkle, S. & Herbst-Kralovetz, M. M. Antimicrobial peptides in the female reproductive tract: a critical component of the mucosal immune barrier with physiological and clinical implications. *Human reproduction update* **21**, 353-377 (2015).

11 Wang, Y.-Y. *et al.* IgG in cervicovaginal mucus traps HSV and prevents vaginal herpes infections. *Mucosal immunology* **7**, 1036-1044 (2014).

12 O'Hanlon, D. E., Come, R. A. & Moench, T. R. Vaginal pH measured in vivo: lactobacilli determine pH and lactic acid concentration. *BMC microbiology* **19**, 1-8 (2019).

13 Hoang, T. *et al.* The cervicovaginal mucus barrier to HIV-1 is diminished in bacterial vaginosis. *PLoS pathogens* **16**, e1008236 (2020).

14 Boskey, E., Cone, R., Whaley, K. & Moench, T. Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. *Human Reproduction* **16**, 1809-1813 (2001).

15 Anderson, D. J., Marathe, J. & Pudney, J. The structure of the human vaginal stratum corneum and its role in immune defense. *American journal of reproductive immunology* **71**, 618-623 (2014).

16 Gorodeski, G. I., Hopfer, U., Liu, C. C. & Margles, E. Estrogen acidifies vaginal pH by up-regulation of proton secretion via the apical membrane of vaginal-ectocervical epithelial cells. *Endocrinology* **146**, 816-824 (2005).

17 Chiazze, L., Brayer, F. T., Macisco, J. J., Parker, M. P. & Duffy, B. J. The length and variability of the human menstrual cycle. *Jama* **203**, 377-380 (1968).

18 Yang, H., Zhou, B., Prinz, M. & Siegel, D. Proteomic analysis of menstrual blood. *Molecular & Cellular Proteomics* **11**, 1024-1035 (2012).

19 WAGNER, G. & OTTESEN, B. Vaginal physiology during menstruation. *Annals of internal medicine* **96**, 921-923 (1982).

20 Owen, D. H. & Katz, D. F. A review of the physical and chemical properties of human semen and the formulation of a semen simulant. *Journal of andrology* **26**, 459-469 (2005).

21 Wolters-Everhardt, E., Dony, J. M., Lemmens, W. A., Doesburg, W. H. & De Pont, J.-J. H. Buffering capacity of human semen. *Fertility and sterility* **46**, 114-119 (1986).
22 Muzny, C. A., Austin, E. L., Harbison, H. S. & Hook, E. W. Sexual partnership characteristics of African American women who have sex with women; impact on sexually transmitted infection risk. *Sexually Transmitted Diseases* **41**, 611-617 (2014).
23 Ewald, H. A. S. & Ewald, P. W. Focus: Ecology and evolution: Natural selection, the microbiome, and public health. *The Yale journal of biology and medicine* **91**, 445 (2018).
24 Iebba, V. *et al.* Eubiosis and dysbiosis: the two sides of the microbiota. *New Microbiol* **39**, 1-12 (2016).
25 Srinivasan, S. *et al.* Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PloS one* **7**, e37818 (2012).
26 Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. Molecular identification of bacteria associated with bacterial vaginosis. *New England Journal of Medicine* **353**, 1899-1911 (2005).
27 Rocha, J. *et al.* Lactobacillus mulieris sp. nov., a new species of Lactobacillus delbrueckii group. *International Journal of Systematic and Evolutionary Microbiology* **70**, 1522-1527 (2020).
28 Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences* **108**, 4680-4687 (2011).
29 Hickey, R. J. *et al.* Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women. *MBio* **6**, 10.1128/mbio. 00097-00015 (2015).
30 Jacobsen, C. N. *et al.* Screening of probiotic activities of forty-seven strains of Lactobacillus spp. by in vitro techniques and evaluation of the colonization ability of five selected strains in humans. *Applied and environmental microbiology* **65**, 4949-4956 (1999).
31 O'Hanlon, D. E., Moench, T. R. & Cone, R. A. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PloS one* **8**, e80074 (2013).
32 Kortum, G., Vogel, W., Andrussow, K., International Union of, P. & Applied Chemistry Commission on Electrochemical, D. *Dissociation constants of organic acids in aqueous solution*. (Butterworths, 1961).
33 Patnaik, R. *et al.* Genome shuffling of Lactobacillus for improved acid tolerance. *Nature biotechnology* **20**, 707-712 (2002).
34 Behera, S. S., Ray, R. C. & Zdolec, N. Lactobacillus plantarum with functional properties: an approach to increase safety and shelf-life of fermented foods. *BioMed research international* **2018** (2018).
35 Gong, Z., Luna, Y., Yu, P. & Fan, H. Lactobacilli inactivate Chlamydia trachomatis through lactic acid but not H2O2. *PloS one* **9**, e107758 (2014).
36 Aldunate, M. *et al.* Vaginal concentrations of lactic acid potently inactivate HIV. *Journal of Antimicrobial Chemotherapy* **68**, 2015-2025 (2013).
37 Brittingham, A. & Wilson, W. A. The antimicrobial effect of boric acid on Trichomonas vaginalis. *Sexually transmitted diseases* **41**, 718-722 (2014).
38 Isaacs, C. E. & Xu, W. Theaflavin-3, 3'-digallate and lactic acid combinations reduce herpes simplex virus infectivity. *Antimicrobial agents and chemotherapy* **57**, 3806-3814 (2013).
39 O'Hanlon, D. E., Moench, T. R. & Cone, R. A. In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. *BMC infectious diseases* **11**, 1-8 (2011).
40 Graver, M. A. & Wade, J. J. The role of acidification in the inhibition of Neisseria gonorrhoeae by vaginal lactobacilli during anaerobic growth. *Annals of Clinical Microbiology and Antimicrobials* **10**, 1-5 (2011).

41 Lai, S. K. *et al.* Human immunodeficiency virus type 1 is trapped by acidic but not by neutralized human cervicovaginal mucus. *Journal of virology* **83**, 11196-11200 (2009).
42 Alakomi, H. L. *et al.* Lactic Acid Permeabilizes Gram-Negative Bacteria by Disrupting the Outer Membrane. *Applied and Environmental Microbiology* **66**, 2001-2005, doi:10.1128/AEM.66.5.2001-2005.2000 (2000).
43 Ray, B. & Daeschel, M. *Food biopreservatives of microbial origin*. (CRC Press, 1992).
44 Bradford, L. L. & Ravel, J. The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. *Virulence* **8**, 342-351 (2017).
45 Stoyancheva, G., Marzotto, M., Dellaglio, F. & Torriani, S. Bacteriocin production and gene sequencing analysis from vaginal Lactobacillus strains. *Archives of microbiology* **196**, 645-653, doi:10.1007/s00203-014-1003-1 (2014).
46 Acedo, J. Z. *et al.* Solution structure of acidocin B, a circular bacteriocin produced by Lactobacillus acidophilus M46. *Applied and Environmental Microbiology* **81**, 2910-2918, doi:10.1128/AEM.04265-14

10.1128/AEM.04265-14. (2015).
47 Ojala, T. *et al.* Comparative genomics of Lactobacillus crispatus suggests novel mechanisms for the competitive exclusion of Gardnerella vaginalis. *BMC genomics* **15**, 1070 (2014).
48 Al Kassaa, I., Hamze, M., Hober, D., Chihib, N.-E. & Drider, D. Identification of Vaginal Lactobacilli with Potential Probiotic Properties Isolated from Women in North Lebanon. *Microbial ecology* **67**, 722-734, doi:10.1007/s00248-014-0384-7 (2014).
49 Munoz, A. *et al.* Modeling the temporal dynamics of cervicovaginal microbiota identifies targets that may promote reproductive health. *Microbiome* **9**, 1-12 (2021).
50 Petrova, M. I., Reid, G., Vaneechoutte, M. & Lebeer, S. Lactobacillus iners: friend or foe? *Trends in microbiology* **25**, 182-191 (2017).
51 Verstraelen, H. *et al.* Longitudinal analysis of the vaginal microflora in pregnancy suggests that L. crispatus promotes the stability of the normal vaginal microflora and that L. gasseri and/or L. iners are more conducive to the occurrence of abnormal vaginal microflora. *BMC microbiology* **9**, 1-10 (2009).
52 Rampersaud, R. *et al.* Inerolysin, a cholesterol-dependent cytolysin produced by Lactobacillus iners. *Journal of bacteriology* **193**, 1034-1041 (2011).
53 Koumans, E. H. *et al.* The prevalence of bacterial vaginosis in the United States, 2001–2004; associations with symptoms, sexual behaviors, and reproductive health. *Sexually transmitted diseases* **34**, 864-869 (2007).
54 Fredricks, D. N., Fiedler, T. L., Thomas, K. K., Oakley, B. B. & Marrazzo, J. M. Targeted PCR for detection of vaginal bacteria associated with bacterial vaginosis. *Journal of clinical microbiology* **45**, 3270-3276 (2007).
55 Verstraelen, H. & Verhelst, R. Bacterial vaginosis: an update on diagnosis and treatment. *Expert review of anti-infective therapy* **7**, 1109-1124, doi:10.1586/eri.09.87 (2009).
56 Amsel, R. *et al.* Nonspecific vaginitis: diagnostic criteria and microbial and epidemiologic associations. *The American journal of medicine* **74**, 14-22 (1983).
57 Nugent, R. P., Krohn, M. A. & Hillier, S. L. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *Journal of clinical microbiology* **29**, 297-301 (1991).
58 Cauci, S. *et al.* Prevalence of Bacterial Vaginosis and Vaginal Flora Changes in Peri- and Postmenopausal Women. *Journal of Clinical Microbiology* **40**, 2147-2152, doi:10.1128/JCM.40.6.2147-2152.2002 (2002).

59 Bahram, A., Hamid, B. & Zohre, T. Prevalence of bacterial vaginosis and impact of genital hygiene practices in non-pregnant women in zanjan, iran. *Oman medical journal* **24**, 288-293, doi:10.5001/omj.2009.58 (2009).
60 Bhalla, P. *et al.* Prevalence of bacterial vaginosis among women in Delhi, India. *Indian journal of medical research (New Delhi, India : 1994)* **125**, 167-172 (2007).
61 Sewankambo, N. *et al.* HIV-1 infection associated with abnormal vaginal flora morphology and bacterial vaginosis. *The Lancet (British edition)* **350**, 546-550, doi:10.1016/S0140-6736(97)01063-5 (1997).
62 Svare, J. A., Schmidt, H., Hansen, B. B. & Lose, G. Bacterial vaginosis in a cohort of Danish pregnant women: prevalence and relationship with preterm delivery, low birthweight and perinatal infections. *BJOG : an international journal of obstetrics and gynaecology* **113**, 1419-1425, doi:10.1111/j.1471-0528.2006.01087.x (2006).
63 Klebanoff, M. A. *et al.* Vulvovaginal symptoms in women with bacterial vaginosis. *Obstetrics and gynecology (New York. 1953)* **104**, 267-272, doi:10.1097/01.AOG.0000134783.98382.b0 (2004).
64 Marrazzo, J. M., Thomas, K. K., Agnew, K. & Ringwood, K. Prevalence and risks for bacterial vaginosis in women who have sex with women. *Sexually transmitted diseases* **37**, 335 (2010).
65 Haggerty, C. L. *et al.* Bacterial vaginosis and anaerobic bacteria are associated with endometritis. *Clinical Infectious Diseases* **39**, 990-995 (2004).
66 Fettweis, J. M. *et al.* The vaginal microbiome and preterm birth. *Nature medicine* **25**, 1012-1021 (2019).
67 Borgdorff, H. *et al.* Lactobacillus-dominated cervicovaginal microbiota associated with reduced HIV/STI prevalence and genital HIV viral load in African women. *The ISME journal* **8**, 1781-1793 (2014).
68 Taha, T. E. *et al.* Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *Aids* **12**, 1699-1706 (1998).
69 Schwebke, J. R. & Desmond, R. A RANDOMIZED TRIAL OF METRONIDAZOLE IN ASYMPTOMATIC BV TO PREVENT ACQUISITION OF STDS. *American journal of obstetrics and gynecology* **196**, 517.e511-517.e516, doi:10.1016/j.ajog.2007.02.048 (2007).
70 De Seta, F. *et al.* The vaginal microbiome: III. the vaginal microbiome in various urogenital disorders. *Journal of Lower Genital Tract Disease* **26**, 85 (2022).
71 Bilardi, J. E. *et al.* The burden of bacterial vaginosis: women's experience of the physical, emotional, sexual and social impact of living with recurrent bacterial vaginosis. *PloS one* **8**, e74378 (2013).
72 Bilardi, J. *et al.* Women's Management of Recurrent Bacterial Vaginosis and Experiences of Clinical Care: A Qualitative Study. *PloS one* **11**, e0151794-e0151794, doi:10.1371/journal.pone.0151794 (2016).
73 Fashemi, B., Delaney, M. L., Onderdonk, A. B. & Fichorova, R. N. Effects of feminine hygiene products on the vaginal mucosal biome. *Microbial ecology in health and disease* **24**, 19703 (2013).
74 Ness, R. B. *et al.* Douching in relation to bacterial vaginosis, lactobacilli, and facultative bacteria in the vagina. *Obstetrics & Gynecology* **100**, 765-772 (2002).
75 Brotman, R. M. *et al.* The effect of vaginal douching cessation on bacterial vaginosis: a pilot study. *American journal of obstetrics and gynecology* **198**, 628. e621-628. e627 (2008).
76 Brotman, R. M. *et al.* A longitudinal study of vaginal douching and bacterial vaginosis—a marginal structural modeling analysis. *American journal of epidemiology* **168**, 188-196 (2008).
77 Workowski, K. A. & Bolan, G. A. Sexually Transmitted Diseases Treatment Guidelines, 2015. *MMWR. Recommendations and reports* **64**, 1-137 (2015).

78 Koumans, E. H., Markowitz, L. E., Hogan, V. & group, C. B. w. Indications for therapy and treatment recommendations for bacterial vaginosis in nonpregnant and pregnant women: a synthesis of data. *Clinical Infectious Diseases* **35**, S152-S172 (2002).
79 Bradshaw, C. S. *et al.* High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *The Journal of infectious diseases* **193**, 1478-1486 (2006).
80 Plummer, E. L. *et al.* Lactic acid-containing products for bacterial vaginosis and their impact on the vaginal microbiota: A systematic review. *PloS one* **16**, e0246953 (2021).
81 Pino, A. *et al.* Bacterial biota of women with bacterial vaginosis treated with lactoferrin: an open prospective randomized trial. *Microbial ecology in health and disease* **28**, 1357417 (2017).
82 Miranda, M. *et al.* Vaginal lactoferrin in prevention of preterm birth in women with bacterial vaginosis. *The journal of maternal-fetal & neonatal medicine* **34**, 3704-3708, doi:10.1080/14767058.2019.1690445 (2021).
83 Reichman, O., Akins, R. & Sobel, J. D. Boric acid addition to suppressive antimicrobial therapy for recurrent bacterial vaginosis. *Sexually transmitted diseases*, 732-734 (2009).
84 Cohen, C. R. *et al.* Randomized trial of Lactin-V to prevent recurrence of bacterial vaginosis. *New England Journal of Medicine* **382**, 1906-1915 (2020).
85 Vujic, G., Jajac Knez, A., Despot Stefanovic, V. & Kuzmic Vrbanovic, V. Efficacy of orally applied probiotic capsules for bacterial vaginosis and other vaginal infections: a double-blind, randomized, placebo-controlled study. *European journal of obstetrics & gynecology and reproductive biology* **168**, 75-79, doi:10.1016/j.ejogrb.2012.12.031 (2013).
86 Lev-Sagie, A. *et al.* Vaginal microbiome transplantation in women with intractable bacterial vaginosis. *Nature medicine* **25**, 1500-1504 (2019).
87 Gardner, H. L. & Dukes, C. D. Haemophilus vaginalis vaginitis: a newly defined specific infection previously classified “nonspecific” vaginitis. *American journal of obstetrics and gynecology* **69**, 962-976 (1955).
88 CRISWELL, B. S., LADWIG, C. L., GARDNER, H. L. & Dukes, C. Haemophilus vaginalis: vaginitis by inoculation from culture. *Obstetrics & Gynecology* **33**, 195-199 (1969).
89 Lewis, W. G. *et al.* Hydrolysis of secreted sialoglycoprotein immunoglobulin A (IgA) in ex vivo and biochemical models of bacterial vaginosis. *Journal of Biological Chemistry* **287**, 2079-2089 (2012).
90 Lewis, W. G., Robinson, L. S., Gilbert, N. M., Perry, J. C. & Lewis, A. L. Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted Actinobacterium Gardnerella vaginalis. *Journal of Biological Chemistry* **288**, 12067-12079 (2013).
91 Robinson, L. S., Schwebke, J., Lewis, W. G. & Lewis, A. L. Identification and characterization of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium Gardnerella vaginalis. *Journal of Biological Chemistry* **294**, 5230-5245 (2019).
92 Briselden, A. M., Moncla, B. J., Stevens, C. E. & Hillier, S. L. Sialidases (neuraminidases) in bacterial vaginosis and bacterial vaginosis-associated microflora. *Journal of clinical microbiology* **30**, 663-666 (1992).
93 Udayalaxmi, J., Bhat, G. & Kotigadde, S. Biotypes and virulence factors of Gardnerella vaginalis isolated from cases of bacterial vaginosis. *Indian Journal of Medical Microbiology* **29**, 165-168 (2011).
94 Lithgow, K., Bagheri, S. & Sycuro, L. Secreted proteolytic activity of vaginal prevotella species remodels structural components of cervical and uterine tissues. *American Journal of Obstetrics & Gynecology* **228**, S789-S790 (2023).

95 Gelber, S. E., Aguilar, J. L., Lewis, K. L. & Ratner, A. J. Functional and phylogenetic characterization of Vaginolysin, the human-specific cytolysin from Gardnerella vaginalis. *Journal of bacteriology* **190**, 3896-3903 (2008).
96 Garcia, E. M., Kraskauskiene, V., Koblinski, J. E. & Jefferson, K. K. Interaction of Gardnerella vaginalis and vaginolysin with the apical versus basolateral face of a three-dimensional model of vaginal epithelium. *Infection and immunity* **87** (2019).
97 Ragaliauskas, T. *et al.* Inerolysin and vaginolysin, the cytolysins implicated in vaginal dysbiosis, differently impair molecular integrity of phospholipid membranes. *Scientific reports* **9**, 10606 (2019).
98 Patterson, J. L., Girerd, P. H., Karjane, N. W. & Jefferson, K. K. Effect of biofilm phenotype on resistance of Gardnerella vaginalis to hydrogen peroxide and lactic acid. *American journal of obstetrics and gynecology* **197**, 170. e171-170. e177 (2007).
99 Wolrath, H., Forsum, U., Larsson, P.-G. & Borén, H. Analysis of bacterial vaginosis-related amines in vaginal fluid by gas chromatography and mass spectrometry. *Journal of Clinical Microbiology* **39**, 4026-4031 (2001).
100 Wolrath, H., Borén, H., Hallén, A. & Forsum, U. Trimethylamine content in vaginal secretion and its relation to bacterial vaginosis. *Apmis* **110**, 819-824 (2002).
101 Brand, J. & Galask, R. Trimethylamine: the substance mainly responsible for the fishy odor often associated with bacterial vaginosis. *Obstetrics & Gynecology* **68**, 682-685 (1986).
102 Hardy, L. *et al.* A fruitful alliance: the synergy between Atopobium vaginae and Gardnerella vaginalis in bacterial vaginosis-associated biofilm. *Sexually transmitted infections* **92**, 487-491 (2016).
103 Castro, J., Rosca, A. S., Cools, P., Vaneechoutte, M. & Cerca, N. Gardnerella vaginalis enhances Atopobium vaginae viability in an in vitro model. *Frontiers in Cellular and Infection Microbiology* **10**, 83 (2020).
104 Pybus, V. & Onderdonk, A. B. Evidence for a commensal, symbiotic relationship between Gardnerella vaginalis and Prevotella bivia involving ammonia: potential significance for bacterial vaginosis. *Journal of infectious diseases* **175**, 406-413 (1997).
105 Randis, T. M. & Ratner, A. J. Vol. 220 1085-1088 (Oxford University Press US, 2019).
106 Pybus, V. & Onderdonk, A. B. A commensal symbiosis between Prevotella bivia and Peptostreptococcus anaerobius involves amino acids: potential significance to the pathogenesis of bacterial vaginosis. *FEMS Immunology & Medical Microbiology* **22**, 317-327 (1998).
107 Schwebke, J. R., Muzny, C. A. & Josey, W. E. Role of Gardnerella vaginalis in the pathogenesis of bacterial vaginosis: a conceptual model. *The Journal of infectious diseases* **210**, 338-343 (2014).
108 Swidsinski, A. *et al.* Adherent biofilms in bacterial vaginosis. *Obstetrics & Gynecology* **106**, 1013-1023 (2005).
109 Eren, A. M. *et al.* Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PloS one* **6**, e26732 (2011).
110 Vodstrcil, L. A., Muzny, C. A., Plummer, E. L., Sobel, J. D. & Bradshaw, C. S. Bacterial vaginosis: drivers of recurrence and challenges and opportunities in partner treatment. *BMC medicine* **19**, 1-12 (2021).
111 Muth, T., Renard, B. Y. & Martens, L. Metaproteomic data analysis at a glance: advances in computational microbial community proteomics. *Expert review of proteomics* **13**, 757-769 (2016).
112 Wilmes, P. & Bond, P. L. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends in microbiology* **14**, 92-97 (2006).

113 Wang, X. *et al.* Protein identification using customized protein sequence databases derived from RNA-Seq data. *Journal of proteome research* **11**, 1009-1017 (2012).
114 Zhang, X. & Figeys, D. Perspective and guidelines for metaproteomics in microbiome studies. *Journal of proteome research* **18**, 2370-2380 (2019).
115 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* **10**, 57-63 (2009).
116 Hettich, R. L., Pan, C., Chourey, K. & Giannone, R. J. Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Analytical chemistry* **85**, 4203-4214 (2013).
117 Cappadona, S., Baker, P. R., Cutillas, P. R., Heck, A. J. & van Breukelen, B. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino acids* **43**, 1087-1108 (2012).
118 Nesvizhskii, A. I. Protein identification by tandem mass spectrometry and sequence database searching. *Mass Spectrometry Data Analysis in Proteomics*, 87-119 (2007).
119 Muth, T., Benndorf, D., Reichl, U., Rapp, E. & Martens, L. Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. *Molecular BioSystems* **9**, 578-585 (2013).
120 Muth, T. *et al.* Navigating through metaproteomics data: a logbook of database searching. *Proteomics* **15**, 3439-3453 (2015).
121 Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods* **4**, 207-214 (2007).
122 Granholm, V. & Käll, L. Quality assessments of peptide–spectrum matches in shotgun proteomics. *Proteomics* **11**, 1086-1093 (2011).
123 Na, S., Bandeira, N. & Paek, E. Fast multi-blind modification search through tandem mass spectrometry. *Molecular & Cellular Proteomics* **11** (2012).
124 Knudsen, G. M. & Chalkley, R. J. The effect of using an inappropriate protein database for proteomic data analysis. *PloS one* **6**, e20873 (2011).
125 Timmins-Schiffman, E. *et al.* Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *The ISME journal* **11**, 309-314 (2017).
126 Rechenberger, J. *et al.* Challenges in clinical metaproteomics highlighted by the analysis of acute leukemia patients with gut colonization by multidrug-resistant enterobacteriaceae. *Proteomes* **7**, 2 (2019).
127 Tanca, A. *et al.* Evaluating the impact of different sequence databases on metaproteome analysis: insights from a lab-assembled microbial mixture. *PloS one* **8**, e82981 (2013).
128 Tanca, A. *et al.* The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome* **4**, 51 (2016).
129 Verberkmoes, N. C. *et al.* Shotgun metaproteomics of the human distal gut microbiota. *The ISME journal* **3**, 179-189 (2009).
130 Lee, E. M. *et al.* Optimizing metaproteomics database construction: lessons from a study of the vaginal microbiome. *Msystems*, e00678-00622 (2023).
131 Bittremieux, W. *et al.* Quality control in mass spectrometry-based proteomics. *Mass Spectrometry Reviews* **37**, 697-711 (2018).
132 Frankenfield, A. M., Ni, J., Ahmed, M. & Hao, L. Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. *Journal of Proteome Research* **21**, 2104-2113 (2022).
133 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature genetics* **25**, 25-29 (2000).
134 The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* **49**, D325-D334 (2021).

135 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *bioRxiv* (2021).
136 Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research* **47**, D309-D314 (2019).
137 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research* **44**, D733-D745 (2016).
138 Gevers, D. *et al.* The Human Microbiome Project: a community resource for the healthy human microbiome. (2012).
139 Stock, S. J. *et al.* Elafin (SKALP/Trappin-2/proteinase inhibitor-3) is produced by the cervix in pregnancy and cervicovaginal levels are diminished in bacterial vaginosis. *Reproductive sciences* **16**, 1125-1134 (2009).
140 Borgdorff, H. *et al.* Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. *Mucosal Immunol* **9**, 621-633, doi:10.1038/mi.2015.86 (2016).
141 Zevin, A. S. *et al.* Microbiome composition and function drives wound-healing impairment in the female genital tract. *PLoS pathogens* **12** (2016).
142 Ferreira, C. S. T., da Silva, M. G., de Pontes, L. G., Dos Santos, L. D. & Marconi, C. Protein content of cervicovaginal fluid is altered during bacterial vaginosis. *Journal of lower genital tract disease* **22**, 147-151 (2018).
143 Bradley, F. *et al.* The vaginal microbiome amplifies sex hormone-associated cyclic changes in cervicovaginal inflammation and epithelial barrier disruption. *American Journal of Reproductive Immunology* **80**, e12863 (2018).
144 Dasari, S. *et al.* Comprehensive Proteomic Analysis of Human Cervical− Vaginal Fluid. *Journal of proteome research* **6**, 1258-1268 (2007).
145 Shaw, J. L., Smith, C. R. & Diamandis, E. P. Proteomic analysis of human cervico-vaginal fluid. *Journal of proteome research* **6**, 2859-2865 (2007).
146 Zegels, G., Van Raemdonck, G. A., Coen, E. P., Tjalma, W. A. & Van Ostade, X. W. Comprehensive proteomic analysis of human cervical-vaginal fluid using colposcopy samples. *Proteome science* **7**, 17 (2009).
147 Burgener, A. *et al.* Comprehensive proteomic study identifies serpin and cystatin antiproteases as novel correlates of HIV-1 resistance in the cervicovaginal mucosa of female sex workers. *Journal of proteome research* **10**, 5139-5149 (2011).
148 Birse, K. *et al.* Molecular signatures of immune activation and epithelial barrier remodeling are enhanced during the luteal phase of the menstrual cycle: implications for HIV susceptibility. *Journal of virology* **89**, 8793-8805 (2015).
149 Muytjens, C. M., Yu, Y. & Diamandis, E. P. Discovery of antimicrobial peptides in cervical-vaginal fluid from healthy nonpregnant women via an integrated proteome and Peptidome analysis. *Proteomics* **17**, 1600461 (2017).
150 Starodubtseva, N. L. *et al.* Label-free cervicovaginal fluid proteome profiling reflects the cervix neoplastic transformation. *Journal of Mass Spectrometry* **54**, 693-703 (2019).
151 Kumar, B. *et al.* Dynamic Alteration in the Vaginal Secretory Proteome across the Early and Mid-Trimesters of Pregnancy. *Journal of Proteome Research* **20**, 1190-1205 (2021).
152 Klatt, N. R. *et al.* Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938-945 (2017).

153 Cruciani, F. *et al.* Proteome profiles of vaginal fluids from women affected by bacterial vaginosis and healthy controls: outcomes of rifaximin treatment. *Journal of Antimicrobial Chemotherapy* **68**, 2648-2659 (2013).
154 Arnold, K. *et al.* Mucosal integrity factors are perturbed during bacterial vaginosis: a proteomic analysis. *AIDS research and human retroviruses* **30**, A30-A30 (2014).
155 Borgdorff, H. *et al.* Unique insights in the cervicovaginal Lactobacillus iners and L. crispatus proteomes and their associations with microbiota dysbiosis. *PloS one* **11** (2016).
156 Farr Zuend, C. *et al.* Pregnancy associates with alterations to the host and microbial proteome in vaginal mucosa. *American Journal of Reproductive Immunology* **83**, e13235 (2020).
157 Alisoltani, A. *et al.* Microbial function and genital inflammation in young South African women at high risk of HIV infection. *Microbiome* **8**, 1-21 (2020).
158 Nunn, K. L. *et al.* Amylases in the Human Vagina. *Msphere* **5**, e00943-00920 (2020).
159 Nesvizhskii, A. I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature methods* **4**, 787-797 (2007).
160 Jagtap, P. *et al.* A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* **13**, 1352-1357 (2013).
161 Johnson, R. S. *et al.* Assessing protein sequence database suitability using de novo sequencing. *Molecular & Cellular Proteomics* **19**, 198-208 (2020).
162 Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382-387 (2014).
163 Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55-62 (2016).
164 May, D. H. *et al.* An alignment-free "metapeptide" strategy for metaproteomic characterization of microbiome samples using shotgun metagenomic sequencing. *Journal of proteome research* **15**, 2697-2705 (2016).
165 Xiao, J. *et al.* Metagenomic taxonomy-guided database-searching strategy for improving metaproteomic analysis. *Journal of proteome research* **17**, 1596-1605 (2018).
166 Géron, A., Werner, J., Wattiez, R., Lebaron, P. & Matallana Surget, S. Deciphering the functioning of microbial communities: shedding light on the critical steps in metaproteomics. *Frontiers in microbiology* **10**, 2395 (2019).
167 Srinivasan, S. *et al.* Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PloS one* **5**, e10197 (2010).
168 Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *PLoS biology* **14**, e1002533 (2016).
169 Li, K., Bihan, M., Yooseph, S. & Methe, B. A. Analyses of the microbial diversity across the human microbiome. *PloS one* **7**, e32118 (2012).
170 Matthiesen, R., Prieto, G. & Beck, H. C. in *Mass Spectrometry Data Analysis in Proteomics* 133-143 (Springer, 2020).
171 Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature communications* **5**, 5277 (2014).
172 Wu, C. *et al.* An optimized informatics pipeline for mass spectrometry-based peptidomics. *Journal of the American Society for Mass Spectrometry* **26**, 2002-2008 (2015).
173 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of molecular biology* **428**, 726-731 (2016).
174 Kleiner, M. Metaproteomics: much more than measuring gene expression in microbial communities. *Msystems* **4**, 10.1128/msystems. 00115-00119 (2019).

175 Kleiner, M. *et al.* Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proceedings of the National Academy of Sciences* **109**, E1173-E1182 (2012).
176 Bolam, D. N. & Koropatkin, N. M. Glycan recognition by the Bacteroidetes Sus-like systems. *Current opinion in structural biology* **22**, 563-569 (2012).
177 Srinivasan, S. *et al.* Metabolic signatures of bacterial vaginosis. *MBio* **6** (2015).
178 Kieliszek, M., Pobiega, K., Piwowarek, K. & Kot, A. M. Characteristics of the proteolytic enzymes produced by lactic acid bacteria. *Molecules* **26**, 1858 (2021).
179 REITZER, L. J. Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, L-alanine and D-alanine. *In Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology 2*, 302-320 (1987).
180 Reitzer, L. Nitrogen assimilation and global regulation in Escherichia coli. *Annual Reviews in Microbiology* **57**, 155-176 (2003).
181 Eisenberg, D., Gill, H. S., Pfluegl, G. M. & Rotstein, S. H. Structure–function relationships of glutamine synthetases. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1477**, 122-145 (2000).
182 Reyes, N., Ginter, C. & Boudker, O. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* **462**, 880-885 (2009).
183 Caldwell, R. W., Rodriguez, P. C., Toque, H. A., Narayanan, S. P. & Caldwell, R. B. Arginase: a multifaceted enzyme important in health and disease. *Physiological reviews* **98**, 641-665 (2018).
184 Barber, M. F. & Elde, N. C. Buried treasure: evolutionary perspectives on microbial iron piracy. *Trends in Genetics* **31**, 627-636 (2015).
185 Jarosik, G. P., Land, C. B., Duhon, P., Chandler Jr, R. & Mercer, T. Acquisition of iron by Gardnerella vaginalis. *Infection and immunity* **66**, 5041-5047 (1998).
186 Jarosik, G. P. & Land, C. B. Identification of a human lactoferrin-binding protein in Gardnerella vaginalis. *Infection and immunity* **68**, 3443-3447 (2000).
187 Tolosano, E. & Altruda, F. Hemopexin: structure, function, and regulation. *DNA and cell biology* **21**, 297-306 (2002).
188 Tolosano, E., Fagoonee, S., Morello, N., Vinchi, F. & Fiorito, V. Heme scavenging and the other facets of hemopexin. *Antioxidants & redox signaling* **12**, 305-320 (2010).
189 Abraham, N. G. & Kappas, A. Pharmacological and clinical aspects of heme oxygenase. *Pharmacological reviews* **60**, 79-127 (2008).
190 Eschenbach, D. A. *et al.* Prevalence of hydrogen peroxide-producing Lactobacillus species in normal women and women with bacterial vaginosis. *Journal of clinical microbiology* **27**, 251-256 (1989).
191 Hillier, S. L., Krohn, M. A., Klebanoff, S. J. & Eschenbach, D. A. The relationship of hydrogen peroxide-producing lactobacilli to bacterial vaginosis and genital microflora in pregnant women. *Obstetrics and Gynecology* **79**, 369-373 (1992).
192 Hawes, S. E. *et al.* Hydrogen peroxide—producing lactobacilli and acquisition of vaginal infections. *Journal of Infectious Diseases* **174**, 1058-1063 (1996).
193 Vallor, A. C., Antonio, M. A., Hawes, S. E. & Hillier, S. L. Factors associated with acquisition of, or persistent colonization by, vaginal lactobacilli: role of hydrogen peroxide production. *The Journal of infectious diseases* **184**, 1431-1436 (2001).
194 O'Hanlon, D. E., Lanier, B. R., Moench, T. R. & Cone, R. A. Cervicovaginal fluid and semen block the microbicidal activity of hydrogen peroxide produced by vaginal lactobacilli. *BMC infectious diseases* **10**, 1-8 (2010).
195 Holmes, K. K., Chen, K. C., Lipinski, C. M. & Eschenbach, D. A. Vaginal redox potential in bacterial vaginosis (nonspecific vaginitis). *Journal of Infectious Diseases* **152**, 379-382 (1985).

196 Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *Journal of molecular biology* **160**, 181-217 (1982).
197 Fujii, J., Ikeda, Y., Kurahashi, T. & Homma, T. Physiological and pathological views of peroxiredoxin 4. *Free Radical Biology and Medicine* **83**, 373-379 (2015).
198 Janoff, A. Elastase in tissue injury. *Annual review of medicine* **36**, 207-216 (1985).
199 Sahoo, M., Del Barrio, L., Miller, M. A. & Re, F. Neutrophil elastase causes tissue damage that decreases host tolerance to lung infection with burkholderia species. *PLoS Pathogens* **10**, e1004327 (2014).
200 Fahrbach, K. M., Malykhina, O., Stieh, D. J. & Hope, T. J. Differential binding of IgG and IgA to mucus of the female reproductive tract. *PloS one* **8**, e76176 (2013).
201 Sarma, J. V. & Ward, P. A. The complement system. *Cell and tissue research* **343**, 227-235 (2011).
202 Delgado-Diaz, D. J. *et al.* Lactic acid from vaginal microbiota enhances cervicovaginal epithelial barrier integrity by promoting tight junction protein expression. *Microbiome* **10**, 1-16 (2022).
203 Zegels, G., Van Raemdonck, G. A., Tjalma, W. A. & Van Ostade, X. W. Use of cervicovaginal fluid for the identification of biomarkers for pathologies of the female genital tract. *Proteome science* **8**, 1-23 (2010).
204 Marconi, C., Donders, G. G., Parada, C. M., Giraldo, P. C. & da Silva, M. G. Do Atopobium vaginae, Megasphaera sp. and Leptotrichia sp. change the local innate immune response and sialidase activity in bacterial vaginosis? *Sexually transmitted infections* **89**, 167-173 (2013).
205 Foley, M. H., Cockburn, D. W. & Koropatkin, N. M. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cellular and Molecular Life Sciences* **73**, 2603-2617 (2016).
206 Shipman, J. A., Berleman, J. E. & Salyers, A. A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *Journal of bacteriology* **182**, 5365-5372 (2000).
207 Pollet, R. M., Martin, L. M. & Koropatkin, N. M. Tonb-dependent transporters in the bacteroidetes: Unique domain structures and potential functions. *Molecular Microbiology* **115**, 490-501 (2021).
208 Lithgow, K. V. *et al.* Resolution of glycogen and glycogen-degrading activities reveals correlates of Lactobacillus crispatus dominance in a cohort of young African women. *bioRxiv*, 2022.2003. 2029.486257 (2022).
209 Spear, G. T. *et al.* Human α-amylase present in lower-genital-tract mucosal fluid processes glycogen to support vaginal colonization by Lactobacillus. *The Journal of infectious diseases* **210**, 1019-1028 (2014).
210 Newgard, C. B., Hwang, P. K. & Fletterick, R. J. The family of glycogen phosphorylases: structure and function. *Critical reviews in biochemistry and molecular biology* **24**, 69-99 (1989).
211 Stewart-Tull, D. Evidence that vaginal lactobacilli do not ferment glycogen. *American journal of obstetrics and gynecology* **88**, 676-679 (1964).
212 Wylie, J. G. & Henderson, A. Identity and glycogen-fermenting ability of lactobacilli isolated from the vagina of pregnant women. *Journal of medical microbiology* **2**, 363-366 (1969).
213 Van Der Veer, C. *et al.* Comparative genomics of human Lactobacillus crispatus isolates reveals genes for glycosylation and glycogen degradation: implications for in vivo dominance of the vaginal microbiota. *Microbiome* **7**, 1-14 (2019).
214 Hertzberger, R. *et al.* Genetic elements orchestrating Lactobacillus crispatus glycogen metabolism in the vagina. *International Journal of Molecular Sciences* **23**, 5590 (2022).

215 Woolston, B. M., Jenkins, D. J., Hood-Pishchany, M. I., Nahoum, S. R. & Balskus, E. P. Characterization of vaginal microbial enzymes identifies amylopullulanases that support growth of Lactobacillus crispatus on glycogen. *bioRxiv*, 2021.2007. 2019.452977 (2021).
216 Bhandari, P., Tingley, J. P., Abbott, D. W. & Hill, J. E. Characterization of an α-glucosidase enzyme conserved in Gardnerella spp. isolated from the human vaginal microbiome. *bioRxiv* (2020).
217 Bhandari, P., Tingley, J., Abbott, D. W. & Hill, J. E. Glycogen-Degrading Activities of Catalytic Domains of α-Amylase and α-Amylase-Pullulanase Enzymes Conserved in Gardnerella spp. from the Vaginal Microbiome. *Journal of Bacteriology* **205**, e00393-00322 (2023).
218 McMillan, A. *et al.* A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. *Scientific reports* **5**, 1-14 (2015).
219 Yeoman, C. J. *et al.* A multi-omic systems-based approach reveals metabolic markers of bacterial vaginosis and insight into the disease. *PloS one* **8**, e56111 (2013).
220 Featherstone, J. & Rodgers, B. Effect of acetic, lactic and other organic acids on the formation of artificial carious lesions. *Caries research* **15**, 377-385 (1981).
221 Nghiem, N. P., Kleff, S. & Schwegmann, S. Succinic acid: technology development and commercialization. *Fermentation* **3**, 26 (2017).
222 Riddick, J. A., Bunger, W. B. & Sakano, T. K. Organic solvents: physical properties and methods of purification. (1986).
223 Al-Mushrif, S., Eley, A. & Jones, B. Inhibition of chemotaxis by organic acids from anaerobes may prevent a purulent response in bacterial vaginosis. *Journal of medical microbiology* **49**, 1023-1030 (2000).
224 Prue, J. & Read, A. Acidity constant of formic acid. *Transactions of the Faraday Society* **62**, 1271-1274 (1966).
225 Latham, K. G., Ferguson, A. & Donne, S. W. Influence of ammonium salts and temperature on the yield, morphology and chemical structure of hydrothermally carbonized saccharides. *SN Applied Sciences* **1**, 1-13 (2019).
226 Girinathan, B. P., Braun, S., Sirigireddy, A. R., Lopez, J. E. & Govind, R. Importance of glutamate dehydrogenase (GDH) in Clostridium difficile colonization in vivo. *PLoS One* **11**, e0160107 (2016).
227 Murray, D. S. *et al.* Structures of the Bacillus subtilis glutamine synthetase dodecamer reveal large intersubunit catalytic conformational changes linked to a unique feedback inhibition mechanism. *Journal of Biological Chemistry* **288**, 35801-35811 (2013).
228 Nelson, T. M. *et al.* Vaginal biogenic amines: biomarkers of bacterial vaginosis or precursors to vaginal dysbiosis? *Frontiers in physiology* **6**, 253 (2015).
229 Macklaim, J. M. *et al.* Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis. *Microbiome* **1**, 12 (2013).
230 Murata, T. *et al.* Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination. *Breast cancer research and treatment* **177**, 591-601 (2019).
231 Fernández, M. & Zúñiga, M. Amino acid catabolic pathways of lactic acid bacteria. *Critical reviews in microbiology* **32**, 155-183 (2006).
232 Kimberly, M. M. & Goldstein, J. Determination of pKa values and total proton distribution pattern of spermidine by carbon-13 nuclear magnetic resonance titrations. *Analytical chemistry* **53**, 789-793 (1981).
233 Watson, N., Dunyak, D., Rosey, E., Slonczewski, J. & Olson, E. Identification of elements involved in transcriptional regulation of the Escherichia coli cad operon by external pH. *Journal of bacteriology* **174**, 530-540 (1992).

234 Lian, J. *et al.* The role of polyamine metabolism in remodeling immune responses and blocking therapy within the tumor immune microenvironment. *Frontiers in Immunology* **13**, 912279 (2022).
235 Folk, J. & Cole, P. Transglutaminase: mechanistic features of the active site as determined by kinetic and inhibitor studies. *Biochimica et Biophysica Acta (BBA)-Enzymology and Biological Oxidation* **122**, 244-264 (1966).
236 Dolynchuk, K. N., Bendor-Samuel, R. & Bowness, J. M. Effect of putrescine on tissue transglutaminase activity in wounds: decreased breaking strength and increased matrix fucoprotein solubility. *Plastic and reconstructive surgery* **93**, 567-573 (1994).
237 Upchurch, H. F., Conway, E., Patterson Jr, M. & Maxwell, M. D. Localization of cellular transglutaminase on the extracellular matrix after wounding: characteristics of the matrix bound enzyme. *Journal of cellular physiology* **149**, 375-382 (1991).
238 Wagner, G., Bohr, L., Wagner, P. & Petersen, L. N. Tampon-induced changes in vaginal oxygen and carbon dioxide tensions. *American journal of obstetrics and gynecology* **148**, 147-150 (1984).
239 Hill, D. R. *et al.* In vivo assessment of human vaginal oxygen and carbon dioxide levels during and post menses. *Journal of Applied Physiology* **99**, 1582-1591 (2005).
240 Aft, R. L. & Mueller, G. Hemin-mediated DNA strand scission. *Journal of Biological Chemistry* **258**, 12069-12072 (1983).
241 Aft, R. L. & Mueller, G. Hemin-mediated oxidative degradation of proteins. *Journal of Biological Chemistry* **259**, 301-305 (1984).
242 Gutteridge, J. & Smith, A. Antioxidant protection by haemopexin of haem-stimulated lipid peroxidation. *Biochemical Journal* **256**, 861-865 (1988).
243 Nagababu, E. & Rifkind, J. M. Heme degradation by reactive oxygen species. *Antioxidants & redox signaling* **6**, 967-978 (2004).
244 Shearer, J. & Graham, T. E. New perspectives on the storage and organization of muscle glycogen. *Canadian journal of applied physiology* **27**, 179-203 (2002).
245 Adeva-Andany, M. M., González-Lucán, M., Donapetry-García, C., Fernández-Fernández, C. & Ameneiros-Rodríguez, E. Glycogen metabolism in humans. *BBA clinical* **5**, 85-100 (2016).
246 Goldsmith, E., Sprang, S. & Fletterick, R. Structure of maltoheptaose by difference Fourier methods and a model for glycogen. *Journal of molecular biology* **156**, 411-427 (1982).
247 Brown, A. M. & Ransom, B. R. Astrocyte glycogen and brain energy metabolism. *Glia* **55**, 1263-1271 (2007).
248 Møller, M. S., Henriksen, A. & Svensson, B. Structure and function of α-glucan debranching enzymes. *Cellular and Molecular Life Sciences* **73**, 2619-2641 (2016).
249 Janeček, Š. & Zámocká, B. A new GH13 subfamily represented by the α-amylase from the halophilic archaeon Haloarcula hispanica. *Extremophiles* **24**, 207-217 (2020).
250 Stam, M. R., Danchin, E. G., Rancurel, C., Coutinho, P. M. & Henrissat, B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Engineering, Design and Selection* **19**, 555-562 (2006).
251 Mirmonsef, P. *et al.* Glycogen levels in undiluted genital fluid and their relationship to vaginal pH, estrogen, and progesterone. *PloS one* **11**, e0153553 (2016).
252 Greenwood, J. & Pickett, M. Salient features of Haemophilus vaginalis. *Journal of Clinical Microbiology* **9**, 200-204 (1979).
253 Spear, G. T. *et al.* Effect of pH on cleavage of glycogen by vaginal enzymes. *PLoS One* **10**, e0132646 (2015).
254 Martín Rosique, R. *et al.* Characterization of indigenous vaginal lactobacilli from healthy women as probiotic candidates. *International Microbiology* (2008).

255 Zhang, J., Li, L., Zhang, T. & Zhong, J. Characterization of a novel type of glycogen-degrading amylopullulanase from Lactobacillus crispatus. *Applied Microbiology and Biotechnology* **106**, 4053-4064 (2022).
256 Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **40**, W445-W451 (2012).
257 Zheng, J. *et al.* dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research*, gkad328 (2023).
258 Käll, L., Krogh, A. & Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic acids research* **35**, W429-W432 (2007).
259 Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39**, W29-W37 (2011).
260 Strength, G. Maximum Recovery Diluent (DM531).
261 Krisman, C. R. A method for the colorimetric estimation of glycogen with lodine. *Analytical biochemistry* **4**, 17-23 (1962).
262 Tester, R. & Al-Ghazzewi, F. H. Intrinsic and extrinsic carbohydrates in the vagina: a short review on vaginal glycogen. *International journal of biological macromolecules* **112**, 203-206 (2018).
263 Jenkins, D. J. *et al.* Identification and characterization of bacterial glycogen-degrading enzymes in the vaginal microbiome. *bioRxiv*, 2021.2007. 2019.452977 (2021).
264 Vodstrcil, L. A. *et al.* The influence of sexual activity on the vaginal microbiota and Gardnerella vaginalis clade diversity in young women. *PloS one* **12**, e0171856 (2017).
265 Fethers, K. A., Fairley, C. K., Hocking, J. S., Gurrin, L. C. & Bradshaw, C. S. Sexual risk factors and bacterial vaginosis: a systematic review and meta-analysis. *Clinical Infectious Diseases* **47**, 1426-1435 (2008).
266 Lambert, J. A., John, S., Sobel, J. D. & Akins, R. A. Longitudinal analysis of vaginal microbiome dynamics in women with recurrent bacterial vaginosis: recognition of the conversion process. *PloS one* **8**, e82599 (2013).
267 FIENBERG, R. & COHEN, R. B. Enzymes of glycogen metabolism in the squamous epithelium of the cervix: A histochemical study. *Obstetrics & Gynecology* **31**, 608-616 (1968).
268 De Seta, F. *et al.* The Vaginal Microbiome: III. The Vaginal Microbiome in Various Urogenital Disorders. *Journal of lower genital tract disease* **26**, 85-92, doi:10.1097/LGT.0000000000000645 (2022).
269 Vaneechoutte, M. *et al.* Emended description of Gardnerella vaginalis and description of Gardnerella leopoldii sp. nov., Gardnerella piotii sp. nov. and Gardnerella swidsinskii sp. nov., with delineation of 13 genomic species within the genus Gardnerella. *International journal of systematic and evolutionary microbiology* **69**, 679-687 (2019).
270 Johnston, C. D. *et al.* Systematic evasion of the restriction-modification barrier in bacteria. *Proceedings of the National Academy of Sciences* **116**, 11454-11459 (2019).
271 Kharsany, A., Hoosen, A. A. & Van den Ende, J. Antimicrobial susceptibilities of Gardnerella vaginalis. *Antimicrobial agents and chemotherapy* **37**, 2733-2735 (1993).
272 Schuetz, A. N. Antimicrobial resistance and susceptibility testing of anaerobic bacteria. *Clinical infectious diseases* **59**, 698-705 (2014).
273 Egervärn, M., Roos, S. & Lindmark, H. Identification and characterization of antibiotic resistance genes in Lactobacillus reuteri and Lactobacillus plantarum. *Journal of applied microbiology* **107**, 1658-1668 (2009).
274 Kay, M. A., He, C.-Y. & Chen, Z.-Y. A robust system for production of minicircle DNA vectors. *Nature biotechnology* **28**, 1287-1289 (2010).

275 Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Current opinion in microbiology* **8**, 466-472 (2005).
276 Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic acids research* **43**, D298-D299 (2015).
277 O'Callaghan, A. & van Sinderen, D. Bifidobacteria and their role as members of the human gut microbiota. *Frontiers in microbiology* **7**, 925 (2016).
278 Pritchard, J. R. *et al.* ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. *PLoS genetics* **10**, e1004782 (2014).
279 Park, M. J., Park, M. S. & Ji, G. E. Improvement of electroporation-mediated transformation efficiency for a Bifidobacterium strain to a reproducibly high level. *Journal of microbiological methods* **159**, 112-119 (2019).
280 Plamont, M.-A. *et al.* Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo. *Proceedings of the National Academy of Sciences* **113**, 497-502 (2016).
281 Arroyo-Olarte, R. D., Bravo Rodriguez, R. & Morales-Ríos, E. Genome editing in bacteria: CRISPR-Cas and beyond. *Microorganisms* **9**, 844 (2021).
282 Khot, P. D., Ko, D. L., Hackman, R. C. & Fredricks, D. N. Development and optimization of quantitative PCR for the diagnosis of invasive aspergillosis with bronchoalveolar lavage fluid. *BMC infectious diseases* **8**, 73 (2008).
283 Golob, J. L. *et al.* Stool microbiota at neutrophil recovery is predictive for severe acute graft vs host disease after hematopoietic cell transplantation. *Clinical Infectious Diseases* **65**, 1984-1991 (2017).
284 Callahan, B. J. *et al.* DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods* **13**, 581-583 (2016).
285 Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics* **11**, 538 (2010).
286 Minot, S. S., Barry, K. C., Kasman, C., Golob, J. L. & Willis, A. D. geneshot: gene-level metagenomics identifies genome islands associated with immunotherapy response. *Genome biology* **22**, 1-10 (2021).
287 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome research* **27**, 824-834 (2017).
288 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014).
289 Kelly, R. T. *et al.* Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Analytical chemistry* **78**, 7796-7801 (2006).
290 Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC bioinformatics* **14**, 60 (2013).
291 Swaney, D. L., Wenger, C. D. & Coon, J. J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *Journal of proteome research* **9**, 1323-1329 (2010).
292 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **8**, e61217 (2013).
293 Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature methods* **12**, 902-903 (2015).
294 Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic acids research* **36**, W5-W9 (2008).
295 Srinivasan, S. *et al.* Megasphaera lornae sp. nov., Megasphaera hutchinsoni sp. nov., and Megasphaera vaginalis sp. nov.: novel bacteria isolated from the female genital tract. *International Journal of Systematic and Evolutionary Microbiology* **71** (2021).

296 Fredricks, D. N., Fiedler, T. L., Thomas, K. K., Mitchell, C. M. & Marrazzo, J. M. Changes in vaginal bacterial concentrations with intravaginal metronidazole therapy for bacterial vaginosis as assessed by quantitative PCR. *Journal of clinical microbiology* **47**, 721-726 (2009).
297 O'Brien, V. P. *et al.* Helicobacter pylori chronic infection selects for effective colonizers of metaplastic glands. *Mbio* **14**, e03116-03122 (2023).
298 Lorian, V. (Lippincott Williams & Wilkins, 2015).
299 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345 (2009).
300 Douwe van Sinderen, M. V. *Bifidobacteria: Methods and Protocols*. 1 edn, (Humana, 2021).