

1 Vaginal metatranscriptome meta-analysis reveals  
2 functional BV subgroups and novel colonisation  
3 strategies

4 Scott J. Dos Santos<sup>1</sup>, Clara Copeland<sup>1</sup>, Jean M. Macklaim<sup>1</sup>,  
5 Gregor Reid<sup>2</sup>, Gregory B. Gloor<sup>1\*</sup>

6 <sup>1</sup>\*Department of Biochemistry, Western University, Middlesex Drive,  
7 London, N6G 2V4, Ontario, Canada.

8 <sup>2</sup>Lawson Health Research Institute, 268 Grosvenor Street, London, N6A  
9 4V2, Ontario, Canada.

10 \*Corresponding author(s). E-mail(s): [gloor@uwo.ca](mailto:gloor@uwo.ca);  
11 Contributing authors: [sdossa5@uwo.ca](mailto:sdossa5@uwo.ca); [ccopela6@uwo.ca](mailto:ccopela6@uwo.ca);  
12 [jean.macklaim@gmail.com](mailto:jean.macklaim@gmail.com); [gregor@uwo.ca](mailto:gregor@uwo.ca);

13 **Abstract**

14 **Background:** The application of ‘-omics’ technologies to study bacterial vagi-  
15 nosis (BV) has uncovered vast differences in composition and scale between the  
16 vaginal microbiomes of healthy and BV patients. Compared to amplicon sequenc-  
17 ing and shotgun metagenomic approaches focusing on a single or few species,  
18 investigating the transcriptome of the vaginal microbiome at a system-wide level  
19 can provide insight into the functions which are actively expressed and differential  
20 between states of health and disease.

21 **Results:** We conducted a meta-analysis of vaginal metatranscriptomes from  
22 three studies, split into exploratory ( $n = 42$ ) and validation ( $n = 297$ ) datasets,  
23 accounting for the compositional nature of sequencing data and differences in  
24 scale between healthy and BV microbiomes. Conducting differential expression  
25 analyses on the exploratory dataset, we identified a multitude of strategies  
26 employed by microbes associated with states of health and BV to evade host  
27 cationic antimicrobial peptides (CAMPs); putative mechanisms used by BV-  
28 associated species to resist and counteract the low vaginal pH; and potential  
29 approaches to disrupt vaginal epithelial integrity so as to establish sites for adhe-  
30 rence and biofilm formation. Moreover, we identified several distinct functional  
31 subgroups within the BV population, distinguished by genes involved in motility,

32 chemotaxis, biofilm formation and co-factor biosynthesis. After defining molecular  
33 states of health and BV in the validation dataset using KEGG orthology terms  
34 rather than community state types, differential expression analysis confirmed  
35 earlier observations regarding CAMP resistance and compromising epithelial barrier  
36 integrity in healthy and BV microbiomes, and also supported the existence  
37 of motile vs. non-motile subgroups in the BV population. These findings were  
38 independent of the enzyme classification system used (KEGG or EggNOG).

39 **Conclusions:** Our findings highlight a need to focus on functional rather than  
40 taxonomic differences when considering the role of microbiomes in disease and  
41 identify pathways for further research as potential BV treatment targets.

42 **Keywords:** ALDEx2, bacterial vaginosis, metatranscriptome, scale reliant inference,  
43 transcriptomics, VIRGO, vaginal microbiome.

## 44 1 Introduction

45 Bacterial vaginosis (BV) is a polymicrobial condition of poorly-understood aetiology  
46 and is the most common cause of vaginal concern worldwide [1]. Prevalence rates vary  
47 by country and are typically between 20-30 % [2], though reported rates exceed 50 %  
48 in some rural populations of developing nations [3]. Symptoms of the condition include  
49 a thin and potentially malodorous discharge, vaginal itching or irritation, pain on  
50 urination and/or sexual intercourse, and elevated vaginal pH; however many cases can  
51 be asymptomatic. Untreated, BV can have profound consequences, including higher  
52 acquisition rates of sexually-transmitted infections [4, 5] (including HIV and oncogenic  
53 HPV), in addition to a marked negative impact on patients' self-esteem, feelings of  
54 shame or anxiety, reduced intimacy and relationship problems [6] and an increased  
55 incidence of pre-term birth [7]. Recommended treatment with oral or intravaginal  
56 metronidazole or clindamycin often achieves symptom resolution in the short term,  
57 but is associated with high recurrence rates of 50-80 % within a year [8, 9].

58 Vaginal microbiome dysbiosis is a hallmark of BV and typically manifests as a  
59 paucity of *Lactobacillus* spp., with a corresponding overgrowth of various obligate  
60 anaerobic species [10]. Typically, total bacterial loads in BV are one to two orders of  
61 magnitude higher than in healthy patients [11], and species from the genera *Gard-*  
62 *nerella*, *Prevotella*, *Famyhessea* (formerly *Atopobium*[12]), *Megasphaera*, *Sneathia*,  
63 and *Mobiluncus*– among others– predominate. Landmark studies of the vaginal micro-  
64 biome using amplicon sequencing defined five canonical ‘community state types’  
65 (CSTs), with CST IV showing a significant correlation with BV [13, 14]. This find-  
66 ing has since been extensively replicated, including studies employing alternative  
67 molecular barcoding genes [15, 16] and shotgun metagenomic methods [17].

68 Characterising microbial gene expression in the vaginal microbiome remains an  
69 uncommon approach, largely due to the costs associated with obtaining sufficient  
70 sequencing depth for a large number of samples and the difficulty of analysis because of  
71 the very different species present in the two conditions. However, defining the vaginal  
72 metatranscriptome via RNA-seq enables a functional assessment of gene expression  
73 within the vaginal niche, in addition to taxonomic characterisation of the community.

74 Identifying differences in actively expressed genes and pathways at the systems level  
75 between states of health and BV could further our understanding of the complex aeti-  
76 ology of BV or yield novel treatment targets. The initial taxonomic [13] and functional  
77 [18] explorations of the vaginal microbiome by culture-independent means represented  
78 a huge advance in BV research; however, while replication of the taxonomic differences  
79 in BV is reassuring, such studies have not yet advanced currently available treatments;  
80 although vaginal microbiome ‘transplantation’ has been explored [19]. Meta-‘omics’  
81 approaches are also limited by a lack of standardised methods, and findings can some-  
82 times be difficult to reconcile or replicate between datasets—especially when different  
83 sample collection methods, bioinformatic pipelines, and statistical analyses have been  
84 used. Moreover, small study populations and inappropriate analyses in such studies  
85 may lead to false-positive findings which are spuriously correlated with disease states,  
86 rather than truly differential [20].

87 One final issue with metatranscriptomic analysis is marked asymmetry between  
88 datasets. The vaginal microbiome is known to be exceedingly asymmetric with bacte-  
89 rial loads differing by up to two orders of magnitude [11] in addition to a large gene  
90 content difference between conditions [21–23]. These differences cause an asymmetry in  
91 the location of ‘housekeeping’ functions during the data analysis that is not accounted  
92 for by the existing standard normalisation methods. This often results in a large pro-  
93 portion of all genes in the dataset to be incorrectly called as significantly different  
94 between experimental conditions [24, 25]. The issue of asymmetry can be overcome  
95 by incorporating a new statistical approach termed scale reliant inference (SRI) [20].  
96 This new statistical approach included in the updated ALDEx2 R package [24, 26],  
97 allows researchers to build a Bayesian posterior model of the data that enables robust  
98 and reproducible inference even in the face of extreme data asymmetry.

99 We carried out a meta-analysis of three publicly available vaginal metatranscrip-  
100 tomic datasets. We identified expressed genes and functions which differentiate BV  
101 and healthy vaginal microbiomes, and which are reproducibly distinct across differ-  
102 ent study populations. Collecting and re-processing three RNA-seq datasets using a  
103 standardised workflow including annotation with the VIRGO pipeline [22], we found  
104 that *Lactobacillus* spp. and BV-associated taxa employ different strategies to evade  
105 host anti-microbial defense systems such as cationic antimicrobial peptides (CAMPs).  
106 Furthermore, we identified putative mechanisms for BV-associated species to counter-  
107 act the low vaginal pH and compromise vaginal epithelial barrier integrity. We also  
108 observed at least two distinct subgroups of BV-associated metatranscriptome profiles  
109 differing by expression of motility, chemotaxis and co-factor biosynthesis genes. These  
110 findings were replicated in a large, independent dataset; various strategies for CAMP  
111 resistance and the existence of motile vs. non-motile BV sub-populations were evident  
112 among differentially expressed pathways. Overall, our findings highlight potential tar-  
113 gets for novel therapeutic strategies in the treatment of BV, opening several avenues  
114 of investigation for future work made possible only by incorporating scale uncertainty  
115 in our analytical approach.

## 116 2 Results

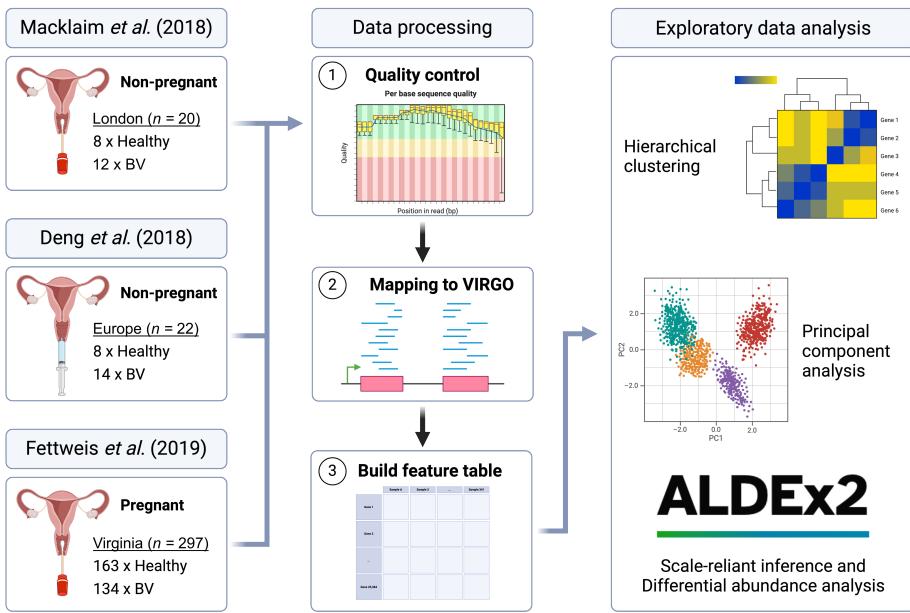
### 117 2.1 Study cohort and final dataset

118 Five vaginal metatranscriptomic datasets were identified for meta-analysis, three of  
119 which were collected for analysis (Figure 1) and are described by [7, 27, 28]. Two  
120 other datasets exist but were not used. The first [23] was an extended time-series  
121 survey of 39 participants and was not included as its size would not significantly  
122 increment our power. The second [29] interrogated the vaginal metatranscriptome and  
123 host response during severe SARS-CoV-2 infection; raw data for this study were not  
124 available at the accession number given in the article. Owing to this, and the potential  
125 confounding introduced by viral infection and intensive care admission, this dataset  
126 was also excluded.

127 The three study populations analysed exhibited considerable variation, suggesting  
128 that any findings reported would be general and not specific to any one dataset. Mack-  
129 laim *et al.* (London dataset [27]) studied a convenience sample of 20 non-pregnant,  
130 volunteers from London, Ontario (Canada) with the aim of identifying functional dif-  
131 ferences between states of health and BV. We used a subset of 22 non-pregnant,  
132 symptomatic BV patient samples from Deng *et al.* (Europe dataset [28]) which orig-  
133 inate from German BV patients who underwent metronidazole treatment. Samples  
134 from this study classed as ‘BV’ are pre-treatment samples, whereas samples classed  
135 as ‘healthy’ are post-treatment samples. Demographically, 84% of the women were  
136 Caucasian and the remainder had some African ancestry [30]. Fettweis *et al.* (Virginia  
137 dataset [7]) collected 297 vaginal metatranscriptome samples from pregnant volun-  
138 teers as part of the MOMS-PI dataset through the integrative Human Microbiome  
139 Project in a USA-based, prospective, longitudinal cohort study aiming to ascertain  
140 the effect of the vaginal microbiome on preterm birth. Samples from London and  
141 Europe datasets were obtained primarily from Caucasian women, whilst those from the  
142 Virginia dataset were primarily from African-American women. A total of 341 meta-  
143 transcriptome samples were included in the meta-analysis; median post-processing  
144 read counts were relatively similar across London and Europe datasets but higher in  
145 the Virginia dataset (Table 1). The smaller London and Europe datasets were used as  
146 an initial ‘exploratory’ dataset, while the larger Virginia dataset was used to validate  
147 observations made in the latter.

**Table 1** Pre-and post-processing read counts for the three datasets included in the meta-analysis. H, healthy; BV, bacterial vaginosis

| Dataset      | No. samples |     |       | Median no. reads |                 |
|--------------|-------------|-----|-------|------------------|-----------------|
|              | H           | BV  | Total | Raw              | Post-processing |
| London [27]  | 8           | 12  | 22    | 14,479,786       | 11,107,052      |
| Europe [28]  | 8           | 14  | 22    | 16,584,536       | 8,096,724       |
| Virginia [7] | 163         | 134 | 297   | 46,368,206       | 19,967,915      |



**Fig. 1 Study workflow:** Raw FASTQ files were downloaded from their respective repositories and were trimmed for quality and length. Reads mapping to GRCh38 and T2T human reference genomes were discarded as presumptive human contamination; the remainder were mapped to v138.1 of the SILVA rRNA database, and any aligned reads were also discarded. The remaining reads were mapped against the VIRGO catalogue of ~1 million non-redundant, microbial genes from the human vagina [22]. Feature tables for each dataset were constructed separately and batch-corrected [31] when used in combination. These data tables served as input for all downstream analyses using the scale reliant inference (SRI) approach to normalisation [20] and the expected values of Bayesian posterior models of the data were used as starting points for all exploratory data analysis and statistical methods.

## 148 2.2 Applying scale to healthy and BV metatranscriptomes

149 The vaginal microbiome of healthy and BV subjects differs greatly in both composition  
 150 and scale; the BV microbiome is both more complex and has about two orders of  
 151 magnitude more bacteria per collected volume [11]. Commonly used data normalisa-  
 152 tions assume that the majority of the parts being sampled are relatively invariant in  
 153 the sampled environment. To account for this asymmetry, we used a new statistical  
 154 model for high throughput sequencing datasets called scale-reliant inference (SRI) [20].  
 155 SRI allows the explicit modelling of both the amount of variation in the underlying  
 156 datasets and the proper location of marker functions, and this feature is incorporated  
 157 into the latest version of ALDEX2 [24, 32].

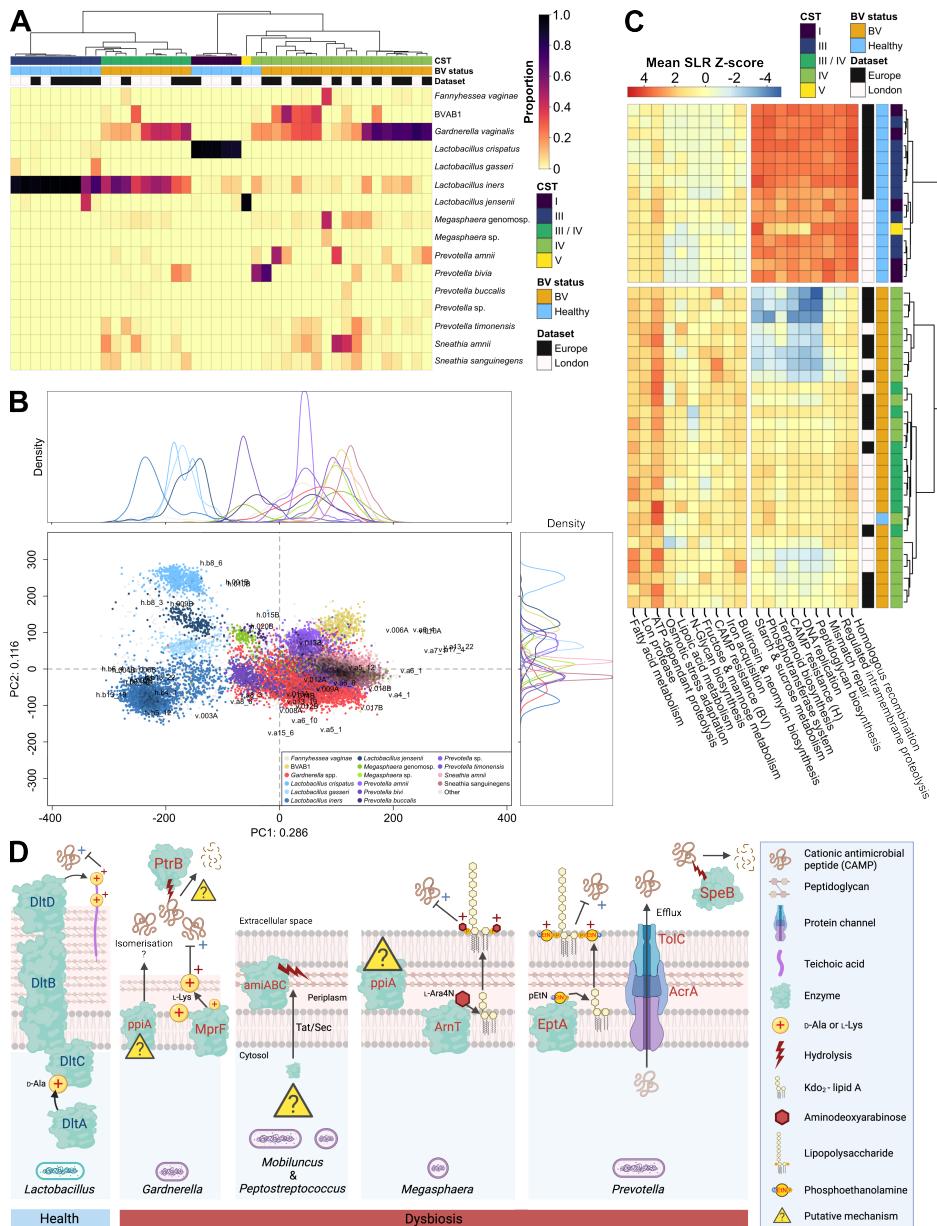
158 Applying SRI through ALDEX2 generates a full Bayesian posterior model of the  
 159 underlying environment from the observed table of counts by incorporating uncertainty  
 160 in both the observed count table and the count normalisation (i.e. scaled log-ratio  
 161 transformation). This approach prevents converging on a precise but incorrect set of  
 162 differentially abundant features [20]. The model is built by explicitly acknowledging  
 163 two *a priori* assumptions about the table of counts. Firstly, that the expression of  
 164 genes encoding standard housekeeping functions required by all organisms should be

165 relatively invariant between groups; that is, the location of housekeeping functions  
166 should exhibit no difference between groups. We found that the naïve estimate of scale  
167 produced by ALDEx2 had an offset of about 8-fold in scale that resulted in the location  
168 of housekeeping functions being off the location of no difference. We found that adding  
169 a 1.15-fold offset to the scale parameter was able to centre the housekeeping functions  
170 (Methods, and Supplementary Figure S1). This is biologically intuitive in that the  
171 small difference in scale needed to centre the housekeeping functions indicates that  
172 they can be assumed to be relatively invariant. We therefore applied this 15 % offset  
173 as part of all differential expression analyses conducted with ALDEx2. The second  
174 assumption is that the amount of variation estimated by the statistical normalisation  
175 used by ALDEx2, the centred log-ratio normalisation, was lower than optimal. Nixon  
176 et al. [20] showed that the measured amount of variation was low for essentially all  
177 high-throughput sequencing datasets and that including some uncertainty was always  
178 optimal. As such, we included scale uncertainty of 0.5 standard deviations, which is  
179 the minimum recommended for this approach [24].

### 180 2.3 Vaginal metatranscriptomics captures the canonical 181 community state types (CSTs)

182 Initial analysis of the London and Europe datasets recapitulated numerous prior  
183 studies on compositional differences between healthy and BV microbiomes. Hierarchi-  
184 cal clustering of species-level metatranscriptome profiles exhibited clusters strongly  
185 reminiscent of the canonical CSTs first reported by Ravel and colleagues [13, 14]  
186 (Figure 2A). We identified three *Lactobacillus*-dominant clusters (*L. crispatus*, *L.*  
187 *iners*, *L. jensenii*), a cluster of anaerobic species comprising various proportions of  
188 the expressed transcriptome, and a final cluster characterised by a mix of *L. iners* and  
189 several BV-associated species. Although many samples in this study were obtained  
190 from otherwise healthy volunteers (e.g. many of the London and Virginia samples),  
191 the terms ‘BV patient’ and ‘healthy patient’ are used hereafter to distinguish between  
192 groups when performing differential expression and other analyses.

193 Metatranscriptome reads that mapped to the gene and organism level of healthy  
194 and BV patients were readily observed to group by species on a compositional principal  
195 component biplot [33] (hereafter PCA plot). Health-associated *Lactobacillus* species  
196 were cleanly separated across the first principal component from the various anaerobic  
197 species associated with BV. Likewise, BV status was easily distinguished for all sam-  
198 ples across PC1 (Figure 2B). We further observed that genes assigned to *Gardnerella*  
199 *vaginalis* and either phylogroup of *Megasphaera* did not form the discrete clusters  
200 observed in other species and were instead dispersed across PC1. While *Megasphaera*  
201 phylogroups have not been formally distinguished as individual taxa [34], the genus,  
202 *Gardnerella*, has recently been delineated into at least 13 genomic species, six of  
203 which have official standing in nomenclature [35, 36] (though they are not presently  
204 included in the VIRGO database taxonomy).



**Fig. 2 Health and BV-associated species of the vaginal microbiome employ differing strategies for host CAMP resistance:** Vaginal microbiome composition (gene-level mapping, coloured by species) within London and Europe datasets was assessed by **(A)** hierarchical clustering of metatranscriptomes and **(B)** compositional principal component analysis (post-SRI); only species represented by  $\geq 75$  genes are shown. Marginal density plots show gene distribution by species across PC1 and PC2. **(C)** Gene assignments for each sample were grouped by KEGG orthology (KO) term regardless of species and median scaled log-ratio (SLR) values calculated by ALDEx2 were aggregated into a Z-score of the mean SLR value per KEGG pathway. Data for the top 10 differential pathways in both health and BV are plotted (all post-SRI absolute effect sizes and expected false discovery rates are  $>1$  and  $<0.01$ , respectively). **(D)** Known and putative mechanisms of CAMP resistance within vaginal metatranscriptomes of the London/Europe datasets (created with BioRender).

205 **2.4 Resistance mechanisms against host cationic antimicrobial  
206 peptides differ between healthy and BV  
207 metatranscriptomes**

208 Differential expression analysis between healthy and BV metatranscriptomes with  
209 ALDEx2 demonstrated significant differences in expression for a multitude of path-  
210 ways, many of which reflect previously reported differences between the two conditions  
211 (Figure 2 & Supplementary Figure S2). For example, expression of genes involved  
212 in butanoate metabolism [37] and sialidase activity [38]– hallmarks of BV and some  
213 species of *Gardnerella*, respectively– were elevated among BV metatranscriptomes,  
214 while peptidoglycan biosynthesis genes were highly expressed among healthy samples  
215 dominated by Gram-positive lactobacilli.

216 One of the most striking newly observed differences between healthy vs. BV sam-  
217 ples was observed among functions related to resistance to host-encoded cationic  
218 antimicrobial peptides (CAMPs) (Figure 2C). Five KEGG orthology (KO) terms  
219 assigned to this pathway were defined as differential (four over-expressed in healthy  
220 samples, one over-expressed in BV samples) and absolute effect sizes for these KO  
221 terms were large (range 1.02 - 1.77). Differential CAMP resistance KO terms in healthy  
222 samples corresponded to all four genes of the DltABCD operon– a ubiquitous pathway  
223 among Gram-positive bacteria [39]– which enables D-alanylation of peptidoglycan-  
224 anchored teichoic acids, resulting in a reduction of the negative surface charge. The  
225 26 VIRGO genes assigned to these four KO terms primarily have taxonomic assign-  
226 ments to lactobacilli in the VIRGO database and a majority of these reads originated  
227 from non-*L. iners* lactobacilli (Supplementary Table S1). Notably, these KO terms  
228 were initially annotated in KEGG as ‘*Staphylococcus aureus* infection’.

229 Among metatranscriptomes from the BV group, only a single KO term for CAMP  
230 resistance was defined as differentially abundant, corresponding to the periplasmic  
231 N-acetylmuramoyl-L-alanine amidases AmiABC, which are critical for cleaving pepti-  
232 doglycan during cell division and separation–particularly under acidic conditions [40].  
233 However, we were unable to find any reports corroborating a link between any of these  
234 amidases with CAMP resistance. Accordingly, we speculate that it may be a misas-  
235 signment in KEGG which also reflects the higher proportion of Gram-negative species  
236 within BV microbiomes.

237 Given the importance of host-microbe interactions during microbial colonisation,  
238 we next investigated the expression of all VIRGO entries present in the KEGG  
239 CAMP resistance pathway (map01503) among BV samples in the London and Europe  
240 datasets. From this, we identified a further eight genes in this pathway present in the  
241 London and Europe datasets (Figure 2D), with functions including those that decrease  
242 the net negative surface charge for repulsion of CAMPs, efflux of CAMPs, and direct  
243 CAMP hydrolysis; however, we were unable to find confirmatory studies linking two  
244 genes (PpiA/PtrB) to CAMP resistance. In multiple instances, a single genus was  
245 responsible for the majority of genes assigned to a given KO term. *Gardnerella* spp.,  
246 for example, contributed 17 of the 34 genes with known taxonomy which were assigned  
247 to the phosphatidylglycerol lysyltransferase, MprF (KO14205; [41]), while 10 of 12  
248 genes assigned to the lipid A ethanolaminephosphotransferase, EptA (K03760; [42]),

249 originated from several species of *Prevotella*. In both cases, other species typically  
250 contributed only a single gene to both KO terms.

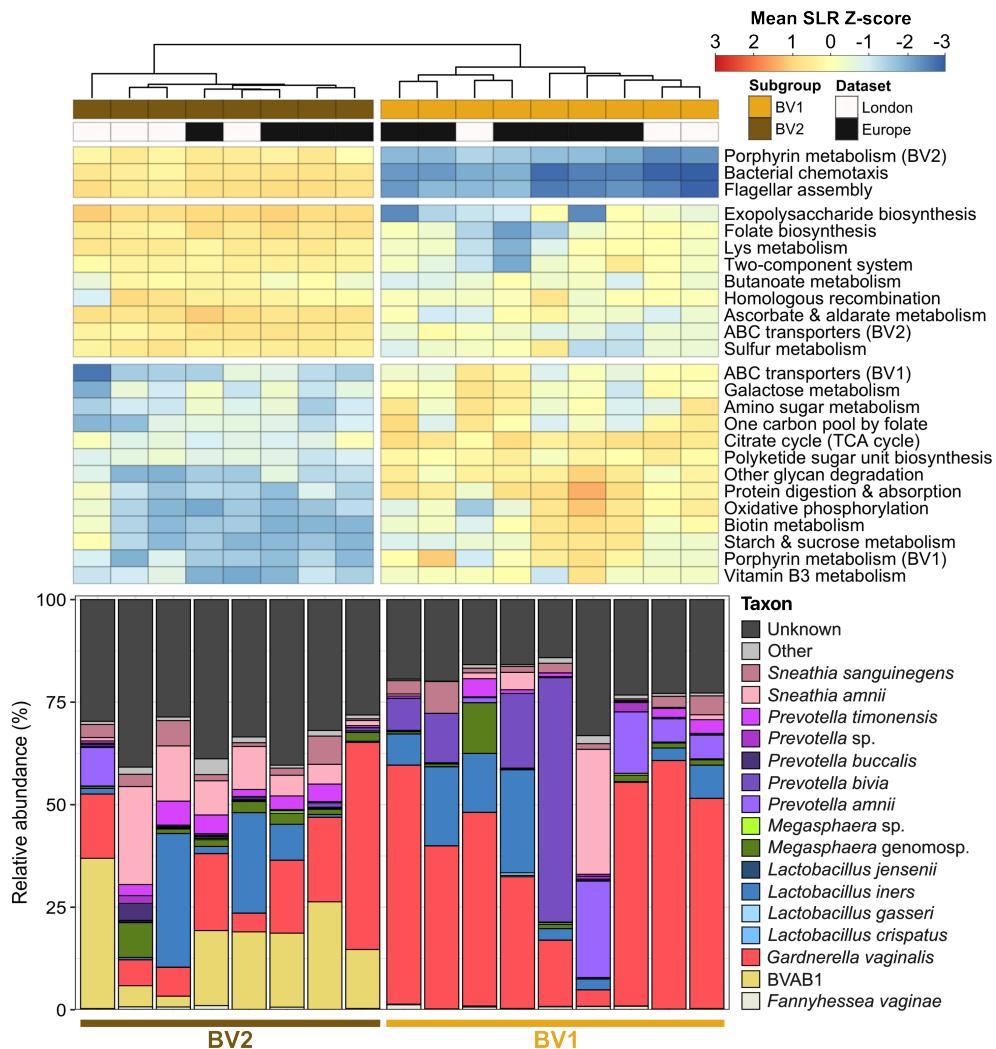
251 During our search for CAMP resistance functions, we also noted the presence of  
252 several VIRGO entries assigned to the KO term, ‘African trypanosomiasis’ (K01354),  
253 which was unexpected considering the vaginal niche. This KO term represents the pro-  
254 lyl oligopeptidase, PtrB, which degrades peptides smaller than ~30 amino acids via  
255 hydrolysis at the carboxyl side of basic amino acids [43]. A total of 58 genes in the  
256 London and Europe datasets were assigned to this KO term: 31 were of unknown tax-  
257 onomy, 23 were assigned to *G. vaginalis*, 2 to *Bifidobacterium longum*, 1 to *Mobiluncus*  
258 *mulieris* and 1 to *Streptococcus pneumoniae*. Although we could not find literature  
259 supporting a role for PtrB in CAMP resistance, given the abundance of basic amino  
260 acids in CAMPs, we hypothesise that such oligopeptidase activity might represent an  
261 additional strategy used by BV-associated taxa for evading host immune responses.

262 The potential misassignment of multiple KO terms for CAMP resistance functions  
263 led us to screen the list of differential KOs for similarly suspicious functions, and con-  
264 firm that KO terms with no assigned function were truly unknown. Two KO terms  
265 flagged due to an assignment of ‘Epithelial cell signaling in *Helicobacter pylori* infec-  
266 tion’ corresponded to the acid-activated urea channel, UreI (K03191), and the putative  
267 U32-family collagenase, PrtC (K08303). Genes contributing to the latter function were  
268 abundant among several BV-associated taxa, including several species of *Prevotella*  
269 and *Sneathia*, *Atopobium vaginae*, and BVAB1. Both terms were over-represented  
270 among BV metatranscriptomes (Supplementary Figure S2) and may bear relevance to  
271 BV, given the involvement of UreI in pH resistance among clinically important human  
272 pathogens [44] and the identification of PrtC in several vaginal *Prevotella* and *Porphy-*  
273 *romonas* and isolates capable of type I collagen degradation [45, 46]. Likewise three  
274 terms initially lacking functional information, were respectively found to correspond  
275 to the periplasmic iron-binding shuttle, FbpA (K02012), the non-heme ferritin iron  
276 storage protein, FtnA (K02217), and the acid-inducible, cytoplasmic membrane-bound  
277 iron transporter, EfeU (K07243). Iron acquisition is directly linked to colonisation  
278 and virulence in many important human pathogens [47]; therefore, the differences in  
279 iron requirements between BV-associated taxa and health-associated lactobacilli may  
280 present a therapeutic opportunity [48].

## 281 2.5 Subgroups of BV metatranscriptomes are delineated by 282 expression of genes involved in motility, chemotaxis, and 283 co-factor biosynthesis

284 Hierarchical clustering of samples based on all differentially abundant pathways  
285 revealed a clear split among BV metatranscriptomes, suggesting the existence of at  
286 least two functional subgroups within the BV population (Supplementary Figure S2).  
287 A comparison of the two main BV subgroups demonstrated that functions involved  
288 in motility (flagellar assembly), heme and cobalamin/vitamin B<sub>12</sub> biosynthesis (por-  
289 phyrin metabolism) and chemotaxis were the main drivers of this difference. Based  
290 on the taxonomic composition of these samples, the presence or absence of BVAB1  
291 (BV-associated bacterium-1; recently reclassified as “*Candidatus Lachnocurva vagi-*  
292 *nae*” [49]) appeared to explain these differences almost entirely (Figure 3, bottom),

similar to the taxonomic split between previously recognised CSTs IV-A and IV-B [14]. However, repeating this analysis after removing all genes assigned to BVAB1 by VIRGO, the major functional differences persisted, suggesting that these differences are system-wide and not solely related to the presence of BVAB1 (Supplementary Figure S3).



**Fig. 3 Distinct BV subgroups differ by expression of motility, chemotaxis and co-factor biosynthesis genes:** The top 25 differentially abundant pathways among samples from the two main BV subgroups in London/Europe datasets are expressed as a Z-score calculated from the mean scaled log-ratio (SLR) values of all KO terms assigned to a given pathway (individual SLR values represent a median across 128 Monte-Carlo instances). All post-SRI absolute effect sizes and false discovery rates calculated by ALDEEx2 were  $>1$  and  $<0.01$ , respectively). Stacked bars below represent the proportional taxonomic composition of the expressed transcriptome. Taxa represented by  $<75$  genes are collapsed into 'Other'.

When considering the taxonomy of reads associated with the KO terms for ‘bacterial chemotaxis’, reads originating from BVAB1 comprise the majority. The eight KO terms over-represented in BV2 in the London/Europe datasets corresponded to various genes within the Che signalling pathway [50], including the sensor histidine kinase (CheA), signal transducers (CheW/V), regulatory methyltransferases/demethylases (CheR/B), and methyl-accepting chemoreceptors (mcp). The KO term for the Che response regulator (CheY; activates the flagellar motor switch upon phosphorylation by CheA) was detected at higher proportions in BV2 samples than in BV1 samples, but this difference was marginally below our effect size threshold of significance (effect = -0.897).

Likewise, reads assigned to the KO terms for ‘flagellar assembly’ mostly originated from BVAB1, as well as *Lactobacillus iners*, *Mobiluncus curtisii* and *Mobiluncus mulieris*. A variety of flagellar components were represented by these KO terms [51], including the major flagellar filament subunit (fliC), rotary protein (motA), hook proteins (flgE/K/L and fliD/K), motor switch proteins (fliM/G), regulatory factors (flgM and flaG), type III protein export proteins (flhA and fliH/I), and C- and MS-ring proteins (fliF/N).

The presence of a KO term corresponding to ‘exopolysaccharide biosynthesis’ among the differential pathways was particularly notable, given its function as poly- $\beta$ -1,6-acetyl-D-glucosamine synthase (PgaC/IcaA; K11936 [52]). This polysaccharide is a key component of biofilm extracellular polymeric substance produced by both Gram-negative and Gram-positive species, and has been implicated in initial cell-surface attachment and later lateral agglomeration during biofilm development and expansion [53]. Biofilm involvement is a prominent feature of BV [54] and vaginal epithelial cells with adherent biofilms (clue cells) are a diagnostic feature during microscopic analysis of Gram-stained, wet-mount vaginal smears [55]. Once more, reads belonging to this KO term were strongly associated with BVAB1 (and to a far lesser extent, *Prevotella amnii*).

Reads corresponding to the KO terms for ‘porphyrin metabolism’ showed no association with BVAB1. Of the three KO terms over-represented in the BV2 subgroup, almost all reads assigned to the cobaltochelatase, CobN, originated from *Prevotella timonensis*, while the majority of reads assigned to the cobalamin adenosyltransferase, PduO, and the porphobilinogen synthase, HemB, were of unknown taxonomy. The KO term elevated in BV1 metatranscriptomes corresponds to uroporphyrinogen III synthase (hemD), and reads assigned to this term were largely contributed by *Prevotella bivia* and *Prevotella amnii*. HemB/D are involved in both heme biosynthesis via  $\delta$ -aminolevulinate and *de novo* production of the biologically-active (co-factor) form of vitamin B<sub>12</sub>, adenosylcobalamin. PduO plays a role in the cobalamin salvage pathway, converting exogenous cobamide taken up via the BtuBCDF transport system into adenosylcobamide, while CobN oversees the incorporation of cobalt into the tetrapyrrole ring of hydrogenobyrinic acid a,c diamide as part of the *de novo* biosynthesis of adenosylcobalamin [56].

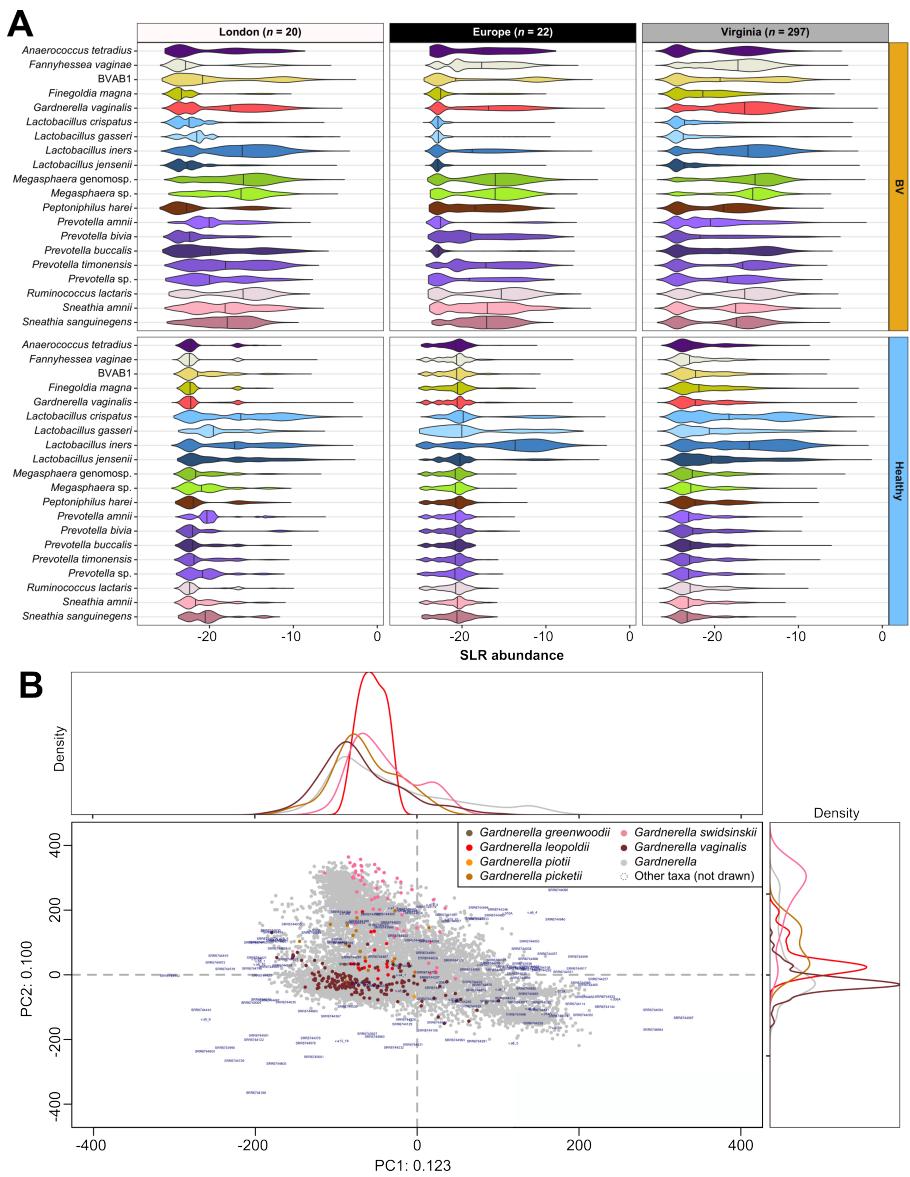
The existence of distinct subgroups of BV – particularly those delineated by expression of genes involved in motility and biofilm formation – led us to speculate that

342 symptomatic vs. asymptomatic BV may be linked to one subgroup or the other. Unfor-  
343 tunately, clinical metadata for the London/Europe datasets was limited to Nugent  
344 scores and vaginal pH, which precluded this analysis. However, extensive self-reported  
345 data on BV symptoms and vaginal pH was collected for the Virginia dataset as part of  
346 the MOMS-PI study on pre-term birth [7]. Therefore, we sought to replicate all of the  
347 above observations in the Virginia dataset and determine if there was any association  
348 of BV symptoms with any identified subgroups.

349 **2.6 Pregnancy-related compositional changes within vaginal  
350 metatranscriptomes and lack of functional correlation with  
351 pre-term birth**

352 Prior to validating any previous observations, we investigated if pregnancy introduced  
353 major taxonomic shifts which could complicate comparisons between London/Europe  
354 (non-pregnant) and Virginia (pregnant) datasets. We determined the CSTs present  
355 within the Virginia dataset and proceeded to define ‘molecular’ states of health and  
356 BV based on hierarchical clustering of metatranscriptomes at the functional (KO)  
357 level (Supplementary Figure S4). Virginia metatranscriptomes separated into two clear  
358 groups– the first, characterised by higher vaginal pH, a preponderance of CST IV  
359 samples and the absence of CST I samples, was taken to represent molecular BV. The  
360 other largely contained samples belonging to CSTs I, III, and V and was defined as  
361 molecular health. Notably, there appeared to be no pattern linking CST or molecular  
362 heath/BV with self-reported abnormal vaginal discharge, odor, or irritation.

363 After processing the three datasets together, we compared the post-SRI SLR val-  
364 ues returned by ALDEx2 for all features from the species represented by at least 75  
365 genes, stratified by dataset and BV status (Figure 4A). *Lactobacillus* depletion and  
366 anaerobic overgrowth in BV metatranscriptomes of all datasets was evident, and the  
367 effect of recent antimicrobial therapy for symptomatic BV within the Europe dataset  
368 was clear: metatranscriptomes comprised a lower proportion of transcripts from non-  
369 *L. iners* lactobacilli and a greater proportion of transcripts from BV-associated taxa  
370 even among healthy samples. However the most striking differences between pregnancy  
371 states lay in the proportion of transcripts from *Sneathia amnii*, *Sneathia sanguinegens*,  
372 and *Gardnerella vaginalis* in BV samples. Expression of transcripts assigned to the  
373 two *Sneathia* species– particularly *S. sanguinegens*– was markedly lower in the Vir-  
374 ginia dataset than in London or Europe, with proportionally more samples containing  
375 only a few reads from these taxa. Concomitantly, *Gardnerella vaginalis* transcripts  
376 and to a lesser degree, those of *Fannyhessea vaginae*, were detected at higher levels  
377 within the Virginia dataset than either non-pregnant dataset.



**Fig. 4 Subtle changes in vaginal microbiome composition and delineation of *Gardnerella* species within the Virginia validation dataset:** Vaginal metatranscriptomes from all datasets in this study underwent filtering and scaled log-ratio (SLR) transformation with ALDEEx2. **(A)** SLR values for all genes grouped by species were visualised as violin plots, stratified by dataset and BV status. Vertical lines on violins indicate taxa medians. **(B)** Genes unique to each of the four *Gardnerella* species with standing in nomenclature were identified and highlighted on a PCA plot of the same SLR-transformed data used in (A). Genes from all other species are not visible, but were included in the ordination. Density plots show gene distribution across PCs.

378 **2.7 Discrete clusters of reads from named *Gardnerella* species  
379 are detectable within metatranscriptomes of all datasets**

380 The prevalence of *Gardnerella vaginalis* in the Virginia dataset prompted us to ask  
381 whether the six named *Gardnerella* species with current standing in nomenclature[35,  
382 36] could be identified and delineated within the broadly distributed group of reads  
383 assigned to the genus, *Gardnerella*, in the VIRGO database. Using publicly avail-  
384 able, complete genome sequences (see methods), we identified genes unique to *G.*  
385 *greenwoodii*, *G. leopoldii*, *G. pickettii*, *G. piotii*, *G. swidsinskii* and *G. vaginalis* with  
386 Paneroo [57] and aligned sequences in the combined London/Europe/Virginia dataset  
387 which were labelled as *Gardnerella vaginalis* to these genes. Re-labelling the gene  
388 taxonomy according to these BLAST hits (conservatively filtered by E-value and align-  
389 ment coverage), we then visualised reads assigned to the genus, *Gardnerella*, within  
390 the ordination of all features in the combined dataset (Figure 4B).

391 Separation of named *Gardnerella* species was seen across all features assigned to  
392 the genus, particularly those belonging to *G. vaginalis* and *G. swidsinskii*. Similar  
393 separation at the species level was also seen when considering the London/Europe and  
394 Virginia datasets, separately (Supplementary Figure S5). Curiously, we observed an  
395 absence of genes unique to the named species for samples with positive PC1 scores,  
396 despite the discrete clustering of several hundred *Gardnerella* features. Such clusters  
397 may represent the multiple, unnamed genome species of *Gardnerella*, for which there  
398 are a paucity of representative complete genome sequences.

399 **2.8 Key findings from discovery datasets can be replicated  
400 in the validation dataset**

401 Having shown that the vaginal metatranscriptome datasets were comparable despite  
402 differing pregnancy status, we asked whether the discoveries made in the London/Eu-  
403 rope datasets could be replicated in the Virginia dataset.

404 Using the previously defined groups of molecular health and BV for SLR-  
405 transformation and data asymmetry correction, we visualised functional metatran-  
406 scriptomes via PCA (Supplementary Figure S6). Following scale simulation with the  
407 same parameters as before, typical housekeeping KO terms were again centred around  
408 the location of no difference. Metatranscriptomes clearly separated across PC1 based  
409 on molecular health vs. molecular BV status; however, we also observed an evident  
410 split within the molecular health group, driven by dominance of either *L. iners* (health  
411 subgroup 1) or *L. crispatus* (health subgroup 2). For consistency with earlier analy-  
412 ses, subgroups of molecular health were treated as a single group for comparison with  
413 molecular BV.

414 Formal differential expression analysis mirrored the results obtained from the Lon-  
415 don/Europe datasets. CAMP resistance was once again among the most differential  
416 pathways, with the DltABCD operon highly expressed among healthy metatranscrip-  
417 tomes (Figure 5A). As before, genes representing the four Dlt KO terms were primarily  
418 assigned to *Lactobacillus* spp. in VIRGO and a majority of these reads were assigned  
419 to non-*L. iners* lactobacilli (Supplementary Table S1). All CAMP resistance KOs  
420 summarised in Figure 2D were also present in the Virginia dataset, and the same

421 species-specific associations were observed (e.g. MprF with *Gardnerella vaginalis*).  
 422 Several additional KOs belonging to the KEGG ‘CAMP resistance’ pathway were  
 423 also detected, including several two-component regulatory systems (PhoPQ, CpxAR,  
 424 BasRS), although these were represented by <100 reads each (likely a result of the  
 425 greater sequencing depth in this dataset).



**Fig. 5 Prior observations on CAMP resistance and BV subgroups can be replicated in a large-scale validation dataset:** Filtered metatranscriptome feature tables underwent scaled log-ratio (SLR) transformation with ALDEx2. Gene assignments for each sample were grouped by KEGG orthology (KO) term regardless of species and median SLR values from ALDEx2 were aggregated into a Z-score of the mean SLR value per KEGG pathway. These data were plotted for three comparisons of Virginia dataset samples: (A) molecular health vs. molecular BV; (B) molecular BV subgroups; (C) molecular health subgroups. For (A) and (C), the top 10 most differential pathways for health and BV are shown. All absolute effect sizes and expected false-discovery rates are  $\geq 1$ ,  $< 1\%$ , respectively.

426 Almost all KO terms over-represented among BV metatranscriptomes in the  
427 discovery dataset were also differentially expressed in the validation dataset. KOs  
428 corresponding to butanoate metabolism, the U32-family putative type I collagenase  
429 (PrtC), the iron transport and storage proteins (EfeU, FbpA, FtnA), but not the  
430 acid-activated urea channel (UreI) exhibited significantly higher expression among BV  
431 metatranscriptomes. Unlike the London/Europe dataset, the single CAMP resistance  
432 KO term over-represented among BV metatranscriptomes corresponded to the phos-  
433 phatidylglycerol lysyltransferase, MprF (absolute effect size = -1.23), rather than the  
434 N-acetylmuramoyl-L-alanine amidases, AmiABC (Supplementary Figure S7). Assess-  
435 ment of the signature functions defining health and BV groups overlapped well with  
436 differential expression analyses and discriminated these groups with high accuracy  
437 (AUC = 0.997, Supplementary Figure S8A-C):CAMP resistance was a key functional  
438 signature of healthy metatranscriptomes, whereas butanoate metabolism and TCA  
439 cycle genes were among the characteristic features of BV.

440 As in the discovery dataset, the initial clustering of Virginia metatranscriptomes  
441 at the functional level revealed multiple sub-populations of BV metatranscriptomes.  
442 Focusing on the two largest subgroups (BV1,  $n = 57$ ; BV2,  $n = 56$ ), we tested  
443 whether the same differences between BV subgroups identified in the London/Europe  
444 dataset could be observed among Virginia BV samples. Differences in expression of  
445 genes involved in flagellar assembly, chemotaxis, and exopolysaccharide biosynthesis  
446 were again the among the largest drivers of subgroup separation (Figure 5B), with  
447 flagellar assembly identified as the sole contributor to the functional signature of one  
448 subgroup (AUC = 0.994, Supplementary Figure S8D-F). All KO terms from flagel-  
449 lar assembly and chemotaxis pathways identified as differential within the discovery  
450 dataset were again differentially expressed between BV subgroups in the validation  
451 dataset. Additionally, several KO terms from these pathways not identified as dif-  
452 ferential previously were among the differentially expressed KOs in the validation  
453 dataset, including the flagellar basal body rod protein, FlgD (K2389), the sigma fac-  
454 tor regulating transcription of flagellar genes, FliA (K02405), the chemotaxis response  
455 regulator that interacts with the flagellar motor switch, CheY (K03413), and the heme-  
456 based aerotactic oxygen sensor, HemA/T (K06595). Exopolysaccharide biosynthesis  
457 was represented by a singular KO term (K11936) corresponding to the same poly- $\beta$ -  
458 1,6-acetyl-D-glucosamine synthase identified in the discovery dataset. The only major  
459 difference between BV subgroup analyses in the discovery and validation datasets  
460 concerned porphyrin metabolism. Only a single KO term from this pathway was  
461 differentially expressed between subgroups: the porphobilinogen synthase, HemB.

462 All previous species-specific associations were recapitulated among BV sub-  
463 groups: BVAB1/*L. iners/Mobiluncus* spp. with flagellar assembly and chemotaxis,  
464 and BVAB1/*P. amnii* with exopolysaccharide biosynthesis. In fact, the presence of  
465 BVAB1 in BV1 but not BV2 samples again appeared to be responsible for a large  
466 degree of the separation of these subgroups (Supplementary Figure S9). We did how-  
467 ever note that the expression of flagellar genes between BV1 and BV2 groups appeared  
468 to be strain-specific, such that the difference in the total number of reads which  
469 were assigned to this pathway and originated from *L. iners* spanned three orders of  
470 magnitude (1,211,315 reads vs. 2,344 reads). Critically, we were able to replicate all

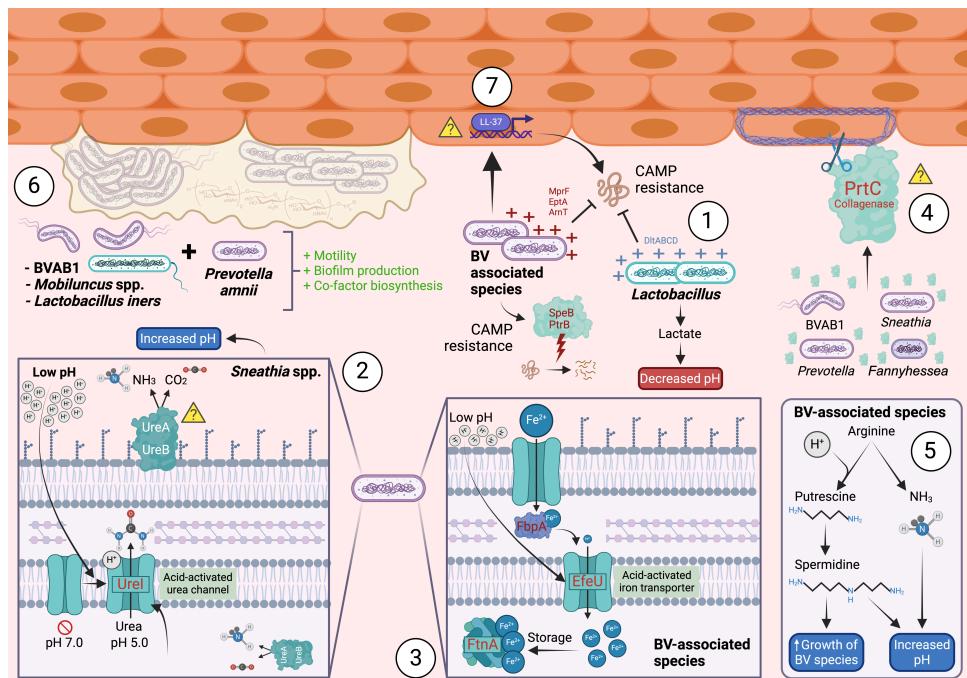
471 of the above findings when genes were aggregated by function using the eggNOG  
472 classification system (Supplementary Figure S10). EggNOG terms corresponding to  
473 flagellar components, chemotactic signalling machinery and exopolysaccharide biosyn-  
474 thesis/biofilm formation were among the most differential functions when comparing  
475 the same BV subgroups.

476 Finally, the size of the validation cohort allowed us to investigate the clear separa-  
477 tion among metatranscriptomes within the molecular health group which appeared to  
478 be due to dominance by either *L. crispatus* or *L. iners*. Differentially expressed func-  
479 tions generally reflected metabolic differences between these two taxa (Figure 5C &  
480 Supplementary Figure S11). Of interest, expression levels of genes involved in flagel-  
481 lar assembly among *L. iners*-dominant samples were significantly higher, while those  
482 encoding a putative polyamine transport system specific for putrescine and spermidine  
483 were among the most differentially expressed transcripts for *L. crispatus*-dominant  
484 samples. Previous work has demonstrated that vaginal isolates of *L. crispatus* are  
485 capable of producing or degrading biogenic amines, including putrescine, depending on  
486 the strain [58]. Moreover, this transport system was also detected in a recent RNA-seq  
487 study of the vaginal environment [23]. Equally of note was a single KO term assigned  
488 to ‘dioxin degradation’ (K01821; PraC/XylH); of all reads in the validation dataset  
489 assigned to this KO term, 99.9% originated from lactobacilli (74.6% from *L. crispa-*  
490 *tus*). That dioxins are classed as persistent chemical pollutants generated as industrial  
491 by-products makes it highly unlikely that they are truly metabolised within the vagi-  
492 nal environment [59]. We note that this KO term corresponds to 2-hydroxymuconate  
493 tautomerase activity and that 2-hydroxymuconate and catechol are the products of  
494 enzymatic reactions involving enzymes of the EC class 3.7.1 (hydrolases acting on C-C  
495 bonds in ketonic substances). Furthermore, 2-hydroxymuconate can be metabolised to  
496 pyruvate via  $\gamma$ -oxalocrotonate and 4-hydroxy-2-oxopentanoate. Though entirely spec-  
497 ular, this could theoretically represent a mechanism by which vaginal estrogen may  
498 be used as an energy source by *L. crispatus*.

### 499 3 Discussion

500 Overall, our meta-analysis of three independent vaginal metatranscriptome datasets  
501 identified differing mechanisms by which species associated with health and BV evade  
502 local host immune responses and persist within the vaginal niche. We also observed dis-  
503 distinct functional subgroups within the BV population of all datasets, largely driven by  
504 the presence of BVAB1 and differing in expression of pathways potentially of relevance  
505 to BV pathogenesis, including motility and biofilm formation. Critically, these findings  
506 were robust to the different methodological approaches employed by the original study  
507 authors, including the large discrepancy between study population demographics, as  
508 well as different functional classification systems employed in our analyses. Our find-  
509 ings, together with other known host-microbe interactions in the vaginal environment,  
510 are summarised in Figure 6.

511 The adoption of the principles of SRI into the Bayesian models constructed by  
512 ALDEx2 was critical for analysis of these vaginal metatranscriptome datasets. The  
513 original ALDEx2 algorithm acknowledged that sequencing datasets represent a single



**Fig. 6 Summary of microbial persistence mechanisms and host-microbe interactions in the vaginal environment:** All functions assigned according to the Kyoto Encyclopedia of Genes and Genomes. Yellow triangles indicate putative mechanisms. **1.** Resistance to host cationic antimicrobial peptides (CAMPs; see Figure 2D) involves reducing net negative surface charge or direct CAMP degradation [39]. **2.** Low vaginal pH triggers the opening of the inner membrane acid-activated urea channel, UreI (K03191). Cytoplasmic and/or surface-adsorbed UreA/B urease may then serve to raise local pH [44]. **3.** Low vaginal pH also triggers opening of the acid-activated periplasmic iron channel, EfeU, permitting transport of iron across the periplasm by FbpA (K02012) and its intracellular accumulation by the non-heme ferritin, FtnA (K02217) [47]. **4.** Production of a U32-family peptidase orthologous to PrtC (K08303) found in vaginal *Porphyromonas* isolates may degrade collagen in the extracellular matrix of the vaginal mucosa [46]. **5.** Production of the biogenic amines, putrescine and spermidine, from arginine. These polyamines, along with the liberation of ammonia during their synthesis, raise vaginal pH and can promote the growth of BV-associated taxa [60]. **6.** Subgroups of BV are delineated by biofilm formation, motility and co-factor biosynthesis; BVAB1 and *Prevotella amnii* synthesize poly- $\beta$ -1,6-acetyl-D-glucosamine via PgaC (K11936) [52]. BVAB1, *Mobiluncus* spp., and *L. iners* express various flagellar [51] and chemotaxis [50] genes as part of the ‘motile’ BV genotype. **7.** Induction of LL-37 production and secretion by *Gardnerella* spp., as occurs at the urothelium during urinary tract infections. Similar mechanisms may be employed by *Gardnerella* spp. in the vagina to out-compete lactobacilli or other taxa [61]. Created with BioRender.

514 point estimate of the sampled environment which does not reflect the true commu-  
 515 nity composition. ALDEX2 generates a posterior distribution of the observed data by  
 516 repeated sampling from a multinomial Dirichlet distribution [26]. This posterior dis-  
 517 tribution undergoes a centred log-ratio normalisation using the geometric mean ( $G_n$ )  
 518 of all features in a given sample. Recently, Nixon *et al.* [20] demonstrated that in doing  
 519 so, ALDEX2 makes an inherent assumption about the scale of the community; that  
 520 the inverse of the geometric mean of the observed counts is exactly equal to the scale.  
 521 However, the estimate of the geometric mean calculated by ALDEX2 is, by definition,

522 a precise but inaccurate measure of scale since all information about community scale  
523 is lost during the sequencing process. Thus, even small errors in the scale assumption  
524 can have a drastic effect on estimates of the true log fold-change of a given feature  
525 between conditions. For this reason, these authors recommended that models should  
526 account for errors in the scale assumption during normalisation to prevent converging  
527 on a precise but inaccurate estimate of scale and hence differential expression.

528 We identified several genes involved in CAMP resistance among the most differ-  
529 entially expressed pathways between healthy and BV metatranscriptomes. Crucially,  
530 this finding was robust to replication across three independent datasets which were  
531 distinct in terms of geography,sampling methodology, pregnancy status and ethnicity.  
532 Given the well-reported effect of ethnicity on vaginal microbiome composition [13, 62],  
533 the replication of almost all findings across exploratory and validation datasets is  
534 remarkable and underscores the value of correct data normalisation. In addition to  
535 uncovering several mechanisms by which health- and BV-associated species persist  
536 within the vaginal niche, this finding also sheds light on important co-evolutionary  
537 processes within the vaginal microbiome. Kiattiburut *et al.* [61] demonstrated how  
538 BV-associated species implicated in urogenital infections (including *G. vaginalis*; W.  
539 Kiattiburut & J. Burton, personal communication) are capable of inducing the pro-  
540 duction and secretion of host CAMPs such as LL-37 by urothelial cells. As lactobacilli  
541 do not evoke the same response [61], it is feasible that this represents a unique  
542 strategy employed by *Gardnerella* to redirect host LL-37 towards lactobacilli. That  
543 both taxa— along with a multitude of other vaginal microbiome constituents— express  
544 CAMP resistance genes points to ongoing co-evolution within these species and also  
545 explains earlier observations regarding the general resistance of vaginal lactobacilli to  
546 CAMPs [63]. Given the importance of *Gardnerella* species in the context of BV, target-  
547 ing L-lysylation of membrane phospholipids by MprF and increasing susceptibility to  
548 CAMP-mediated lysis may prove to be of benefit when considering novel therapeutics  
549 for BV.

550 Many surveys of the vaginal microbiome employing amplicon sequencing have  
551 attempted to define subtypes of BV-associated CST IV, largely based around what was  
552 at the time thought to be subgroups/clades of *G. vaginalis* and other BV taxa [64, 65].  
553 Our findings regarding functional subgroups of metatranscriptomes suggest that tax-  
554 onomic classification using canonical CSTs is largely redundant— especially amongst  
555 BV populations. While expression levels of transcripts from BVAB1 appeared to differ  
556 between BV subgroups, the expression of transcripts from other species contributing  
557 to differences in expression of flagellar (*L. iners*) or exopolysaccharide biosynthesis  
558 (*P. amnii*) genes did not. For these species, differences between subgroups were likely  
559 due to strain-specific differences in gene content, as was the case in older culture-based  
560 studies reporting the isolation of flagellated and non-flagellated lactobacilli from vagi-  
561 nal swabs [66]. Given the importance of polymicrobial biofilms in BV pathogenesis  
562 (and more generally in resistance to antimicrobial agents), the existence of a motile,  
563 biofilm-forming BV subtype may be of importance when considering the triggers for  
564 the onset of active BV symptoms and when investigating causes of treatment failure.

565 Differences in the expression of genes involved in iron uptake and storage is not  
566 particularly surprising, given that lactobacilli are not dependent on iron for growth.

567 Although sensitivity to iron levels may depend on nucleotide availability in some lac-  
568 tobacilli [67], most others are unaffected and instead rely on manganese and cobalt as  
569 co-factors in their essential redox enzymes [48]. It is noteworthy then that some BV  
570 subgroups express enzymes involved in cobalamin biosynthesis which, by definition,  
571 requires uptake of cobamide or other forms of cobalt. This may represent a strategy  
572 employed by BV-associated species to out-compete lactobacilli through the sequestra-  
573 tion of a valuable resource. Likewise, expression of pH-dependent urea channels at the  
574 higher levels noted among BV metatranscriptomes compared to their healthy coun-  
575 terparts could reflect attempts by BV-associated taxa to counteract the low vaginal  
576 pH. Although detected at low levels, both urease subunits were also detected in our  
577 meta-analysis; this enzyme is central to similar pH-raising mechanisms that underpin  
578 virulence in pathogens such as *Helicobacter pylori* [44].

579 Finally, the identification of a collagenase present in several BV-associated species  
580 raises questions about the role played by collagen degradation BV pathogenesis, par-  
581 ticularly regarding the cause vs. effect of dysbiosis. Several reviews [68, 69] point to  
582 the production of “collagenases and fibrinolysins” by various species of *Sneathia* and  
583 *Prevotella*, citing the work of Al-Mushrif *et al.*; however this work investigated the  
584 *in vitro* effects of organic acids on monocytes and only stated the fact that several  
585 pathogenic *Prevotella* produce these enzymes [70]. In the chain of papers cited, Hof-  
586 stad ultimately demonstrated in 1984 the collagen- and fibrin-degrading activities of  
587 various *Prevotella* species (then classified as *Bacteroides*), though vaginal *Prevotella*  
588 were not studied [71]. More recently, preliminary work by Lithgow and colleagues  
589 confirmed that vaginal isolates of *Porphyromonas uenonis*, *Porphyromonas asaccha-*  
590 *rolytica* *Prevotella bivia* and *Sneathia amnii* all possess secretory collagenase activity  
591 [46]. While it remains to be experimentally proven how these enzymes access collagen  
592 within the extracellular matrix of the vaginal mucosa, damage to the epithelial bar-  
593 rier could favour bacterial attachment and influence biofilm formation, as is the case  
594 in the respiratory tract [72, 73]. Collagenase expression may also contribute to the  
595 production of the detached ‘clue cells’ [55] observed during clinical diagnosis, as these  
596 biofilm-covered epithelial cells could be released upon degradation of the extracellular  
597 matrix.

## 598 4 Conclusion

599 Overall, we have demonstrated the advantages of incorporating tools such as SRI  
600 into ‘omics analyses, showing how it can be used to identify differentially expressed  
601 functions which are robust to replication even across independent and highly divergent  
602 datasets. Our application of these tools to vaginal metatranscriptomic datasets opens  
603 several avenues of investigation in relation to potential therapeutic targets for BV  
604 which require further investigation to confirm their involvement in pathogenesis. The  
605 inclusion of SRI into the ALDEx2 platform should be of general use when examining  
606 multiple other metatranscriptomes and other problematic high throughput sequencing  
607 datasets.

## 608 5 Methods

### 609 5.1 Meta-analysis datasets

610 Three previously published vaginal metatranscriptome datasets were included in our  
611 meta-analysis: Macklaim *et al.* (London [27]), Deng *et al.* (European [28]), and  
612 Fettweiss *et al.* (Virginia [7]) (see "Data availability and code" section for access  
613 information).

614 The London dataset was produced from vaginal swabs of the posterior fornix and  
615 mid-vaginal wall, collected from 22 women in 2013. Total RNA was extracted from  
616 swab fluid using a standard workflow [27] and RNA libraries were sequenced on an  
617 Illumina HiSeq 2000 instrument.

618 The European dataset was produced from vaginal lavage samples (2 mL saline)  
619 collected from 22 women between April 2014 and September 2015. Total RNA was  
620 extracted from 1 mL of vaginal fluid suspension using a Mo Bio PowerMicrobiome  
621 RNA Isolation kit and rRNA was depleted using a Ribo-Zero Gold rRNA Removal  
622 kit [28]. Paired-end RNA libraries were constructed with an Illumina ScriptSeq kit (2  
623 x 110 bp reads) and sequenced on a HiSeq 2500 instrument.

624 The Virginia dataset was produced from vaginal swabs of the mid-vaginal wall, col-  
625 lected from pregnant women as part of the MOMS-PI study. Total RNA was extracted  
626 using a Mo Bio PowerMicrobiome RNA Isolation kit and rRNA was depleted using  
627 an Epicentre/ Illumina Ribo-Zero Magnetic Epidemiology Kit [7]. Paired-end RNA  
628 libraries were constructed with a KAPA Biosystems KAPA RNA HyperPrep Kit (2 x  
629 150 bp reads) and sequenced on an Illumina HiSeq4000 instrument.

### 630 5.2 Re-processing of raw data

631 All FASTQ files were re-processed in the same manner to minimise differences across  
632 datasets due to bioinformatic pipelines. Raw reads were trimmed for quality (minimum  
633 Q20 across a 4bp sliding window) and length (cropped to 75 bp) using Trimmomatic  
634 (v0.39). Trimmed reads were then sequentially mapped to two different human genome  
635 assemblies (GRCh38 and T2T-CHM13) and all mapped reads were discarded. The  
636 remaining non-human reads were then mapped to the SILVA rRNA database (release  
637 138) and all putative rRNA reads were discarded. All mapping steps were performed  
638 using Bowtie2 (v2.5.1) with the high accuracy option. Finally, non-human, rRNA-  
639 filtered reads were mapped to the VIRGO database—a non-redundant catalogue of  
640 genes from the human vagina [22] for taxonomic and functional assignment, and feature  
641 tables for each dataset were constructed separately, using scripts provided by the  
642 maintainers of VIRGO. Data were generated for gene-level, eggNOG [74] and KEGG  
643 [75] level analyses.

### 644 5.3 Filtering and batch-correction of metatranscriptomes

645 Feature tables from VIRGO were merged together in R and filtered with the CoDaSeq  
646 package to retain only the genes present in  $\geq 30\%$  of samples, with a proportional  
647 abundance of  $\geq 0.005\%$  in at least one sample. Less strict filtering criteria were

648 explored in the R script ‘revisions\_filtering\_exploration.R’, but reducing the prevalence  
649 threshold to  $\geq 10\%$  of samples did not alter any study conclusions (see GitHub).  
650 Filtered London and European datasets then underwent batch correction using the  
651 ComBat\_seq function from the sva package [31], prior to all downstream analysis in  
652 R. Batch correction was not performed for the Virginia dataset owing to its use as an  
653 independent confirmatory dataset. Analyses comparing metatranscriptomes by preg-  
654 nancy status were performed on a merged feature table of the London, Europe and  
655 Virginia datasets following filtering and batch-correction as above.

## 656 5.4 Species-level summaries of vaginal metatranscriptomes

657 Taxonomic composition of metatranscriptomes at the species level was summarised  
658 in terms of proportional abundance of expressed transcripts (calculated from fil-  
659 tered, batch-corrected feature tables) using the pheatmap R package and taxonomic  
660 assignments according to the VIRGO database.

661 For visualisation of community composition via principal component analysis  
662 (PCA), batch-corrected feature tables underwent scaled log-ratio (SLR) transfor-  
663 mation using the updated ALDEx2 package in R, prior to PCA with the prcomp  
664 function. The resultant compositional biplots were visualised using the CoDaSeq R  
665 package (<https://github.com/ggloor/CoDaSeq>). Gene functions were assigned using  
666 KEGG orthology (KO) numbers, eggNOG group membership and EC classification,  
667 as provided by VIRGO [22].

668 To differentiate the named species of the genus, *Gardnerella*, we obtained genomes  
669 of the species type strain, plus additional genomes with  $\leq 100$  contigs ( $\leq 200$  for *G.*  
670 *pickettii*), from NCBI for each of the following species (number of genomes indicated  
671 in parentheses): *G. greenwoodii* (4), *G. leopoldii* (5), *G. pickettii* (11, *G. piotii* (4),  
672 *G. swidsinskii* (7), and *G. vaginalis* (13). Accession numbers for all genomes used are  
673 listed in the file ‘gardnerella\_genomes.xlsx’ found in the ‘Supplement’ directory of this  
674 study’s GitHub repository. Genomes were annotated for each species separately using  
675 bakta (v 1.93) [76]. Output .gff files were used as input for pan-genome analysis with  
676 panaroo (v1.3.3) [57] and the resulting presence-absence matrix was used to identify  
677 genes unique to each named *Gardnerella* species. Sequences corresponding to these  
678 ‘species marker genes’ were extracted from the VIRGO catalogue and aligned to the  
679 VIRGO database. BLAST results were filtered to retain only hits with  $\geq 95\%$  sequence  
680 identity and a  $\geq 90\%$  alignment overlap, and the corresponding ‘marker’ genes were  
681 used to estimate the location of the named *Gardnerella* species on PCA plots.

## 682 5.5 Differential expression analysis with ALDEx2 using 683 scale-reliant inference

684 Differentially expressed genes and pathways were identified using the ALDEx2 pack-  
685 age in R using scale simulated random variable methods [20, 24, 26]. The underlying  
686 scale was estimated by running the ‘aldex.clr’ function with gamma set to  $1e^{-3}$  (i.e.  
687 nearly 0). The scale values for the two conditions can be determined as the average  
688 of the rows of the ‘@scaleSamps’ slot in the resulting ‘clr’ class object. We observed  
689 that the mean scale values differed by approximately 8-fold ( $\sim 2^{2.95}$ ), and that this

690 resulted in the housekeeping functions being offset from the line of no difference (Sup-  
691 plementary Figure S1). Examining the scale of just the housekeeping functions in the  
692 same manner suggested that a scale difference of ~ 0.15 would be more appropriate.  
693 Accordingly, all differential expression analyses performed with the updated ALDEX2  
694 package implemented a 15 % scale difference between groups, and a scale uncertainty  
695 of 0.5 standard deviations (as recommended previously [20, 24]). All feature tables  
696 were log<sub>2</sub>-transformed across 128 Monte-Carlo instances with these parameters.

697 All values calculated by ALDEX2 and used for analysis are the median values of  
698 the posterior distribution for each gene or function in each sample. Median scaled  
699 log-ratio (SLR) values, absolute effect sizes and expected *P*-values from Benjamini-  
700 Hochberg-corrected Welch's t-test were calculated across all Monte-Carlo replicates.  
701 and were accessed using the 'include.sample.summary = TRUE' parameter within  
702 the 'aldex.effect' function. Appropriate effect size thresholds were assessed using the  
703 'aldex.plot' function, and KO or EggNOG terms with an absolute effect size  $\leq -1$  or  $\geq 1$   
704 were considered differentially abundant. Note that all parts with these effect sizes also  
705 had an FDR of  $\leq 1\%$ . Normalised Z-scores of effect size were calculated per-sample for  
706 pathways represented by differentially expressed KO/EggNOG terms ( $\frac{x-\mu}{\sigma}$ ), where  $x$   
707 is the median effect size,  $\mu$  is the per-sample mean effect size and  $\sigma$  is the per-sample  
708 effect size standard deviation. Normalised Z-scores were then plotted per-pathway  
709 using the pheatmap package.

## 710 5.6 Identification of signature functions between groups

711 Signature functions were identified for groups of metatranscriptomes (healthy vs. BV  
712 and BV subgroups 1 vs. 2) using the 'coda4microbiome' R package. Feature tables were  
713 aggregated by KEGG pathway and subject to SLR transformation using 'aldex.clr',  
714 calculating the geometric mean using only the low-variance, high-abundance features.  
715 Pathways with a median within-group dispersion  $\leq 4$  and effect score  $\geq 0.5$  were passed  
716 to the 'coda\_glmnet' function to identify signature functions.

## 717 5.7 Re-assigning pathways

718 All gene function assignments were performed automatically using the Kyoto Ency-  
719 clopedia of Genes and Genomes [75]. However, due to the unique environment and  
720 somewhat generic annotation in KEGG, several assignments were clearly incorrect.  
721 Each function was individually investigated and re-assigned using the Kyoto Encyclo-  
722 pedia of Genes and Genomes. First the KO numbers were identified, then each function  
723 was investigated looking at the pathway, enzymatic function, chemical reaction, loca-  
724 tion, and which species was expressing it in our samples. Using that knowledge,  
725 as well as knowledge of the vaginal environment, the most probable function was  
726 inferred. KO terms that were modified, including the rationale for each, are in the  
727 'Supplement' directory under 'pathway-changes.xlsx' and in 'lon\_eur\_health\_vs\_BV.R'  
728 and 'virginia\_definingMolecularBV.R' in the 'code' directory of the GitHub repository  
729 for this study. The latter two R scripts also contain DOIs linking to peer-reviewed  
730 literature supporting the KO term changes at the end of the files.

**731 6 Declarations**

**732 6.1 Ethics approval and consent to participate**

**733** A data access request was submitted to the NCBI Database of Genotypes and Phenotypes (dbGaP) to obtain the FASTQ files associated with vaginal metatranscriptome samples from the MOMS-PI study (dbGaP; accession no. phs001523.v1.p1). Access was authorised by dbGaP following the conditions of the data access policy under the University of Western Ontario Research Ethics Board approval #123506.

**738 6.2 Consent for publication**

**739** Not applicable.

**740 6.3 Availability of data and materials**

**741** Raw FASTQ files are available for each of the datasets used. London and European datasets are immediately available from the European Nucleotide Archive under the accession numbers PRJEB31833 and PRJEB21446, respectively. Raw FASTQ files from the Virginia are available from dbGaP under accession no. phs001523.v1.p1.

**745** Whole genome sequences of *Gardnerella* species used in this study can be found under the RefSeq accession numbers listed in the file, ‘gardnerella\_genomes.xlsx’ in the ‘Supplement’ directory of this study’s GitHub repository.

**748** All code used for the re-processing of FASTQ files at the command line, and data analysis in RStudio is available as a code stable release at: <https://github.com/scottdossantos/dossantos2024study>.

**751 6.4 Competing interests**

**752** The authors declare that they have no competing interests.

**753 6.5 Funding**

**754** SJDS and JMM were supported by an internal grant from the Schulich School of Medicine & Dentistry at Western University, awarded to GG. CC was supported by an award from the Weston Family foundation (award no. 20170705). GR held a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

**758 6.6 Authors’ contributions**

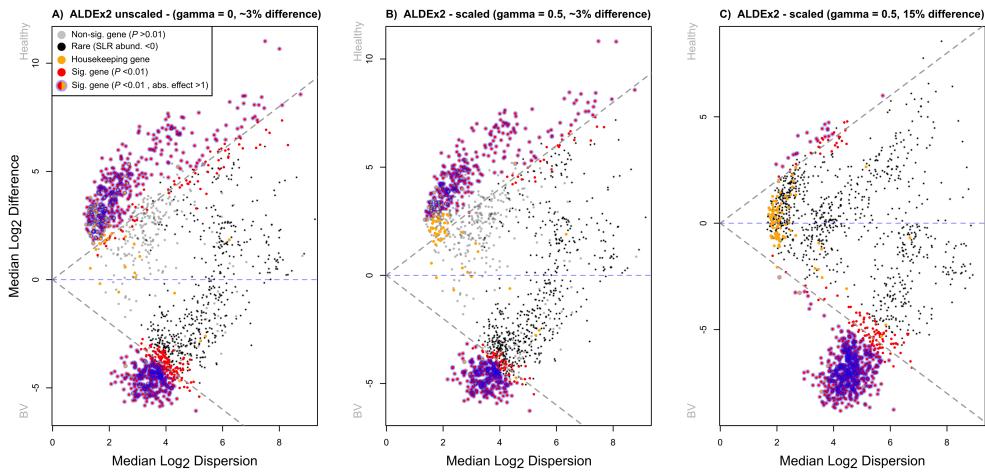
**759** Author contributions are listed below, according to the CRediT taxonomy:

- 760** • Funding acquisition: GR, GBG
- 761** • Conceptualisation: GR, GBG
- 762** • Data curation: SJDS, CC, JMM
- 763** • Formal analysis: SJDS, CC, JMM, GG
- 764** • Writing- original draft: SJDS, CC, GBG
- 765** • Writing- review & editing: All authors

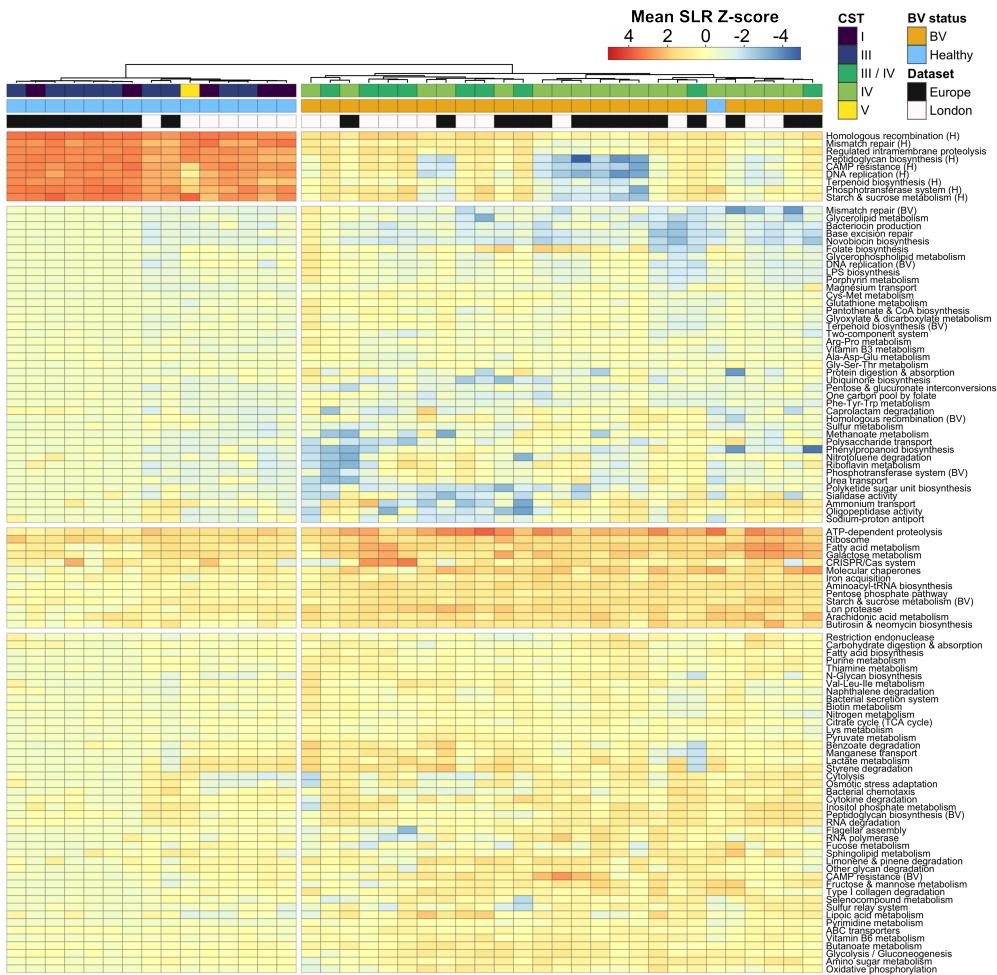
**766 6.7 Acknowledgements**

**767** The authors would like to thank Dr. Jeremy Burton and Wongsakorn Kiattiburut for  
**768** their helpful discussions regarding cationic antimicrobial peptides, and Amy McMillan  
**769** for her assistance with data collection.

## 770 7 Supplementary material



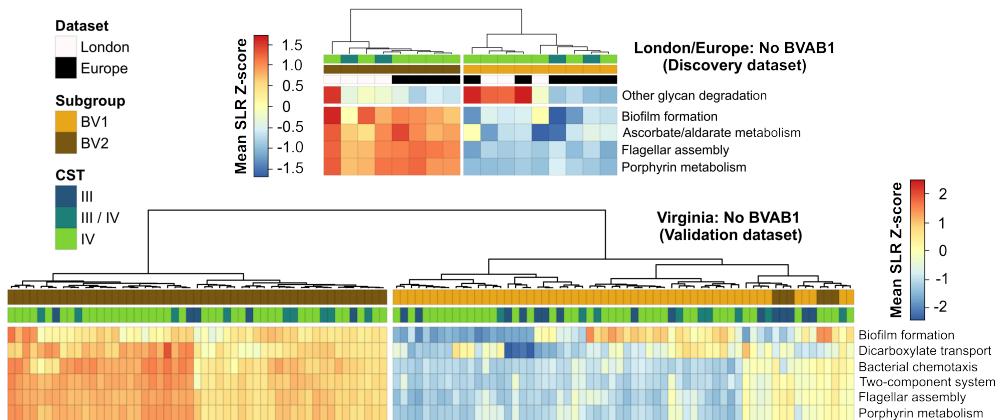
**Supplementary Fig. S1 Using a between-group scale offset of 15 % is sufficient to centre a vaginal metatranscriptomic dataset:** MW effect plots were produced for the London/Europe discovery dataset normalised using (A) ALDEEx2 without scale simulation, (B) ALDEEx2 operating with the default scale model ( $\gamma = 0.5$ ) or (C) ALDEEx2 running with a full scale model ( $\gamma = 0.5$ , 15 % scale offset between healthy and BV groups). Not accounting for scale results in a large number of false-positive results (A) and a clear error in the centre-point of the data which is, by definition, determined by the housekeeping features. Adding a small amount of scale uncertainty (B) increases within-group dispersion of all features such that many housekeeping functions are no longer differential; however, only the full scale model (C) correctly centres the housekeeping functions around the point of no difference between groups and corrects the large number of false-positive and -negative results. Colouring of data points as follows: non-significant features = grey ( $P \geq 0.01$ ), non-significant, rare features = black (median SLR abundance  $<0$ ), housekeeping features = orange (KO terms corresponding to 'Glycolysis', 'Ribosome' or 'Aminoacyl-tRNA biosynthesis'), significant features based on  $P$ -value alone = red ( $P \leq 0.01$ ) and significant features based on  $P$ -value and effect size = red with blue outline ( $P \leq 0.01$ , absolute effect size  $\geq 1$ ). Grey dashed lines indicate equivalence of between- and within- group difference; blue dashed line indicates no difference between groups.



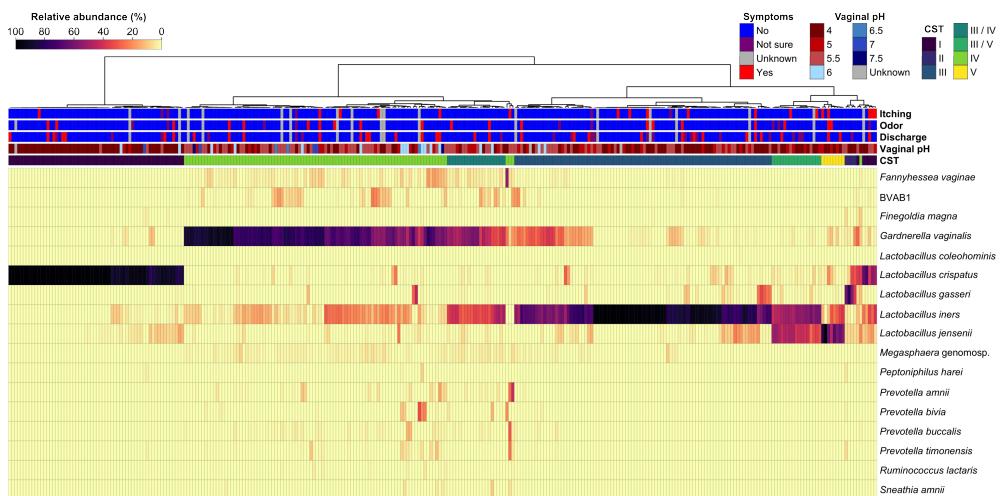
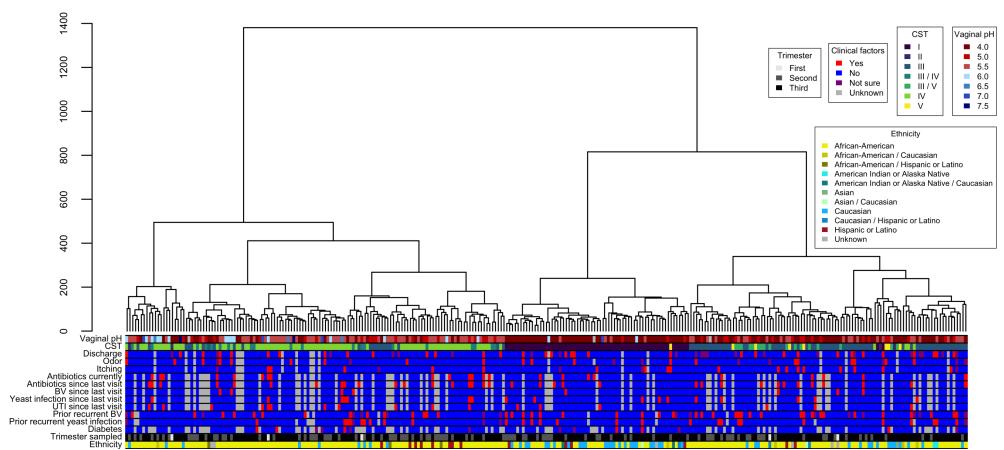
**Supplementary Fig. S2 Differentially expressed pathways between health and BV – London/Europe dataset:** Gene assignments were grouped by KEGG identifiers regardless of species for each sample and all differentially abundant KEGG pathways between healthy and BV metatranscriptomes in the London/Europe dataset are shown. All post-SRI absolute effect sizes and expected false discovery rates calculated by ALDEx2 are  $>1$  and  $<0.01$ , respectively. CST, BV status and dataset indicated by colour bars.

**Supplementary Table S1** Total gene and read counts for all genes assigned to the four KEGG orthology terms that represent the DltABCD operon. Data are separated to show gene and read counts for the London/Europe (discovery) and Virginia (validation) datasets.

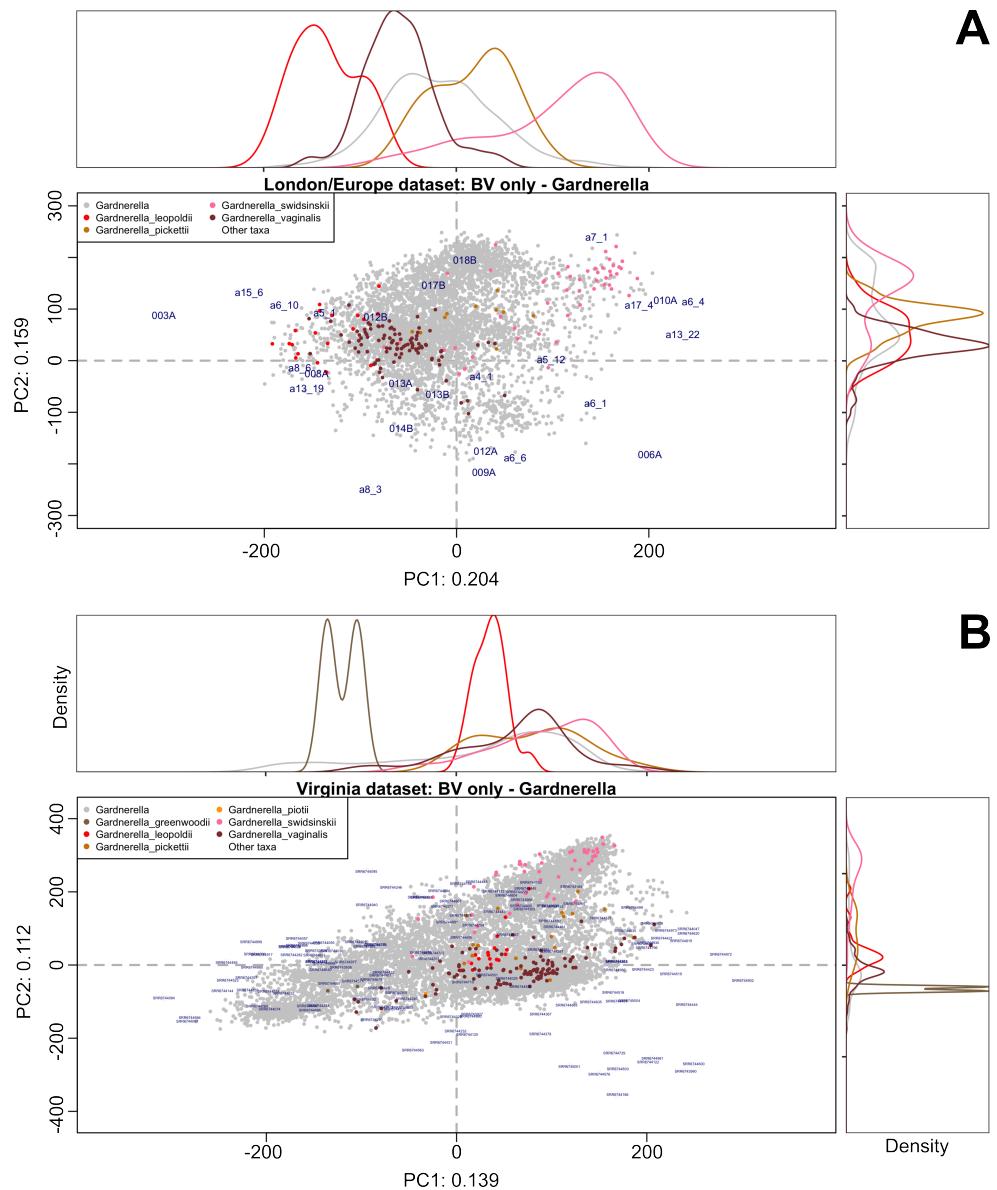
| Dataset             | London/Europe |           | Virginia  |           |
|---------------------|---------------|-----------|-----------|-----------|
|                     | No. genes     | No. reads | No. genes | No. reads |
| <i>L. crispatus</i> | 5             | 115,858   | 8         | 7,191,584 |
| <i>L. iners</i>     | 8             | 177,871   | 8         | 5,429,846 |
| <i>L. jensenii</i>  | 2             | 79,900    | 7         | 1,613,482 |
| <i>L. gasseri</i>   | 2             | 14,043    | 2         | 144,203   |
| Unknown taxonomy    | 9             | 41,381    | 11        | 1,458,737 |



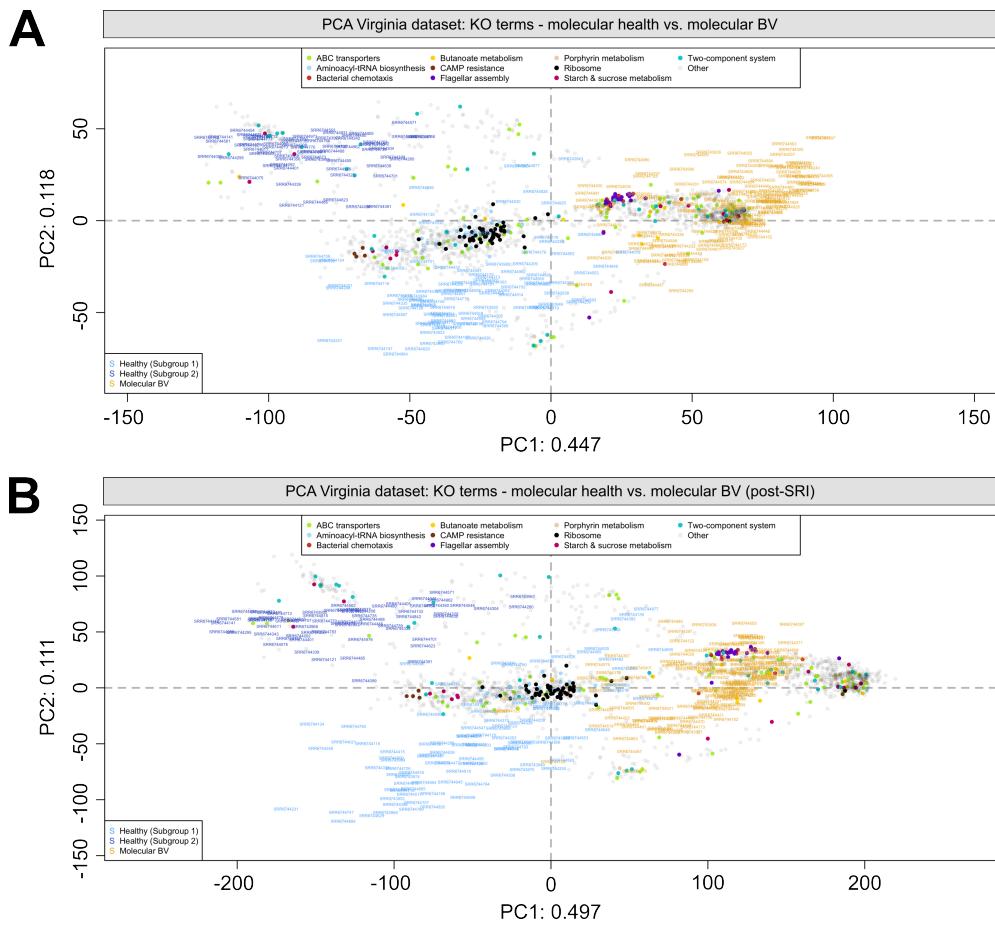
**Supplementary Fig. S3 BV subgroup differences are not solely due to the presence of BVAB1:** Differential expression analyses for both London/Europe (top) and Virginia (bottom) datasets were repeated after removing all genes with a taxonomic assignment to BVAB1. Differentially expressed pathways among samples are expressed as a Z-score calculated from the mean scaled log-ratio (SLR) values of all KO terms assigned to a given pathway (individual SLR values represent a median across 128 Monte-Carlo instances). All post-SRI absolute effect sizes and false discovery rates calculated by ALDEEx2 were  $>1$  and  $<0.01$ , respectively).

**A****B**

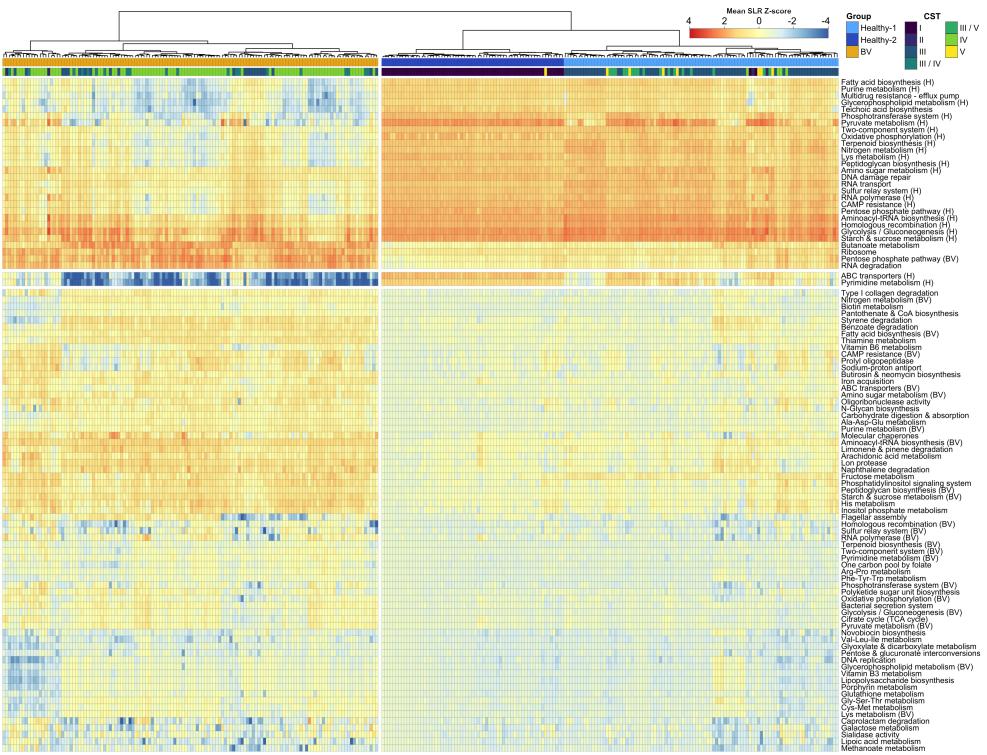
**Supplementary Fig. S4 Definition of molecular states of health and BV by clustering of metatranscriptomes at the functional level:** All reads from the Virginia dataset were aggregated by species (proportional abundance of expressed transcripts; **A**) or function (KO term; **B**), prior to hierarchical clustering of Euclidean distances using Ward's method. Colour bars indicate vaginal pH, CST, clinical metadata, and ethnicity. (**A**) Proportional abundances of expressed metatranscriptome reads with known taxonomy; species represented by  $\geq 75$  genes are shown. (**B**) Dendrogram of SLR-transformed, post-SRI metatranscriptome profiles aggregated by KO term, regardless of taxonomy.



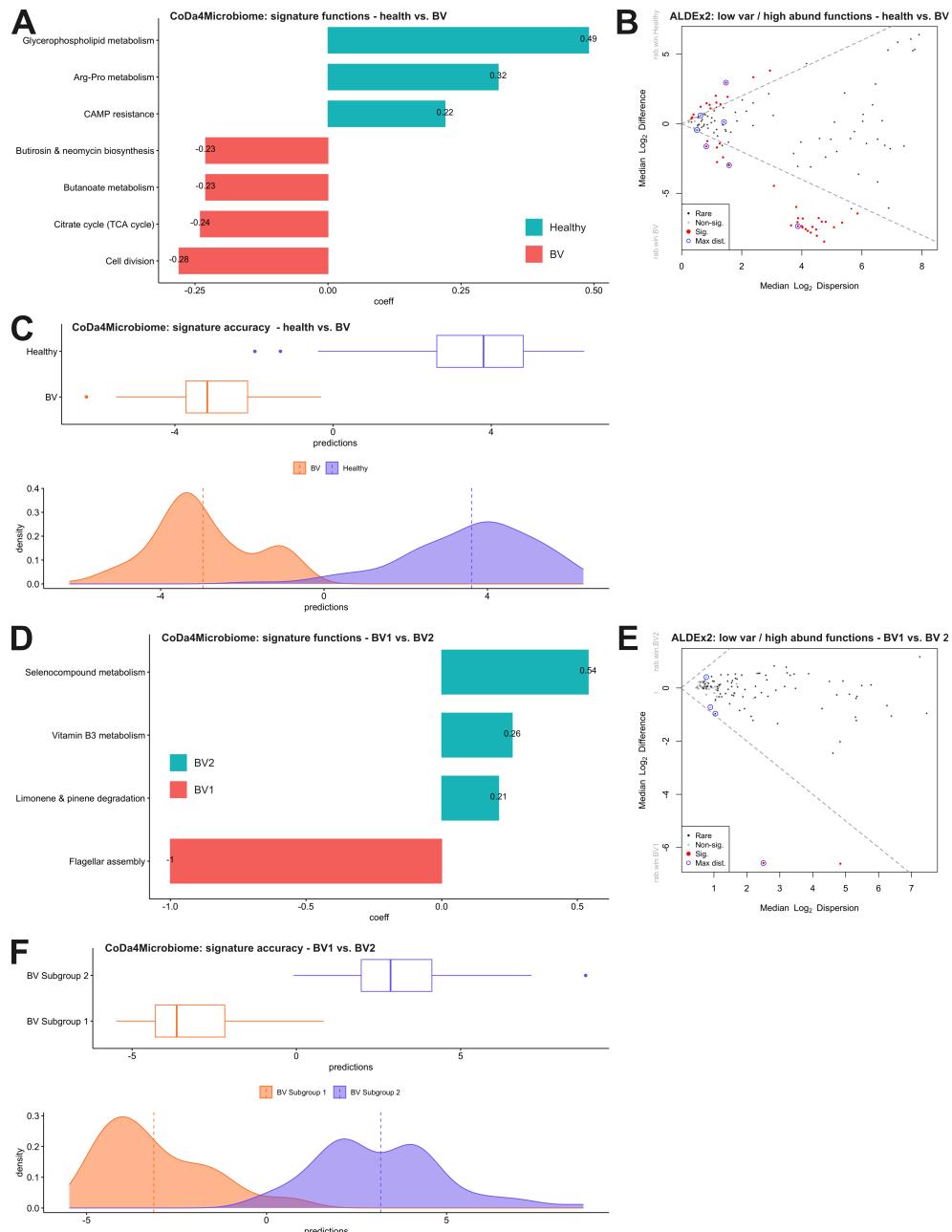
**Supplementary Fig. S5 Individual *Gardnerella* species can be distinguished within metatranscriptome profiles:** Metatranscriptome profiles from London/Europe (A) and Virginia (B) datasets were subject to SLR-transformation prior to PCA. All gene features were included in the ordination; however only reads assigned to *Gardnerella* are shown. Genes found to be unique to each of the named *Gardnerella* species by pangenome analyses are coloured.



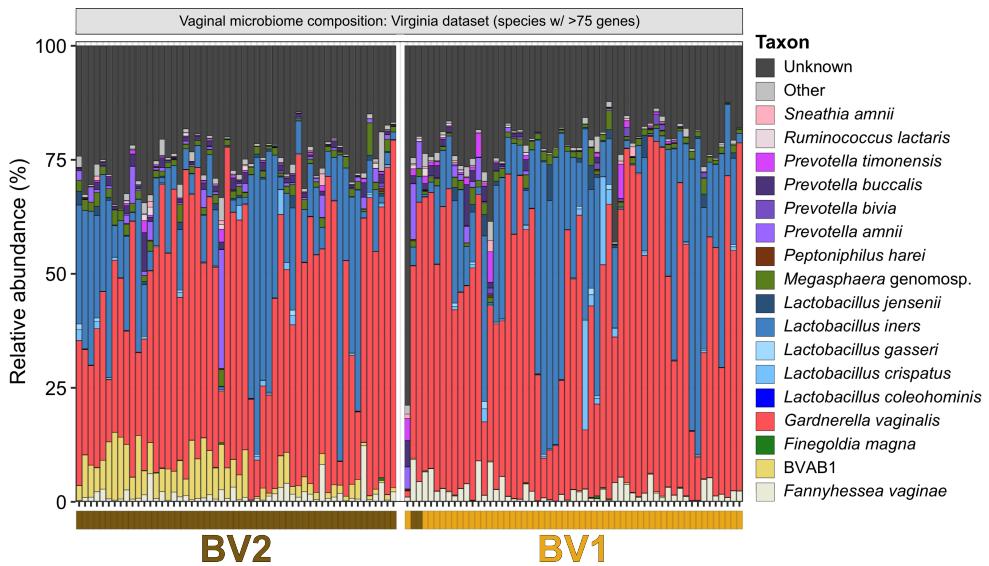
**Supplementary Fig. S6 Correcting data asymmetry with SRI effectively centres genes of no-difference:** KO-aggregated metatranscriptome profiles from the Virginia dataset underwent SLR-transformation either alone (**A**) or with SRI (15 % scale difference, gamma = 0.5) prior to PCA. KO terms coloured by KEGG pathway assignments for a selection of relevant KEGG pathways. Samples coloured by group according to hierarchical clustering in Supplementary Figure [S4](#).



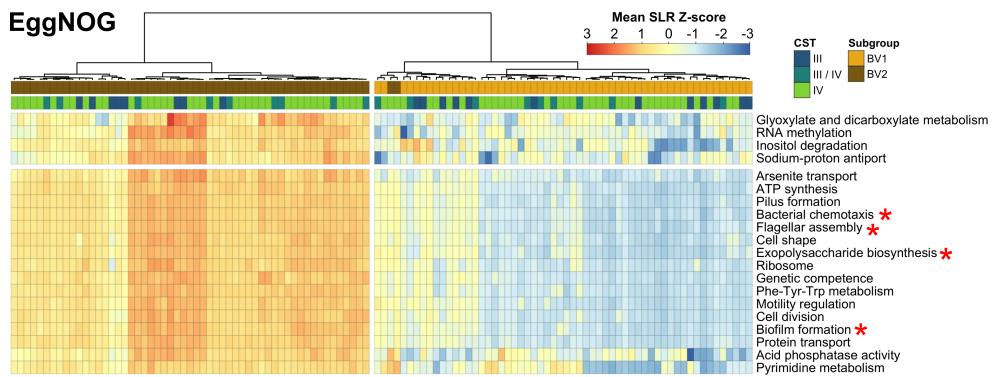
**Supplementary Fig. S7 Major differentially expressed pathways of the vaginal metatranscriptome can be replicated across datasets:** Metatranscriptome profiles from the Virginia dataset were aggregated by KO term prior to SLR-transformation and SRI using ALDEEx2. Analysis was stratified by states of molecular health or BV and Z-scores of mean SLR values per pathway are plotted. Colour bars indicate group and CST. All post-SRI effect sizes and BH-corrected *P*-values calculated by ALDEEx2 are  $\leq 0.01$  and  $< 1\%$ , respectively.



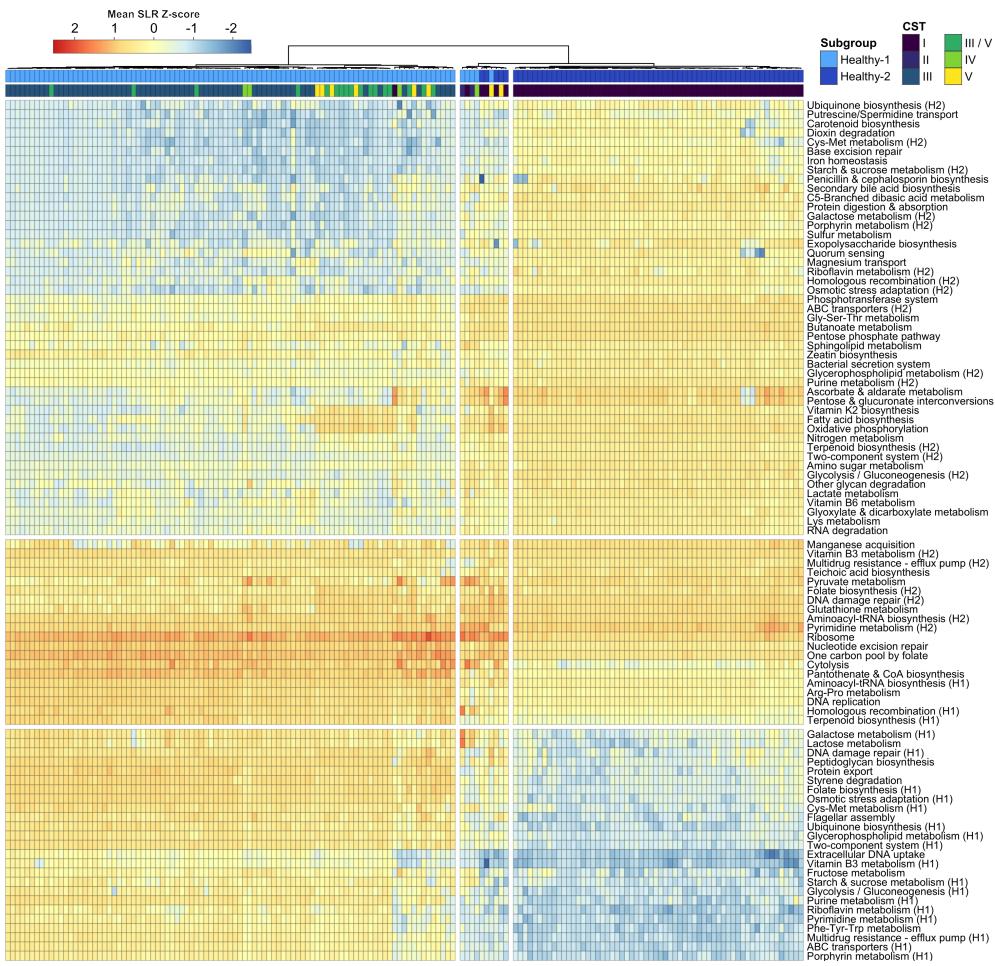
**Supplementary Fig. S8 Prediction of signature functions complements differential expression analyses:** Prediction of signature functions for health vs. BV (A-C) and BV subgroups (D-F) by the 'cod4microbiome' R package. Sets of functions identified in signature plots (A,D) exhibit high AUC values of 0.9975 and 0.9947, respectively, and are indicated on ALDEEx2 effect plots in blue. Prediction plots (C,F) demonstrate a high degree of separation by group for these signatures, indicating that they define the respective groups with a high degree of accuracy. These results should be interpreted with caution, given that calculating the geometric mean from low-variance, high-abundance features does not eliminate the problematic data asymmetry shown in Suppl. Fig S1.



**Supplementary Fig. S9 Proportional abundance of BVAB1 transcripts is not the sole driver of functional differences between molecular BV subgroups:** Metatranscriptome profiles of samples belonging to the two main BV subgroups within the Virginia dataset are expressed as proportional abundance of expressed transcripts. Reads of known and unknown taxonomy are included, with reads corresponding to species represented by  $\leq 75$  genes grouped into ‘Other’. Subgroup membership indicated below plot.



**Supplementary Fig. S10 Functional differences between BV subgroups are not dependent on enzyme classification systems:** Metatranscriptome profiles from the Virginia dataset were aggregated by KO term prior to SLR-transformation and SRI using ALDEEx2. Analysis was stratified by membership of molecular BV subgroups and Z-scores of mean SLR values per pathway are plotted. Red asterisks indicate functions identified in the corresponding KEGG KO term analysis. Colour bars indicate group and CST. All post-SRI effect sizes and BH-corrected  $P$ -values calculated by ALDEEx2 are  $\leq 0.01$  and  $< 1\%$ , respectively.



**Supplementary Fig. S11 Differentially expressed pathways between molecular health subgroups –Virginia dataset:** Metatranscriptome profiles from the Virginia dataset were aggregated by KO term prior to SLR-transformation and SRI using ALDEEx2. Analysis was stratified by membership of molecular health subgroups and Z-scores of mean SLR values per pathway are plotted. Colour bars indicate group and CST. All post-SRI effect sizes and BH-corrected *P*-values calculated by ALDEEx2 are  $\leq 0.01$  and  $< 1\%$ , respectively.

## 771 References

- 772 [1] Srinivasan, S., Fredricks, D.N.: The human vaginal bacterial biota and bacte-  
773 rial vaginosis. *Interdisciplinary Perspectives on Infectious Diseases* **2008**, 750479  
774 (2008) <https://doi.org/10.1155/2008/750479>
- 775 [2] Peebles, K., Velloza, J., Balkus, J.E., McClelland, S.R., Barnabas, R.V.: High  
776 global burden and costs of bacterial vaginosis: a systematic review and meta-  
777 analysis. *Sexually Transmitted Diseases* **46**(5), 304–311 (2019) <https://doi.org/10.1097/OLQ.0000000000000972>
- 779 [3] Kenyon, C., Colebunders, R., Crucitti, T.: The global epidemiology of bacterial  
780 vaginosis: a systematic review. *American Journal of Obstetrics and Gynecology*  
781 **209**(6), 505–23 (2013) <https://doi.org/10.1016/j.ajog.2013.05.006>
- 782 [4] Allsworth, J.E., Peipert, J.F.: Severity of bacterial vaginosis and the risk of  
783 sexually transmitted infection. *American Journal of Obstetrics and Gynecology*  
784 **205**(2), 113–16 (2011) <https://doi.org/10.1016/j.ajog.2011.02.060>
- 785 [5] Abou Chacra, L., Ly, C., Hammoud, A., Iwaza, R., Mediannikov, O., Bretelle, F.,  
786 Fenollar, F.: Relationship between bacterial vaginosis and sexually transmitted  
787 infections: coincidence, consequence or co-transmission? *Microorganisms* **11**(10)  
788 (2023) <https://doi.org/10.3390/microorganisms11102470>
- 789 [6] Adolfsson, A., Hagander, A., Mahjoubipour, F., Larsson, P.-G.: How vaginal  
790 infections impact women's everyday life— women's lived experiences of bacterial  
791 vaginosis and recurrent vulvovaginal candidiasis. *Advances in Sexual Medicine*  
792 **7**(1), 1–19 (2016) <https://doi.org/10.4236/asm.2017.71001>
- 793 [7] Fettweis, J.M., Serrano, M.G., Brooks, J.P., Edwards, D.J., Girerd, P.H., Parikh,  
794 H.I., Huang, B., Arodz, T.J., Edupuganti, L., Glascocock, A.L., Xu, J., Jimenez,  
795 N.R., Vivadelli, S.C., Fong, S.S., Sheth, N.U., Jean, S., Lee, V., Bokhari, Y.A.,  
796 Lara, A.M., Mistry, S.D., Duckworth, R.A. 3rd, Bradley, S.P., Koparde, V.N.,  
797 Orenda, X.V., Milton, S.H., Rozycski, S.K., Matveyev, A.V., Wright, M.L.,  
798 Huzurbazar, S.V., Jackson, E.M., Smirnova, E., Korlach, J., Tsai, Y.-C., Dick-  
799 inson, M.R., Brooks, J.L., Drake, J.I., Chaffin, D.O., Sexton, A.L., Gravett,  
800 M.G., Rubens, C.E., Wijesooriya, N.R., Hendricks-Muñoz, K.D., Jefferson, K.K.,  
801 Strauss, J.F. 3rd, Buck, G.A.: The vaginal microbiome and preterm birth. *Nature  
Medicine* **25**(6), 1012–1021 (2019) <https://doi.org/10.1038/s41591-019-0450-2>
- 803 [8] Oduyaobo, O.O., Anorlu, R.I., Ogunsola, F.T.: The effects of antimicrobial therapy  
804 on bacterial vaginosis in non-pregnant women. *Cochrane Database of Systematic  
805 Reviews* (3), 006055 (2009) <https://doi.org/10.1002/14651858.CD006055.pub2>
- 806 [9] Bradshaw, C., Morton, A., Hocking, J., Garland, S., Morris, M., Moss, L., Hor-  
807 vath, L., Kuzevska, I., Fairley, C.: High recurrence rates of bacterial vaginosis

- 808 over the course of 12 months after oral metronidazole therapy and factors asso-  
809 ciated with recurrence. *Journal of Infectious Diseases* **193**(11), 1478–1486 (2006)  
810 <https://doi.org/10.1086/503780>
- 811 [10] Lev-Sagie, A., De Seta, F., Verstraelen, H., Ventolini, G., Lonnee-Hoffmann, R.,  
812 Vieira-Baptista, P.: The vaginal microbiome II: vaginal dysbiotic conditions. *Jour-*  
813 *nal of Lower Genital Tract Disease* **26**(1), 79–84 (2022) <https://doi.org/10.1097/LGT.0000000000000644>
- 815 [11] Zozaya-Hinchliffe, M., Lillis, R., Martin, D.H., Ferris, M.J.: Quantitative PCR  
816 assessments of bacterial species in women with and without bacterial vaginosis.  
817 *Journal of Clinical Microbiology* **48**(5), 1812–9 (2010) <https://doi.org/10.1128/JCM.00851-09>
- 819 [12] Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J., Woyke, T., Kyrpides,  
820 N., Pukall, R., Klenk, H., Goodfellow, M., Göker, M.: Genome-based taxonomic  
821 classification of the phylum actinobacteria. *Frontiers in Microbiology* **9**, 2007  
822 (2018) <https://doi.org/10.3389/fmicb.2018.02007>
- 823 [13] Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S.K., McCulle, S.L.,  
824 Karlebach, S., Gorle, R., Russell, J., Tacket, C.O., Brotman, R.M., Davis, C.C.,  
825 Ault, K., Peralta, L., Forney, L.J.: Vaginal microbiome of reproductive-age  
826 women. *Proceedings of the National Academy of Sciences USA* (108), 4680–4687  
827 (2011) <https://doi.org/10.1073/pnas.100611107>
- 828 [14] France, M., Bing, M., Gajer, P., Brown, S., Humphrys, M., Holm, J., Waetjen,  
829 L., Brotman, R., Ravel, J.: Valencia: a nearest centroid classification method  
830 for vaginal microbial communities based on composition. *Microbiome* **8**(1), 166  
831 (2020) <https://doi.org/10.1186/s40168-020-00934-6>
- 832 [15] Albert, A.Y.K., Chaban, B., Wagner, E.C., Schellenberg, J.J., Links, M.G.,  
833 Van Schalkwyk, J., Reid, G., Hemmingsen, S.M., Hill, J.E., Money, D.: A study of  
834 the vaginal microbiome in healthy Canadian women utilizing cpn60-based molec-  
835 ular profiling reveals distinct Gardnerella subgroup community state types. *PLoS*  
836 *One* **10**(8), 0135620 (2015) <https://doi.org/10.1371/journal.pone.0135620>
- 837 [16] Dos Santos, S.J., Pakzad, Z., Albert, A.Y.K., Elwood, C.N., Grabowska, K.,  
838 Links, M.G., Hutcheon, J.A., Maan, E.J., Manges, A.R., Dumonceaux, T.J.,  
839 Hodgson, Z.G., Lyons, J., Mitchell-Foster, S.M., Gantt, S., Joseph, K.S.,  
840 Van Schalkwyk, J.E., Hill, J.E., Money, D.M.: Maternal vaginal microbiome  
841 composition does not affect development of the infant gut microbiome in early  
842 life. *Frontiers in Cellular and Infection Microbiology* **13**, 1144254 (2023) <https://doi.org/10.3389/fcimb.2023.1144254>
- 844 [17] Symul, L., Jeganathan, P., Costelloe, E.K., France, M., Bloom, S.M., Kwon, D.S.,  
845 Ravel, J., Relman, D.A., Holmes, S.: Sub-communities of the vaginal microbiota in  
846 pregnant and non-pregnant women. *Proceedings of the Royal Society B: Biological*

- 847 Sciences **290**(2011), 20231461 (2023) <https://doi.org/10.1098/rspb.2023.1461>
- 848 [18] Goltzman, D.S.A., Sun, C.L., Proctor, D.M., DiGiulio, D.B., Robaczewska,  
849 A., Thomas, B.C., Shaw, G.M., Stevenson, D.K., Holmes, S.P., Banfield, J.F.,  
850 Relman, D.A.: Metagenomic analysis with strain-level resolution reveals fine-  
851 scale variation in the human pregnancy microbiome. Genome Research **28**(10),  
852 1467–1480 (2018) <https://doi.org/10.1101/gr.236000.118>
- 853 [19] Lev-Sagie, A., Goldman-Wohl, D., Cohen, Y., Dori-Bachash, M., Leshem, A.,  
854 Mor, U., Strahilevitz, J., Moses, A.E., Shapiro, H., Yagel, S., Elinav, E.: Vaginal  
855 microbiome transplantation in women with intractable bacterial vaginosis. Nature  
856 Medicine **25**(10), 1500–1504 (2019) <https://doi.org/10.1038/s41591-019-0600-6>
- 857 [20] Pistner-Nixon, M., Letourneau, J., David, L.A., Lazar, N.A., Mukherjee, S.,  
858 Silverman, J.D.: Scale reliant inference. arXiv (2023) <https://doi.org/10.48550/arXiv.2201.03616>
- 860 [21] Macklaim, J.M., Fernandes, A.D., Di Bella, J.M., Hammond, J.-A., Reid, G.,  
861 Gloor, G.B.: Comparative meta-RNA-seq of the vaginal microbiota and differenti-  
862 al expression by lactobacillus iners in health and dysbiosis. Microbiome **1**(1), 12  
863 (2013) <https://doi.org/10.1186/2049-2618-1-12>
- 864 [22] Ma, B., France, M.T., Crabtree, J., Holm, J.B., Humphrys, M.S., Brotman, R.M.,  
865 Ravel, J.: A comprehensive non-redundant gene catalog reveals extensive within-  
866 community intraspecies diversity in the human vagina. Nature Communications  
867 **11**(1), 940 (2020) <https://doi.org/10.1038/s41467-020-14677-3>
- 868 [23] France, M.T., Fu, L., Rutt, L., Yang, H., Humphrys, M.S., Narina, S., Gajer,  
869 P.M., Ma, B., Forney, L.J., Ravel, J.: Insight into the ecology of vaginal bacte-  
870 ria through integrative analyses of metagenomic and metatranscriptomic data.  
871 Genome Biology **23**(1), 66 (2022) <https://doi.org/10.1186/s13059-022-02635-9>
- 872 [24] Gloor, G.B., Pistner-Nixon, M., Silverman, J.D.: Normalizations are not what  
873 you think: Explicit scale simulation in ALDEx2. bioArxiv (2023) <https://doi.org/10.1101/2023.10.21.563431>
- 875 [25] McGovern, K.C., Nixon, M.P., Silverman, J.D.: Addressing erroneous scale  
876 assumptions in microbe and gene set enrichment analysis. PLoS Computational  
877 Biology **19**(11), 1011659 (2023) <https://doi.org/10.1371/journal.pcbi.1011659>
- 878 [26] Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G., Gloor, G.B.: ANOVA-like  
879 differential expression (ALDEx) analysis for mixed population RNA-seq. PLoS  
880 One **8**(7), 67019 (2013) <https://doi.org/10.1371/journal.pone.0067019>
- 881 [27] Macklaim, J.M., Gloor, G.B.: From RNA-seq to biological inference: using com-  
882 positional data analysis in meta-transcriptomics. Methods in Molecular Biology  
883 **1849**, 193–213 (2018) [https://doi.org/10.1007/978-1-4939-8728-3\\_13](https://doi.org/10.1007/978-1-4939-8728-3_13)

- 884 [28] Deng, Z.-L., Gottschick, C., Bhuju, S., Masur, C., Abels, C., Wagner-  
885 Döbler, I.: Metatranscriptome analysis of the vaginal microbiota reveals  
886 potential mechanisms for protection against metronidazole in bacterial vagi-  
887 nosis. *mSphere* **3**(3) (2018) <https://doi.org/10.1128/mSphereDirect.00262-18>  
888 <https://msphere.asm.org/content/3/3/e00262-18.full.pdf>
- 889 [29] Xiao, M., Lu, B., Ding, R., Liu, X., Wu, X., Li, Y., Liu, X., Qiu, L., Zhang, Z.,  
890 Xie, J., Chen, Y., Zhang, D., Dong, L., Zhang, M., Peng, J., Yang, H., Kudihna,  
891 T., Xu, Y., Li, T., Yi, C., Zhu, L.: Metatranscriptomic analysis of host response  
892 and vaginal microbiome of patients with severe COVID-19. *Science China Life  
893 Sciences* **65**(7), 1473–1476 (2022) <https://doi.org/10.1007/s11427-021-2091-0>
- 894 [30] Gottschick, C., Deng, Z.-L., Vital, M., Masur, C., Abels, C., Pieper, D.H., Rohde,  
895 M., Mendling, W., Wagner-Döbler, I.: Treatment of biofilms in bacterial vagi-  
896 nosis by an amphoteric tenside pessary-clinical study and microbiota analysis.  
897 *Microbiome* **5**(1), 119 (2017) <https://doi.org/10.1186/s40168-017-0326-y>
- 898 [31] Zhang, Y., Parmigiani, G., Johnson, W.E.: ComBat-seq: batch effect adjustment  
899 for RNA-seq count data. *Nucleic Acids Research: Genomics and Bioinformatics*  
900 **2**(3), 078 (2020) <https://doi.org/10.1093/nargab/lqaa078>
- 901 [32] Fernandes, A.D., Reid, J.N., Macklaim, J.M., McMurrough, T.A., Edgell, D.R.,  
902 Gloor, G.B.: Unifying the analysis of high-throughput sequencing datasets: char-  
903 acterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments  
904 by compositional data analysis. *Microbiome* **2**, 15–11513 (2014) <https://doi.org/10.1186/2049-2618-2-15>
- 906 [33] Aitchison, J., Greenacre, M.: Biplots of compositional data. *Journal of the Royal  
907 Statistical Society: Series C (Applied Statistics)* **51**(4), 375–392 (2002) <https://doi.org/10.1111/1467-9876.00275>
- 909 [34] Glascock, A.L., Jimenez, N.R., Boundy, S., Koparde, V.N., Brooks, J.P., Edwards,  
910 D.J., Strauss III, J.F., Jefferson, K.K., Serrano, M.G., Buck, G.A., Fettweis, J.M.:  
911 Unique roles of vaginal *Megasphaera* phylotypes in reproductive health. *Microbial  
912 Genomics* **7**(12), 000526 (2021) <https://doi.org/10.1099/mgen.0.000526>
- 913 [35] Vaneechoutte, M., Guschin, A., Simaey, L.V., Gansemans, Y., Nieuwerburgh,  
914 F.V., Cools, P.: Emended description of *Gardnerella vaginalis* and description  
915 of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella  
916 swidsinskii* sp. nov., with delineation of 13 genomic species within the genus  
917 *Gardnerella*. *International Journal of Systemic and Evolutionary Microbiology*  
918 **69**(3), 679–87 (2019) <https://doi.org/10.1099/ijsem.0.003200>
- 919 [36] Sousa, M., Ksiezarek, M., Perovic, S., Antunes-Lopes, T., Gross, F., Ribeiro,  
920 T., Peixe, L.: *Gardnerella pickettii* sp. nov. (formerly *gardnerella* genomic species  
921 3) and *gardnerella greenwoodii* sp. nov. (formerly *gardnerella* genomic species 8)  
922 isolated from female urinary microbiome. *International Journal of Systematic and*

- 923 Evolutionary Microbiology **73**(11), 006140 (2023) <https://doi.org/10.1099/ijsem.0.006140>
- 925 [37] Yeoman, C.J., Thomas, S.M., Miller, M.E.B., Ulanov, A.V., Torralba, M., Lucas,  
926 S., Gillis, M., Cregger, M., Gomez, A., Ho, M., Leigh, S.R., Stumpf, R., Creedon,  
927 D.J., Smith, M.A., Weisbaum, J.S., Nelson, K.E., Wilson, B.A., White,  
928 B.A.: A multi-omic systems-based approach reveals metabolic markers of bac-  
929 terial vaginosis and insight into the disease. PLoS One **8**(2), 56111 (2013)  
930 <https://doi.org/10.1371/journal.pone.0056111>
- 931 [38] Hardy, L., Jespers, V., Bulck, M.V.d., Buyze, J., Mwambarangwe, L., Musen-  
932 gamana, V., Vaneechoutte, M., Crucitti, T.: The presence of the putative  
933 Gardnerella vaginalis sialidase A gene in vaginal specimens is associated with  
934 bacterial vaginosis biofilm. PLoS One **12**(2), 0172522 (2017) <https://doi.org/10.1371/journal.pone.0172522>
- 936 [39] Neuhaus, F.C., Baddiley, J.: A continuum of anionic charge: structures and  
937 functions of D-alanyl-teichoic acids in Gram-positive bacteria. Microbiology and  
938 Molecular Biology Reviews **67**(4), 686–723 (2003) <https://doi.org/10.1128/mmbr.67.4.686-723.2003>
- 940 [40] Mueller, E.A., Iken, A.G., Ali Öztürk, M., Winkle, M., Schmitz, M., Vollmer, W.,  
941 Di Ventura, B., Levin, P.A.: The active repertoire of Escherichia coli peptidogly-  
942 can amidases varies with physiochemical environment. Molecular Microbiology  
943 **116**(1), 311–328 (2021) <https://doi.org/10.1111/mmi.14711>
- 944 [41] Ernst, C.M., Staubitz, P., Mishra, N.N., Yang, S.-J., Hornig, G., Kalbacher, H.,  
945 Bayer, A.S., Kraus, D., Peschel, A.: The Bacterial Defensin Resistance Protein  
946 MprF Consists of Separable Domains for Lipid Lysinylation and Antimicrobial  
947 Peptide Repulsion. PLoS Pathogens **5**(11), 1000660 (2009) <https://doi.org/10.1371/journal.ppat.1000660>
- 949 [42] Raetz, C.R.H., Reynolds, C.M., Trent, M.S., Bishop, R.E.: Lipid A modification  
950 systems in Gram-negative bacteria. Annual Review of Biochemistry **76**, 295–329  
951 (2007) <https://doi.org/10.1146/annurev.biochem.76.010307.145803>
- 952 [43] Coetzer, T.H.T., Goldring, J.P.D., Huson, L.E.J.: Oligopeptidase B: a processing  
953 peptidase involved in pathogenesis. Biochimie **90**(2), 336–344 (2008) <https://doi.org/10.1016/j.biochi.2007.10.011>
- 955 [44] Strugatsky, D., McNulty, R., Munson, K., Chen, C.-K., Soltis, S.M., Sachs, G.,  
956 Luecke, H.: Structure of the proton-gated urea channel from the gastric pathogen  
957 Helicobacter pylori. Nature **493**(7431), 255–258 (2013) <https://doi.org/10.1038/nature11684>
- 959 [45] Lithgow, K.V., Buchholz, V.C.H., Ku, E., Konschuch, S., D'Aubeterre, A.,

- 960 Sycuro, L.K.: Protease activities of vaginal porphyromonas species disrupt coag-  
961ulation and extracellular matrix in the cervicovaginal niche. *npj Biofilms and*  
962 *Microbiomes* **8**(1), 1–15 (2022) <https://doi.org/10.1038/s41522-022-00270-7>
- 963 [46] Lithgow, K., Buchholz, V., D'Aubeterre, A., Sycuro, L.: Collagenase activity of  
964 vaginal microbes associated with preterm labour. *American Journal of Obstetrics*  
965 *and Gynecology* **226**(2), 302 (2022) <https://doi.org/10.1016/j.ajog.2021.11.1299>
- 966 [47] Chan, C., Ng, D., Fraser, M.E., Schryvers, A.B.: Structural and functional  
967 insights into iron acquisition from lactoferrin and transferrin in Gram-negative  
968 bacterial pathogens. *Biometals* **36**(3), 683–702 (2023) <https://doi.org/10.1007/s10534-022-00466-6>
- 969 [48] Weinberg, E.D.: The Lactobacillus anomaly: total iron abstinence. *Perspectives in*  
970 *Biology and Medicine* **40**(4), 578–583 (1997) <https://doi.org/10.1353/pbm.1997.0072>
- 971 [49] Holm, J.B., France, M.T., Ma, B., McComb, E., Robinson, C.K., Mehta, A.,  
972 Tallon, L.J., Brotman, R.M., Ravel, J.: Comparative metagenome-assembled  
973 genome analysis of “Candidatus Lachnocurva vaginiae”, formerly known as bacte-  
974 rial vaginosis-associated bacterium-1 (BVAB1). *Frontiers in Cellular and Infection*  
975 *Microbiology* **10**, 117 (2020) <https://doi.org/10.3389/fcimb.2020.00117>
- 976 [50] Miller, L.D., Russell, M.H., Alexandre, G.: Diversity in Bacterial Chemotactic  
977 Responses and Niche Adaptation vol. 66, pp. 53–75. Academic Press, ??? (2009).  
978 [https://doi.org/10.1016/s0065-2164\(08\)00803-4](https://doi.org/10.1016/s0065-2164(08)00803-4)
- 979 [51] Armitage, J.P., Berry, R.M.: Assembly and dynamics of the bacterial flagel-  
980 lum. *Annual Review of Microbiology* **74**, 181–200 (2020) <https://doi.org/10.1146/annurev-micro-090816-093411>
- 981 [52] Choi, A.H.K., Slamti, L., Avci, F.Y., Pier, G.B., Maira-Litrán, T.: The pgaABCD  
982 locus of *Acinetobacter baumannii* encodes the production of poly-β-1-6-N-  
983 acetylglucosamine, which is critical for biofilm formation. *Journal of Bacteriology*  
984 **191**(19), 5953–5963 (2009) <https://doi.org/10.1128/jb.00647-09>
- 985 [53] Agladze, K., Wang, X., Romeo, T.: Spatial periodicity of *Escherichia coli* K-12  
986 biofilm microstructure initiates during a reversible, polar attachment phase of  
987 development and requires the polysaccharide adhesin PGA. *Journal of Bacte-*  
988 *riology* **187**(24), 8237–8246 (2005) <https://doi.org/10.1128/jb.187.24.8237-8246.2005>
- 989 [54] Swidsinski, A., Loening-Baucke, V., Mendling, W., Dörffel, Y., Schilling, J., Hal-  
990 wani, Z., Jiang, X., Verstraelen, H., Swidsinski, S.: Infection through structured  
991 polymicrobial *gardnerella* biofilms (StPM-GB). *Histology and Histopathology*  
992 **29**(5), 567–587 (2014) <https://doi.org/10.14670/HH-29.10.567>

- 997 [55] Simoes, J., Discacciati, M., Brolazo, E., Portugal, P., Dini, D., Dantas, M.: Clinical  
998 diagnosis of bacterial vaginosis. International Journal of Gynecology and  
999 Obstetrics **94**(1), 28–32 (2006) <https://doi.org/10.1016/j.ijgo.2006.04.013>
- 1000 [56] Fang, H., Kang, J., Zhang, D.: Microbial production of vitamin B12: a review and  
1001 future perspectives. Microbial Cell Factories **16**(1), 15 (2017) <https://doi.org/10.1186/s12934-017-0631-y>
- 1003 [57] Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A.,  
1004 Gladstone, R.A., Lo, S., Beaudoin, C., Floto, R.A., Frost, S.D.W., Corander, J.,  
1005 Bentley, S.D., Parkhill, J.: Producing polished prokaryotic pangenomes with the  
1006 panaroo pipeline. Genome Biology **21**(1), 180 (2020) <https://doi.org/10.1186/s13059-020-02090-4>
- 1008 [58] Puebla-Barragan, S., Watson, E., Veer, C., Chmiel, J., Carr, C., Burton, J.,  
1009 Sumarah, M., Kort, R., Reid, G.: Interstrain variability of human vaginal *Lac-*  
1010 *tobacillus crispatus* for metabolism of biogenic amines and antimicrobial activity  
1011 against urogenital pathogens. Molecules **26**(15), 4538 (2021) <https://doi.org/10.3390/molecules26154538>
- 1013 [59] Hutzinger, O., Blumich, M.J., v.d.Berg, M., Olie, K.: Sources and fate of PCDDs  
1014 and PCDFs: an overview. Chemosphere **14**(6), 581–600 (1985) [https://doi.org/10.1016/0045-6535\(85\)90167-5](https://doi.org/10.1016/0045-6535(85)90167-5)
- 1016 [60] Nelson, T., Borgogna, J., Brotman, R., Ravel, J., Walker, S., Yeoman, C.: Vaginal  
1017 biogenic amines: biomarkers of bacterial vaginosis or precursors to vaginal  
1018 dysbiosis? Frontiers in Physiology **6**, 253 (2015) <https://doi.org/10.3389/fphys.2015.00253>
- 1020 [61] Kiattiburut, W., Burton, J., Hickling, D.: Immunomodulatory roles of antimicrobials  
1021 peptide LL-37 in urinary bladder bacterial infection. Journal of Urology  
1022 **206**(3), 491 (2021) <https://doi.org/10.1097/JU.0000000000002026.04>
- 1023 [62] Fettweis, J.M., Brooks, J.P., Serrano, M.G., Sheth, N.U., Girerd, P.H., Edwards,  
1024 D.J., Strauss, J.F., Consortium, T.V.M., Jefferson, K.K., Buck, G.A.: Differences  
1025 in vaginal microbiome in african american women versus women of european  
1026 ancestry. Microbiology **160**(10), 2272–2282 (2014) <https://doi.org/10.1099/mic.0.081034-0>
- 1028 [63] Smeianov, V., Scott, K., Reid, G.: Activity of cecropin P1 and FA-LL-37 against  
1029 urogenital microflora. Microbes and Infection **2**(7), 773–777 (2000) [https://doi.org/10.1016/S1286-4579\(00\)90359-9](https://doi.org/10.1016/S1286-4579(00)90359-9)
- 1031 [64] Freitas, A.C., Chaban, B., Bocking, A., Rocco, M., Yang, S., Hill, J.E., Money,  
1032 D.M., VOGUE Research Group, T.: The vaginal microbiome of pregnant women  
1033 is less rich and diverse, with lower prevalence of Mollicutes, compared to non-  
1034 pregnant women. Scientific Reports **7**(1), 9212 (2017) <https://doi.org/10.1038/>

1035 s41598-017-07790-9

- 1036 [65] Romero, R., Hassan, S.S., Gajer, P., Tarca, A.L., Fadrosh, D.W., Nikita, L.,  
1037 Galuppi, M., Lamont, R.F., Chaemsathong, P., Miranda, J., Chaiworapongsa,  
1038 T., Ravel, J.: The composition and stability of the vaginal microbiota of normal  
1039 pregnant women is different from that of non-pregnant women. *Microbiome* **2**, 4  
1040 (2014) <https://doi.org/10.1186/2049-2618-2-4>
- 1041 [66] McGroarty, J.A.: Cell surface appendages of lactobacilli. *FEMS Microbiology Letters*  
1042 **124**(3), 405–409 (1994) <https://doi.org/10.1111/j.1574-6968.1994.tb07316.x>
- 1044 [67] Elli, M., Zink, R., Rytz, A., Reniero, R., Morelli, L.: Iron requirement of Lacto-  
1045 bacillus spp. in completely chemically defined growth media. *Journal of Applied*  
1046 *Microbiology* **88**(4), 695–703 (2000) <https://doi.org/10.1046/j.1365-2672.2000.01013.x>
- 1048 [68] Onderdonk, A.B., Delaney, M.L., Fichorova, R.N.: The human microbiome during  
1049 bacterial vaginosis. *Clinical Microbiology Reviews* **29**(2), 223–238 (2020) <https://doi.org/10.1128/cmr.00075-15>
- 1051 [69] Africa, C.W.J., Nel, J., Stemmet, M.: Anaerobes and bacterial vaginosis in  
1052 pregnancy: virulence factors contributing to vaginal colonisation. *International*  
1053 *Journal of Environmental Research and Public Health* **11**(7), 6979–7000 (2014)  
1054 <https://doi.org/10.3390/ijerph110706979>
- 1055 [70] Al-Mushrif, S., Eley, A., Jones, B.M.: Inhibition of chemotaxis by organic acids  
1056 from anaerobes may prevent a purulent response in bacterial vaginosis. *Journal*  
1057 *of Medical Microbiology* **49**(11), 1023–1030 (2000) <https://doi.org/10.1099/0022-1317-49-11-1023>
- 1059 [71] Hofstad, T.: Pathogenicity of anaerobic Gram-negative rods: possible mecha-  
1060 nisms. *Reviews of Infectious Diseases* **6**(2), 189–199 (1984) <https://doi.org/10.1093/clinids/6.2.189>
- 1062 [72] Boase, S., Jervis-Bardy, J., Cleland, E., Pant, H., Tan, L., Wormald, P.-J.:  
1063 Bacterial-induced epithelial damage promotes fungal biofilm formation in a sheep  
1064 model of sinusitis. *International Forum of Allergy and Rhinology* **3**(5), 341–348  
1065 (2013) <https://doi.org/10.1002/alr.21138>
- 1066 [73] Chegini, Z., Noei, M., Hemmati, J., Arabestani, M.R., Shariati, A.: The  
1067 destruction of mucosal barriers, epithelial remodeling, and impaired mucocili-  
1068 ary clearance: possible pathogenic mechanisms of *Pseudomonas aeruginosa*  
1069 and *Staphylococcus aureus* in chronic rhinosinusitis. *Cell Communication and*  
1070 *Signaling* **21**(1), 306 (2023) <https://doi.org/10.1186/s12964-023-01347-2>
- 1071 [74] Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K.,

- 1072** Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., Mering, C., Bork, P.:  
**1073** eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology  
**1074** resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**(D1),  
**1075** 309–314 (2019) <https://doi.org/10.1093/nar/gky1085>
- 1076** [75] Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M.: KEGG  
**1077** for representation and analysis of molecular networks involving diseases and  
**1078** drugs. *Nucleic Acids Research* **38**(S1), 355–60 (2010) <https://doi.org/10.1093/nar/gkp896>
- 1080** [76] Schwengers, O., Jelonek, L., Dieckmann, M., Beyvers, S., Blom, J., Goesmann, A.:  
**1081** Prokka: rapid prokaryotic genome annotation. *Microbial Genomics* **7**(11), 000685  
**1082** (2021) <https://doi.org/10.1099/mgen.0.000685>