A generalist lifestyle allows rare *Gardnerella* spp. to persist at low levels in the vaginal microbiome

Salahuddin Khan, Sarah J. Vancuren[1] and Janet E. Hill*

Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, S7N 5B4

[1]Current address: Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON

*To whom correspondence should be addressed

52 Campus Drive

Saskatoon, SK. S7N 5B4

E-mail:

SK: salahuddin.khan@usask.ca

SJV: svancure@uoguelph.ca

JEH: Janet.Hill@usask.ca

ORCID:

SK: 0000-0002-6826-9319

SJV: 0000-0002-6145-9670

JEH: 0000-0002-2187-6277

Abstract

Gardnerella spp. are considered a hallmark of bacterial vaginosis, a dysbiosis of the vaginal microbiome. There are four cpn60 sequence-based subgroups within the genus (A, B, C, and D), and thirteen genome species have been defined recently. *Gardnerella* spp. co-occur in the vaginal microbiome with varying abundance, and these patterns are shaped by a resource-dependent, exploitative competition, which affects the growth rate of subgroup A, B, and C negatively. The growth rate of rarely abundant subgroup D, however, increases with the increasing number of competitors, negatively affecting the growth rate of others. We hypothesized that a nutritional generalist lifestyle and minimal niche overlap with the other, more abundant *Gardnerella* spp. facilitate the maintenance of subgroup D in the vaginal microbiome through negative-frequency dependent selection. Using 40 whole genome sequences from isolates representing all four subgroups we found that they could be distinguished based on content of their predicted proteomes. Proteins associated with carbohydrate and amino acid uptake and metabolism were significant contributors to the separation of subgroups. Subgroup D isolates had significantly more of their proteins assigned to amino acid metabolism than the other subgroups. Subgroup D isolates were also significantly different from others in terms of number and type of carbon sources utilized in a phenotypic assay, while the other three could not be distinguished. Overall, the results suggest that a generalist lifestyle and lack of niche overlap with other *Gardnerella* spp. leads to subgroup D being favoured by negative-frequency dependent selection in the vaginal microbiome.

Keywords: *Gardnerella*, vaginal microbiome, negative frequency dependent selection, pangenome, competition

Introduction

Gardnerella spp. are an important diagnostic marker of bacterial vaginosis (BV), a dysbiosis of the vaginal microbiome characterized by a shift from lactobacilli dominated vaginal microbiome to a more diverse microbiome, containing many aerobic and anaerobic bacterial species, including *Gardnerella* spp.. *Gardnerella* is a diverse genus and at least four subgroups (A, B, C, and D) have been identified using cpn60 universal target barcode sequencing [1], which correspond to four clades defined by Ahmed et al. [2]. Recently, *Gardnerella* subgroups have been reclassified into thirteen genome species, of which four are now named as *G. vaginalis* (Subgroup C/Clade 1), *G. swidsinskii* and *G. leopoldii* (Subgroup A/Clade 4), and *G. piotii* (Subgroup B/Clade 2) [3, 4]. These *Gardnerella* species differ in their phenotypic traits, including sialidase activity and vaginolysin production, which may render some of the subgroups more pathogenic than the others [5–7].

Women with vaginal microbiomes dominated by *Gardnerella* are usually colonized by at least two *Gardnerella* spp. [4, 8]. The relative abundances of these co-occurring species, however, are not equal. Subgroup A (*G. swidsinksii* and *G. leopoldii*) and subgroup C (*G. vaginalis*) are most frequently dominant in reproductive aged women [4, 8]. These two subgroups are also often associated with the clinical symptoms of bacterial vaginosis [4, 9, 10]. Subgroup B has been suggested to be associated with intermediate microbiota [7, 9, 10]. Subgroup D, comprised of several unnamed "genome species", has only been detected at low prevalence and abundance [4, 10].

Several factors can affect the abundance and co-occurrence of *Gardnerella* spp. in the vaginal microbiome, including host physiology, host-microbiota interactions, nutrient availability and ecological interactions among bacteria [11, 12]. Ecological interactions are perhaps the most

70 important factors which may affect the co-occurrence and ecological succession of *Gardnerella*
71 species in the vaginal microbiome. Recently, we demonstrated that an indirect, exploitative
72 competition between subgroups of *Gardnerella* is prevalent in co-cultures *in vitro*. While the
73 growth rates of isolates in subgroups A, B, and C, were negatively affected by competition,
74 growth rates of *Gardnerella* subgroup D isolates increased with the increasing number of
75 competing subgroups in co-culture communities [12].

76 The strength of microbial interactions between bacterial species can be affected by niche overlap
77 [13, 14], and species with similar nutritional requirements will naturally compete over the same
78 resources [15]. In addition to competition for nutritional resources, bacteria may also compete
79 for resources essential for colonizing a specific site. Since isolates from *Gardnerella* subgroups
80 A, B and C are negatively affected by competition, and subgroup D isolates experienced a boost
81 in growth rate, the degree of niche overlap between subgroup A, B, and C is presumably higher
82 than between subgroup D and any of the others.

83 Although the growth rate of subgroup D increases in co-cultures, it does not have an intrinsically
84 high growth rate. In fact, the *in vitro* growth rate of subgroup D is half of that of subgroup C,
85 which may contribute to its low abundance in the vaginal microbiome [12]. Low abundance
86 species are often favoured by negative-frequency-dependent selection [16, 17], which can be
87 governed by nutritional requirements [18]. Bacteria capable of utilizing relatively few,
88 abundantly available nutrients in a particular environment are nutritional specialists in the
89 context of that environment. Generalists, on the contrary, are bacteria capable of utilizing more
90 nutrient sources than their specialist counterparts. In negative frequency-dependent selection, the
91 resources accessible to rapidly growing specialists will dwindle, reducing the fitness of the
92 specialists as their population increases. As a result, the population of more generalist bacteria

93 capable of utilizing a wider range of nutrient sources will expand in a density-dependent manner
94 [18, 19]. Generalists can also negatively affect the growth of specialists by competing for the
95 resources that can be utilized by both of them [14].

96 Although growth of the rarely abundant subgroup D is facilitated in co-cultures, the degree of
97 overlap in nutrient utilization among the subgroups and the range of nutrient utilization by
98 individual subgroups are yet unknown. The objective of our present study was, therefore, to
99 evaluate the amount of genomic and phenotypic overlap in nutrient utilization among the
100 subgroups of *Gardnerella* and to determine if subgroup D is a nutritional generalist relative to
101 the three other subgroups. Findings are interpreted in relation to the hypothesis that subgroup D
102 is maintained in the vaginal microbiome through negative frequency dependent selection.

103 **Methods**

104 **Bacterial isolates**

105 Thirty-nine *Gardnerella* isolates from our culture collection representing all four subgroups
106 (based on cpn60 barcode sequencing) were selected for the study (n = 12 subgroup A, 12
107 subgroup B, 8 subgroup C, and 7 subgroup D isolates) (Table S1). Isolates were streaked on
108 Columbia agar plates with 5% (v/v) sheep blood and were incubated anaerobically at 37° C for
109 48 h. For broth culture, colonies from blood agar plates were suspended in BHI broth
110 supplemented with 10% horse serum and 0.25% (w/v) maltose.

111 **Whole-genome sequencing**

112 Whole genome sequences for 10 of the study isolates had been published previously, and the
113 remaining 29 were sequenced as part of the current study (Table S1). DNA was extracted from
114 isolates using a modified salting out protocol [20] and was stored at -20°C. DNA was quantified

115 using Qubit dsDNA BR assay kit (Invitrogen, Burlington, Ontario) and the quality of the extracts
116 was assessed by the A260/A280 ratio. Isolate identity was confirmed by cpn60 barcode
117 sequencing as follows. cpn60 barcode sequences were amplified from extracted DNA with the
118 primers JH0665 (CGC CAG GGT TTT CCC AGT CAC GAC GAY GTT GCA GGY GAY
119 GGH CHA CAA C) and JH0667 (AGC GGA TAA CAA TTT CAC ACA GGA GGR CGA
120 TCR CCR AAK CCT GGA GCY TT). The reaction contained 2 µL template DNA in 1× PCR
121 buffer (0.2 M Tris-HCl at pH 8.4, 0.5 M KCl), 2.5 mM $MgCl_2$, 200 µM dNTP mixture, 400 nM
122 of each primer, 2 U AccuStart Taq DNA polymerase, and water to bring to a final volume of 50
123 µL. PCR was carried out with incubation at 94°C for 30 seconds, 40 cycles of 94°C 30 sec, 60°C
124 for 1 min, 72°C for 1min, and final extension at 72°C for 10 min. PCR products were purified
125 and sequenced by Sanger sequencing and compared with the chaperonin sequence database
126 cpnDB [21] to confirm identity.

127 Following confirmation of the identity of isolates, sequencing libraries were prepared using the
128 Nextera XT DNA library preparation kit according to the manufacturer's instructions (Illumina,
129 Inc., San Diego, CA). PhiX DNA (15% [vol/vol]) was added to the indexed libraries before
130 loading onto the flow cell. The 500 cycle V2 reagent kit was used for the Illumina MiSeq
131 platform (Illumina, Inc.).

132 Raw sequences were trimmed using Trimmomatic [22] with a minimum quality score of 20 over
133 a sliding window of 4, and minimum read length of 40. Trimmed sequences were assembled
134 using SOAPdenovo2 [23] or SPAdes (NR002, NR043, NR044) [24]. Assembled genomes were
135 annotated using the National Center for Biotechnology Information Prokaryotic Genome
136 Annotation Pipeline [25].

137 **Pangenome analysis**

138 Pangenome analysis of the 39 study isolates and the published genome of *G. vaginalis* strain
139 ATCC 14019 (Accession number: PRJNA55487) was performed using the micropan R package
140 [26]. We used "complete" linkage for clustering, and the cut-off value for the generation of
141 clusters was set to 0.75. For initial visualization of the results, the Jaccard index was used to
142 calculate similarity of patterns of presence and absence of protein clusters among all isolates and
143 a dendrogram was constructed from the results by unweighted pair group method with arithmetic
144 mean (UPGMA) using DendroUPGMA (http://genomes.urv.cat/UPGMA/).

145 **COG analysis**

146 Predicted protein sequences from individual genomes were classified into Clusters of
147 Orthologous Groups (COG) categories using WebMEGA
148 (http://weizhonglab.ucsd.edu/webMEGA). Based on the output from this process, the proportion
149 of proteins in each of the COG categories was calculated for each genome. The distributions of
150 proportional abundances of each category were then used to assess the relationships of the four
151 subgroups in terms of COG category representation.

152 **Carbon source utilization assay**

153 Bacterial isolates from freezer stocks were streaked on 5% sheep blood agar plates and were
154 grown for 48 h anaerobically, prior to inoculation of AN Microplates (Biolog Inc, Hayward,
155 CA). Each plate contained 95 carbon sources and one blank well. Colonies of *Gardnerella*
156 isolates were harvested using a sterile swab and suspended in 14 mL of inoculating fluid
157 supplied by the manufacturer. The cell density was adjusted to 55%T (OD_{595} approximately
158 0.25) using a turbidimeter. Each well was filled with 100 µl of culture suspension and was

159 incubated at 35˚ C anaerobically for 48 h. All inoculations and incubations were performed in an
160 anaerobic chamber containing 10 % CO_2, 5% hydrogen, and 85% nitrogen. All plates were read
161 visually after 48h of incubation. If there was no carbon source utilization, the wells remained
162 colourless. A visual change from colourless to purple indicated carbon source utilization. To
163 avoid bias in interpretation, a subset of the plates was read by a second observer who was
164 blinded to the identity of the isolates. There was no disagreement between independent
165 observers. The entire experiment was performed in two biological replicates.

166 **Carbon source profiling of co-cultures**

167 Representative isolates (VN003 of subgroup A, VN002 of subgroup B, NR001 of subgroup C,
168 and WP012 of subgroup D) from the four subgroups were co-cultured in the Biolog AN
169 Microplate in a pairwise fashion (n =6, AB, AC, AD, BC, BD, CD), by combining 50 µL of each
170 isolate suspended in inoculation fluid in each well. The co-cultured AN Microplates were
171 incubated at 35˚C for 48h before being assessed visually for colour change. The experiment was
172 repeated on separate days.

173 **Statistical analysis**

174 The degree of similarity between the isolates in terms of presence/absence of protein clusters
175 generated in the pangenome analysis, proportional abundance of proteins in various COG
176 categories, and carbon source utilization patterns were calculated using the Bray-Curtis
177 dissimilarity matrix. Principle components analysis (PCA) was performed on the distance
178 matrices and significance of relationships were tested using PERMANOVA with the ADONIS
179 function in the *vegan* package [27]. The SIMPER function was used to identify variables driving
180 the differences between groups.

181 One-way ANOVA, student's t-test and chi-square tests were applied to determine if utilization of
182 particular carbon sources was associated with specific subgroups.

183 All statistical analyses were performed in RStudio (version 3.5.2). Figures were generated using
184 GraphPad Prism 8.0 and RStudio (version 3.5.2).

185 **Results**

186 **Overlap between the subgroups based on pangenome and COG analysis**

187 The purpose of our pangenome analysis was to estimate the degree of niche overlap between
188 *Gardnerella* subgroups based on comparisons of their predicted proteomes. Hierarchical
189 clustering using complete linkage produced 4,868 clusters or predicted proteins in the
190 pangenome of the 40 isolates included (rarefaction curve shown in Fig. S1). The strict core
191 (defined as the protein clusters present in all isolates) included 176 clusters. Most of these core
192 proteins were related to metabolism, transcriptional control, DNA replication, and protein
193 synthesis. Clustering of the genomes by subgroup was apparent in a UPGMA dendrogram based
194 on the presence/absence patterns of the 4,868 protein clusters (Fig. 1a). PCA was performed to
195 determine the extent of overlap between the subgroups. The amount of variance explained by the
196 two principal components was 19.4%, based on which, the four subgroups were separable (Fig.
197 1b). The dissimilarity between the four subgroups was significant (pairwise-ADONIS,
198 Bonferroni adjusted, $p < 0.05$, A vs B, $R^2 = 0.45$; A vs C, $R^2 = 0.48$; A vs D, $R^2 = 0.26$; B vs C,
199 $R^2 = 0.34$; B vs D, $R^2 = 0.45$; and C vs D, $R^2 = 0.55$).

200 Following the identification of core and accessory proteins, we investigated the distribution of
201 functional classifications of proteins encoded by isolates in the four subgroups. COG analysis
202 resulted in assignment of predicted proteins into 23 functional categories. As expected,

203 hierarchical clustering of the COG distribution patterns corresponded to subgroup affiliation
204 (Fig. 2a). PCA was performed on the Bray-Curtis dissimilarity matrix and the differences
205 between all subgroups were found to be significant (pairwise ADONIS, Bonferroni adjusted, p
206 <0.05, A vs B, $R^2 = 0.31$; A vs C, $R^2 = 0.71$; A vs D, $R^2 = 0.26$; B vs C, $R^2 = 0.48$; B vs D, $R^2 =$
207 0.20; and C vs D, $R^2 = 0.74$) (Fig. 2b).

208 We identified the variables which caused the four subgroups to diverge in terms of abundance of
209 different COG categories in a multivariate analysis using SIMPER [28]. SIMPER calculates the
210 contribution of each variable to the dissimilarity observed between two groups and relies on
211 Bray-Curtis dissimilarity matrix for calculating the proportion of contribution of each variable
212 being tested. Thirty-six percent of the differences between subgroup A and subgroup B were
213 accounted for by amino acid transport and metabolism (COG category E), inorganic ion transport
214 metabolism (category P), translation, ribosomal structure and biogenesis proteins (category J).
215 The proportion of proteins with functions related to carbohydrate transport and metabolism
216 (category G) was the major factor that differentiated subgroup A from C, contributing to 34% of
217 the dissimilarity observed. Carbohydrate transport and metabolism also accounted for 31% of the
218 dissimilarity observed between subgroups B and C and 36% of the dissimilarity between
219 subgroups C and D. The major contributing factors that differentiated subgroup A and D were
220 proportional abundance of proteins assigned to functional categories H (co-enzyme transport and
221 metabolism) and E (amino acids transport and metabolism) (23%). Subgroups B from D were
222 separated primarily based on functional categories P (inorganic ion transport and metabolism), J
223 (translation, ribosomal structure and biogenesis), G (carbohydrate transport and metabolism
224 proteins), and E (amino acid transport and metabolism proteins), which together accounted for
225 37% of the dissimilarity observed.

226 **Functional categories of proteins differentiating subgroups of *Gardnerella***

227 We tested if the proportions of individual functional categories of proteins that drive the overall
228 separation of the four subgroups in multivariate analysis were significantly different between
229 pairs of *Gardnerella* subgroups. This analysis revealed that subgroup C has a significantly higher
230 proportion of its encoded proteins associated with carbohydrate transport and metabolism and
231 transport, and transcriptional regulation than the other subgroups (unpaired t-test, $p \leq 0.01$,
232 Bonferroni adjusted, Fig. 3a, 3d). The proportion of proteins associated with amino acid
233 transport and metabolism is significantly higher in subgroup D than subgroups A, B, and C
234 (unpaired t-test, $p \leq 0.01$, Bonferroni adjusted, Fig. 3b). Proteins involved in co-enzyme
235 transport and metabolism were found in significantly higher proportional abundance in subgroup
236 A than in subgroup B, C and D (unpaired t-test, $p \leq 0.0001$, Bonferroni adjusted, Fig. 3c).
237 Subgroup B has a significantly higher abundance of proteins associated with inorganic ion
238 transport and metabolism than subgroup A and D (unpaired t-test, $p \leq 0.0001$, Fig. 3e), but the
239 difference between subgroup B and C was not significant. Subgroup B also has a significantly
240 higher proportion of translation, ribosomal structure and biogenesis proteins (unpaired t-test, $p \leq$
241 0.001, Fig. 3f) compared to subgroup C.

242 **Carbon source utilization phenotypes**

243 We hypothesized that subgroup D, a slow-growing, rarely detected *Gardnerella* subgroup is
244 maintained in the vaginal microbiome at a low level and avoids competitive exclusion through
245 negative-frequency-dependent selection, made possible by being a nutritional generalist. We
246 performed carbon source utilization profiling of thirty-six representative isolates (n= 12,
247 subgroup A; n= 9, subgroup B; n=8, subgroup C (including type strain *G. vaginalis* ATCC
248 14018); and n=7, subgroup D). The number of carbon sources utilized by any *Gardnerella* strain

249 ranged from 5 to 24. Only 25% (9/36) of the isolates utilized more than 17 carbon sources,
250 including two subgroup C (NR001, NR038) and all subgroup D isolates. Twenty isolates utilized
251 at least 13 carbon sources, including three subgroup A (3/12, 25%), four subgroup B (4/8, 50%),
252 six subgroup C (6/7, 86%), and all seven isolates of subgroup D (100%). The average number of
253 carbon sources utilized by isolates in subgroups A, B, C, and D was 10.4±3.1, 11.8±1.75,
254 13.9±3.6, and 20.3±1.9, respectively (Fig. 4). A one-way ANOVA was performed to compare
255 the overall difference in carbon sources utilization among the four subgroups showed significant
256 difference among the subgroups ($F(3,32) = 18.15$, $p < 0.05$). A posthoc comparison between the
257 subgroups revealed that the number of carbon sources utilized by subgroup D was significantly
258 higher than subgroup A, B, and C (Tukey HSD, $p < 0.05$). All of the tested *Gardnerella* isolates
259 were able to utilize pyruvic acid, palatinose, and L-rhamnose. The next most frequently utilized
260 carbon sources were D-fructose (32/36, 97%) and L-fucose (32/36, 97%).

261 Overall, 31/95 carbon sources were utilized by at least one isolate, and the majority (20/31) of
262 these were sugars (mono- or oligosaccharides). Together, subgroup D isolates (n = 7) utilized
263 more of the sugar substrates (18/37 available) than any other subgroup, including subgroup C
264 (n=8), which utilized 15/37 available sugars. Utilization of any of the 11 available amino acids
265 was rarely observed, with only two of the subgroup C isolates positive for L-methionine or L-
266 valine utilization.

267 **Overlap in carbon sources utilization among the subgroups**

268 To determine if subgroups could be distinguished based on carbon source utilization profiles, a
269 principal component analysis was performed (Fig. 5). The overlap between the representative
270 isolates of subgroups A, B, and C was significant. Subgroup D was significantly dissimilar to

271 subgroups A and B (Fig. 5, pairwise-ADONIS, A vs D, $R^2 = 0.55$; B vs D, $R^2 = 0.55$, $p<0.05$).
272 Although the dissimilarity between subgroup C and D was not statistically significant after
273 Bonferroni adjustment, 39% (pairwise ADONIS, C vs D, $R^2= 0.39$) of the variation in carbon
274 source utilization could be explained by subgroup affiliation of the tested isolates, which was
275 higher than between subgroups A, B and C (A vs B: 13%, A vs. C: 21%, and B vs. C: 8%).

276 **Association of carbon source utilization pattern with subgroups**

277 To identify carbon sources that differentiate the subgroups, we selected twelve substrates that
278 were utilized by more than five isolates but fewer than thirty isolates. Chi-square tests were
279 performed to determine if the subgroups significantly differ in the utilization of those twelve
280 carbon sources. The four *Gardnerella* subgroups differed in their use of 3 of the 12 carbon
281 sources: turanose, inosine, and uridine 5-monophosphate (Chi-square test, $p <0.05$, Bonferroni
282 adjusted) (Fig. 6). For each of these three carbon sources, subgroups A and B had low frequency
283 of use (9.5% = 2/21, 0.0% = 0/21, 0.0% = 0/21; subgroups A and B combined), subgroup C had
284 low or intermediate frequency of use (25.0% = 2/8, 50.0% = 4/8, and 62.5% = 5/8), whereas
285 subgroup D had high frequency of use (100.0% = 7/7, 100.0% = 7/7, 100.0% = 7/7).

286 **Carbon source utilization by co-cultured isolates**

287 Since the four subgroups co-exist in the same ecosystem, it is possible that mixing them might
288 facilitate the utilization of certain carbon sources. To detect any such facilitation in carbon
289 sources utilization, we co-cultured isolates from all four subgroups in six pairwise combinations
290 (A+B, A+C, A+D, B+C, B+D, and C+D). The representative isolates of subgroups A-D utilized
291 11, 13, 19 and 24 carbon sources, respectively, when grown alone while co-cultures utilized from
292 12 to a maximum of 22 carbon sources (Table 1). In every case, the co-culture utilized fewer
293 carbon sources than the isolate that utilized the most carbon sources on its own.

294

295 **Discussion**

296 Rarely abundant species can be maintained in the human microbiome through a variety of
297 mechanisms, which include but are not limited to sequestration of essential nutrients from
298 competing species, diversification of phenotype [29], social cheating [30], and negative
299 frequency dependent selection [17]. Differences in nutrient utilization among community
300 members can be a key factor that sets the stage for negative frequency dependent selection [31].
301 The reproductive fitness of nutritional specialist species will remain high as long as the supply of
302 nutrients usable by the specialists is abundant. As soon as the supply of these nutrients drops,
303 slower-growing generalists, by virtue of their greater utilization capacity, will have increased
304 fitness, which will eventually lead to their dominance in the absence of any other negative
305 influences.

306 Among the four subgroups of *Gardnerella* spp. that colonize the vaginal microbiome of
307 reproductive-aged women, subgroup D is the rarest in terms of abundance and prevalence among
308 women [4]. Subgroup D is also relatively slow-growing, yet shows an increased growth rate
309 when the number of competitors in an *in vitro* community increases [12]. We have reported
310 previously that resource-based competition is common among *Gardnerella* spp. and no evidence
311 for contact-dependent interaction was observed. Therefore, we set out to investigate if negative-
312 frequency dependent selection is responsible for persistence of subgroup D, which would require
313 relatively small niche overlap and a more generalist lifestyle than the other *Gardnerella* spp. in
314 the vaginal microbiome.

315 **Predicted niche overlap between the four subgroups**

316 Niche overlap may lead to competition for nutrients and space [31–34] and it has been reported
317 that competition is prevalent among metabolically similar bacterial species [14]. Occupying
318 distinct niches can therefore help bacterial species avoid competition for space, growth factors,
319 and nutrients, resulting in increased reproductive fitness. Since subgroup D isolates have higher
320 growth rates *in vitro* in the presence of competitors compared to when grown alone, these
321 isolates presumably occupy a distinct niche. The pangenome analysis showed that the four
322 subgroups differ significantly based on the composition of their predicted proteomes (Fig. 1),
323 with only 176 proteins comprising the strict core of proteins found in all isolates. This finding is
324 not surprising since the genetic diversity among *Gardnerella* is well established, and genome
325 sequence comparisons formed the basis for the recent reclassification of *Gardnerella* into 13
326 genome species [1–3].

327 Comparisons of the entire predicted proteomes do not, however, focus on the key factors for a
328 resource-based competition: nutrient utilization potential. Proteins involved in nutrient uptake
329 and metabolism account for only a fraction of the 4,868 protein clusters comprising the
330 pangenome. Analysis of the distribution of various functional (COG) categories of proteins
331 revealed significant differences among subgroups in their predicted capacity to utilize
332 carbohydrates and amino acids (Fig. 3a, 3b), with subgroup D having significantly more of its
333 proteome dedicated to amino acid transport and metabolism than any of the other subgroups.
334 Since a resource-based competition encapsulates competition for space, growth factors and
335 nutrients, our findings from the pangenome and COG analyses suggest that the competition
336 among the four subgroups is not spatial but may be primarily for nutrients; a speculation
337 supported by the previous observation that *Gardnerella* spp. form multi-subgroup biofilms [12].

338 **Subgroup D is a nutritional generalist relative to subgroup A, B and C**

339 The diversity of nutrients available to microbiota in the vaginal microbiome is less than in the
340 gastrointestinal microbiome, where food intake provides a constant source of diverse nutrients
341 that affect the assembly of gut microbiota [35, 36]. Vaginal microbiota, on the contrary, are
342 largely dependent upon host-derived nutrients, the most abundant of which is glycogen.
343 Glycogen is deposited in the vaginal lumen by epithelial cells under the influence of estrogen
344 [37], and is digested into maltooligosaccharides, maltodextrins and glucose by the combined
345 activities of host and microbial enzymes prior to uptake and metabolism by the microbiota [38–
346 40]. Given the relatively narrow range of nutrients available in the vaginal microbiome, it is
347 expected that the resident microbiota, including the four subgroups of *Gardnerella,* overlap to a
348 considerable extent in their nutrient utilization capacity, resulting in some level of competition
349 among them [32, 36, 41]. As discussed earlier, subgroup D is an exception since the growth of
350 these isolates was actually facilitated in co-cultures, suggesting that while it may compete with
351 other *Gardnerella* spp. over common nutrients like the breakdown products of glycogen, it may
352 be able to utilize a greater overall diversity of nutrients (i.e. it is a generalist).

353 The AN microplate assay results showed that subgroup D isolates utilized more of the provided
354 carbon sources than isolates in the three other subgroups (Fig. 4). Furthermore, when the patterns
355 of substrate use were considered, subgroup A, B and C were not separable from each other, but
356 subgroup D was significantly different (Fig. 5). The distinct pattern observed in subgroup D was
357 partially driven by utilization of three particular substrates: turanose, inosine, and uridine 5-
358 monophosphate (Fig. 6). Turanose is an isomer of sucrose, known as a non-accumulative
359 osmoprotectant, aiding bacterial growth at high osmolarity [42]. The importance of turanose
360 utilization in the vaginal environment is not known yet, but our observation is an indication that
361 subgroup D isolates can metabolize sucrose-like sugars. The two other carbon sources: inosine

362 and uridine 5-monophosphate are probably used in purine and pyrimidine biosynthesis in
363 *Gardnerella* spp..

364 Some findings of the pangenome and COG analyses could not be reconciled with the phenotypic
365 carbon source utilization assay. For example, although subgroup C and D have higher
366 proportions of their proteomes predicted to be involved in transport and metabolism of
367 carbohydrates (Category G) and amino acids (Category E), respectively, than the other
368 subgroups, subgroup C isolates did not utilize the greatest number of available sugar substrates
369 in the AN microplate and subgroup D isolates did not utilize any of the amino acid substrates
370 available. It is, however, important to consider that the carbon source utilization assay was
371 performed in a plastic environment and included only 95 substrates, many of which are not
372 relevant to the vaginal microbiome. More relevant amino acid sources available in the vagina,
373 including those whose abundance is altered in bacterial vaginosis, such as isoleucine, leucine,
374 proline, and tryptophan, are not included [43–45]. Ideally, this study would have involved a
375 vaginal-microbiome specific nutrient panel, but such reagents were not available. Even with this
376 limitation, our results suggest that subgroup D is a nutritional generalist relative to other
377 *Gardnerella* spp.. Most of the ecological studies that have been performed to date to elucidate
378 mechanisms shaping the assembly of bacterial communities have included either environmental
379 bacterial species or well-characterized model organisms [14, 16, 29, 36, 41, 46–50]. There are
380 understandably fewer studies that focus on interactions among host-associated microbiota [51,
381 52].

382 **Negative-frequency-dependent selection in the vaginal microbiome**

383 The genomic and phenotypic differences we observed between subgroup D and the three other
384 subgroups, including the potential to utilize more amino acids, use of a greater number of carbon

385 sources, and a distinct pattern of substrate utilization, suggest that subgroup D is a candidate for
386 negative frequency dependent selection. Why then are these *Gardnerella* spp. only observed
387 rarely, and in low abundance in reproductive aged women? Among 413 vaginal samples from
388 reproductive aged Canadian women, genome species comprising subgroup D of *Gardnerella*
389 were detected in <10% of samples and never accounted for more than 5% of the microbiota [4].
390 Vaginal environmental dynamics and related host factors, such as menstruation, sloughing of
391 epithelial cells, and fluctuating pH contribute to the turnover of bacterial species, shifting the
392 bacterial population density and changing the nutrients available [53]. A decline in population
393 density would reshuffle the vaginal ecosystem, increasing the supply of abundant nutrients
394 accessible to faster growing, specialists, and checking the growth of slower growing generalist
395 subgroup D.

396 Although subgroup D is likely rare due the factors described above, it could still be a major
397 player in ecological succession and transition of vaginal microbiota between a *Lactobacillus*
398 dominated community and the overgrowth of anaerobes characteristic of bacterial vaginosis.
399 These organisms may also play a particular role in biofilm formation or competition for
400 occupancy of the vaginal mucosa. Rarely abundant species often act as keystone species helping
401 colonization by other bacterial species, which are also essential to maintain homeostasis of an
402 ecosystem [54–57]. Resolution of the role of low abundant *Gardnerella* spp. will depend on the
403 development and application of experimental systems that more closely model the human
404 vaginal microbiome. Rodent models have shown some promise, especially for studies of specific
405 combinations of organisms [58], but there is also potential in bioreactors [59], and cell and tissue
406 culture systems that attempt to recapitulate many of the environmental and physiological aspects

407 of the vaginal microbiome [60]. Further study of rare *Gardnerella* spp. will likely also result in
408 the definition of additional species within this diverse genus.

409 **Acknowledgements**

410 The authors are grateful to Champika Fernando for excellent technical support and to all
411 members of the Hill Lab for their review of the manuscript. Special thanks to Prairie Diagnostic
412 Services for access to their anaerobic chamber. No thanks to COVID-19.

413 **Declarations**

414 **Funding**

415 The research was supported by a Natural Sciences and Engineering Research Council of Canada
416 Discovery Grant to JEH.

417 **Competing interests**

418 None declared.

419 **Availability of data and material**

420 NCBI Bioproject accession numbers for all genome sequence data are included in Table S1.

421 **Authors' contributions**

422 Conceived and designed the study: Salahuddin Khan and Janet E. Hill. Performed the
423 experiments: Salahuddin Khan and Sarah J. Vancuren. Analysed the data: Salahuddin Khan,
424 Sarah J. Vancuren, Janet E. Hill. Wrote and revised the manuscript: Salahuddin Khan, Sarah J.
425 Vancuren, Janet E. Hill.

426 **References**

427 1. Paramel Jayaprakash T, Schellenberg JJ, Hill JE (2012) Resolution and characterization of
428 distinct cpn60-based subgroups of *Gardnerella vaginalis* in the vaginal microbiota. PLoS
429 ONE 7:e43009. https://doi.org/10.1371/journal.pone.0043009

430 2. Ahmed A, Earl J, Retchless A, et al (2012) Comparative genomic analyses of 17 clinical
431 isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades
432 consistent with subspeciation into genovars. J Bacteriol 194:3922–37.
433 https://doi.org/10.1128/JB.00056-12

434 3. Vaneechoutte M, Guschin A, Van Simaey L, et al (2019) Emended description of
435 *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piotii*
436 sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within
437 the genus *Gardnerella*. Int J Syst Evol Microbiol 69:679–687.
438 https://doi.org/10.1099/ijsem.0.003200

439 4. Hill JE, Albert AYK (2019) Resolution and co-occurrence patterns of *Gardnerella leopoldii,*
440 *Gardnerella swidsinskii, Gardnerella piotii* and *Gardnerella vaginalis* within the vaginal
441 microbiome. Infect Immun 87:e00532-19. https://doi.org/10.1128/IAI.00532-19

442 5. Janulaitiene M, Gegzna V, Baranauskiene L, et al (2018) Phenotypic characterization of
443 *Gardnerella vaginalis* subgroups suggests differences in their virulence potential. PLOS
444 ONE 13:e0200625. https://doi.org/10.1371/journal.pone.0200625

445 6. Yeoman CJ, Yildirim S, Thomas SM, et al (2010) Comparative genomics of *Gardnerella*
446 *vaginalis* strains reveals substantial differences in metabolic and virulence potential. PLoS
447 ONE 5:e12411. https://doi.org/10.1371/journal.pone.0012411

448 7. Schellenberg JJ, Paramel Jayaprakash T, Withana Gamage N, et al (2016) *Gardnerella*
449 *vaginalis* subgroups defined by cpn60 sequencing and sialidase activity in isolates from
450 Canada, Belgium and Kenya. PLoS One 11:e0146510.
451 https://doi.org/10.1371/journal.pone.0146510

452 8. Albert AY, Chaban B, Wagner EC, et al (2015) A study of the vaginal microbiome in
453 healthy Canadian women utilizing cpn60-based molecular profiling reveals distinct
454 *Gardnerella* subgroup community state types. PLoS One 10:e0135620.
455 https://doi.org/10.1371/journal.pone.0135620

456 9. Hilbert DW, Schuyler JA, Adelson ME, et al (2017) *Gardnerella vaginalis* population
457 dynamics in bacterial vaginosis. Eur J Clin Microbiol Infect Dis.
458 https://doi.org/10.1007/s10096-017-2933-8

459 10. Shipitsyna E, Krysanova A, Khayrullina G, et al (2019) Quantitation of all four *Gardnerella*
460 *vaginalis* clades detects abnormal vaginal microbiota characteristic of bacterial vaginosis
461 more accurately than putative *G. vaginalis* sialidase A gene count. Mol Diagn Ther 23:139–
462 147. https://doi.org/10.1007/s40291-019-00382-5

463 11. Schellenberg JJ, Patterson MH, Hill JE (2017) *Gardnerella vaginalis* diversity and ecology
464 in relation to vaginal symptoms. Res Microbiol doi:10.1016/j.resmic.2017.02.011:
465 https://doi.org/10.1016/j.resmic.2017.02.011

466 12. Khan S, Voordouw MJ, Hill JE (2019) Competition among *Gardnerella* subgroups from the
467 human vaginal microbiome. Front Cell Infect Microbiol 9:374.
468 https://doi.org/10.3389/fcimb.2019.00374

469 13. Madsen JS, Røder HL, Russel J, et al (2016) Coexistence facilitates interspecific biofilm
470 formation in complex microbial communities. Environ Microbiol 18:2565–2574.
471 https://doi.org/10.1111/1462-2920.13335

472 14. Russel J, Røder HL, Madsen JS, et al (2017) Antagonism correlates with metabolic similarity
473 in diverse bacteria. Proc Natl Acad Sci U S A 114:10684–10688.
474 https://doi.org/10.1073/pnas.1706016114

475 15. Darwin C (1859) On the origin of species by means of natural selection, or preservation of
476 favoured races in the struggle for life. John Murray, London

477 16. Hibbing ME, Fuqua C, Parsek MR, Peterson SB (2010) Bacterial competition: surviving and
478 thriving in the microbial jungle. Nat Rev Microbiol 8:15–25.
479 https://doi.org/10.1038/nrmicro2259

480 17. Levin BR (1988) Frequency-dependent selection in bacterial populations. Philos Trans R Soc
481 Lond B Biol Sci 319:459–472. https://doi.org/10.1098/rstb.1988.0059

482 18. Kurihara Y, Shikano S, Toda M (1990) Trade-off between interspecific competitive ability
483 and growth rate in bacteria. Ecology 71:645–650. https://doi.org/10.2307/1940318

484 19. Ross-Gillespie A, Gardner A, Buckling A, et al (2009) Density dependence and cooperation:
485 theory and a test with bacteria. Evolution 63:2315–2325. https://doi.org/10.1111/j.1558-
486 5646.2009.00723.x

487 20. Martin-Platero AM, Valdivia E, Maqueda M, Martinez-Bueno M (2007) Fast, convenient,
488 and economical method for isolating genomic DNA from lactic acid bacteria using a
489 modification of the protein "salting-out" procedure. Anal Biochem 366:102–4.
490 https://doi.org/S0003-2697(07)00159-5 [pii] 10.1016/j.ab.2007.03.010

491 21. Vancuren SJ, Hill JE (2019) Update on cpnDB: a reference database of chaperonin
492 sequences. Database 2019:baz033 https://doi.org/10.1093/database/baz033

493 22. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina
494 sequence data. Bioinformatics 30:2114–20. https://doi.org/10.1093/bioinformatics/btu170

495 23. Luo R, Liu B, Xie Y, et al (2012) SOAPdenovo2: an empirically improved memory-efficient
496 short-read *de novo* assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18

24. Prjibelski A, Antipov D, Meleshko D, et al (2020) Using SPAdes De Novo Assembler. Curr Protoc Bioinforma 70:e102. https://doi.org/10.1002/cpbi.102

25. Tatusova T, DiCuccio M, Badretdin A, et al (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

26. Snipen L, Liland KH (2015) micropan: an R-package for microbial pan-genomics. BMC Bioinformatics 16:79. https://doi.org/10.1186/s12859-015-0517-0

27. Oksanen J, Blanchet FG, Kindt R, et al (2012) vegan: Community Ecology Package. R package version 2.0-10. Available: http://CRAN.R-project.org/package=vegan

28. Whitfield-Cargile CM, Cohen ND, Suchodolski J, et al (2015) Composition and diversity of the fecal microbiome and inferred fecal metagenome does not predict subsequent pneumonia caused by *Rhodococcus equi* in foals. PLOS ONE 10:e0136586. https://doi.org/10.1371/journal.pone.0136586

29. Martín PV, Muñoz MA, Pigolotti S (2019) Bet-hedging strategies in expanding populations. PLOS Comput Biol 15:e1006529. https://doi.org/10.1371/journal.pcbi.1006529

30. Chen R, Déziel E, Groleau M-C, et al (2019) Social cheating in a *Pseudomonas aeruginosa* quorum-sensing variant. Proc Natl Acad Sci 116:7021–7026. https://doi.org/10.1073/pnas.1819801116

31. Freter R, Brickner H, Botney M, et al (1983) Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. Infect Immun 39:676–685. https://doi.org/10.1128/IAI.39.2.676-685.1983

32. Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. Proc Natl Acad Sci 106:19659–19665. https://doi.org/10.1073/pnas.0905137106

33. Hutchinson GE (1957) Concluding Remarks. Cold Spring Harb Symp Quant Biol 22:415–427. https://doi.org/10.1101/SQB.1957.022.01.039

34. Amarasekare P (2002) Interference competition and species coexistence. Proc R Soc Lond B Biol Sci 269:2541–2550. https://doi.org/10.1098/rspb.2002.2181

35. Graf D, Di Cagno R, Fåk F, et al (2015) Contribution of diet to the composition of the human gut microbiota. Microb Ecol Health Dis 26:26164. https://doi.org/10.3402/mehd.v26.26164

36. Smith CC, Snowberg LK, Gregory Caporaso J, et al (2015) Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. ISME J 9:2515–2526. https://doi.org/10.1038/ismej.2015.64

37. Mirmonsef P, Hotton AL, Gilbert D, et al (2014) Free glycogen in vaginal fluids is associated with *Lactobacillus* colonization and low vaginal pH. PLoS One 9:e102467. https://doi.org/10.1371/journal.pone.0102467

532 38. Bhandari P, Tingley JP, Abbott DW, Hill JE (2020) Characterization of an α-glucosidase
533 enzyme conserved in *Gardnerella* spp. isolated from the human vaginal microbiome.
534 bioRxiv 2020.05.11.086124. https://doi.org/10.1101/2020.05.11.086124

535 39. Nunn KL, Forney LJ (2016) Unraveling the dynamics of the human vaginal microbiome.
536 Yale J Biol Med 89:331–337

537 40. Spear GT, French AL, Gilbert D, et al (2014) Human alpha-amylase present in lower-
538 genital-tract mucosal fluid processes glycogen to support vaginal colonization by
539 *Lactobacillus*. J Infect Dis 210:1019–1028. https://doi.org/10.1093/infdis/jiu231

540 41. Bauer MA, Kainz K, Carmona-Gutierrez D, Madeo F (2018) Microbial wars: Competition in
541 ecological niches and within the microbiome. Microb Cell 5:215–219.
542 https://doi.org/10.15698/mic2018.05.628

543 42. Gouffi K, Blanco C (2000) Is the accumulation of osmoprotectant the unique mechanism
544 involved in bacterial osmoprotection? Int J Food Microbiol 55:171–174.
545 https://doi.org/10.1016/s0168-1605(00)00192-6

546 43. Srinivasan S, Morgan MT, Fiedler TL, et al (2015) Metabolic signatures of bacterial
547 vaginosis. mBio 6:. https://doi.org/10.1128/mBio.00204-15

548 44. Vicente-Muñoz S, Cobo T, Puchades-Carrasco L, et al (2020) Vaginal metabolome: towards
549 a minimally invasive diagnosis of microbial invasion of the amniotic cavity in women with
550 preterm labor. Sci Rep 10:5465. https://doi.org/10.1038/s41598-020-62542-6

551 45. Vitali B, Cruciani F, Picone G, et al (2015) Vaginal microbiome and metabolome highlight
552 specific signatures of bacterial vaginosis. Eur J Clin Microbiol Infect Dis 34:2367–2376.
553 https://doi.org/10.1007/s10096-015-2490-y

554 46. Stubbendieck RM, Straight PD (2015) Escape from lethal bacterial competition through
555 coupled activation of antibiotic resistance and a mobilized subpopulation. PLOS Genet
556 11:e1005722. https://doi.org/10.1371/journal.pgen.1005722

557 47. Gonzalez DJ, Haste NM, Hollands A, et al (2011) Microbial competition between *Bacillus*
558 *subtilis* and *Staphylococcus aureus* monitored by imaging mass spectrometry. Microbiology
559 157:2485–2492. https://doi.org/10.1099/mic.0.048736-0

560 48. Griffin AS, West SA, Buckling A (2004) Cooperation and competition in pathogenic
561 bacteria. Nature 430:1024–1027. https://doi.org/10.1038/nature02744

562 49. Oliveira NM, Martinez-Garcia E, Xavier J, et al (2015) Biofilm formation as a response to
563 ecological competition. PLoS Biol 13:e1002191.
564 https://doi.org/10.1371/journal.pbio.1002191

565 50. Rezzoagli C, Granato ET, Kümmerli R (2020) Harnessing bacterial interactions to manage
566 infections: a review on the opportunistic pathogen *Pseudomonas aeruginosa* as a case
567 example. J Med Microbiol 69:147–161. https://doi.org/10.1099/jmm.0.001134

568 51. García-Bayona L, Comstock LE (2018) Bacterial antagonism in host-associated microbial
569 communities. Science 361:. https://doi.org/10.1126/science.aat2456

570 52. Rakoff-Nahoum S, Foster KR, Comstock LE (2016) The evolution of cooperation within the
571 gut microbiota. Nature 533:255–259. https://doi.org/10.1038/nature17626

572 53. Abreu CI, Friedman J, Andersen Woltz VL, Gore J (2019) Mortality causes universal
573 changes in microbial community composition. Nat Commun 10:2120.
574 https://doi.org/10.1038/s41467-019-09925-0

575 54. Banerjee S, Schlaeppi K, van der Heijden MGA (2018) Keystone taxa as drivers of
576 microbiome structure and functioning. Nat Rev Microbiol 16:567–576.
577 https://doi.org/10.1038/s41579-018-0024-1

578 55. Herren CM, McMahon KD (2018) Keystone taxa predict compositional change in microbial
579 communities. Environ Microbiol 20:2207–2217. https://doi.org/10.1111/1462-2920.14257

580 56. Hajishengallis G, Liang S, Payne MA, et al (2011) A low-abundance biofilm species
581 orchestrates inflammatory periodontal disease through the commensal microbiota and the
582 complement pathway. Cell Host Microbe 10:497–506.
583 https://doi.org/10.1016/j.chom.2011.10.006

584 57. Maldonado-Contreras A, Goldfarb KC, Godoy-Vitorino F, et al (2011) Structure of the
585 human gastric bacterial community in relation to *Helicobacter pylori* status. ISME J 5:574–
586 579. https://doi.org/10.1038/ismej.2010.149

587 58. Teixeira GS, Carvalho FP, Arantes RM, et al (2012) Characteristics of *Lactobacillus* and
588 *Gardnerella vaginalis* from women with or without bacterial vaginosis and their
589 relationships in gnotobiotic mice. J Med Microbiol 61:1074–81.
590 https://doi.org/10.1099/jmm.0.041962-0

591 59. Guzman-Rodriguez M, McDonald JAK, Hyde R, et al (2018) Using bioreactors to study the
592 effects of drugs on the human microbiota. Methods 149:31–41.
593 https://doi.org/10.1016/j.ymeth.2018.08.003

594 60. Herbst-Kralovetz MM, Pyles RB, Ratner AJ, et al (2016) New systems for studying
595 intercellular interactions in bacterial vaginosis. J Infect Dis 214:S6–S13.
596 https://doi.org/10.1093/infdis/jiw130

597

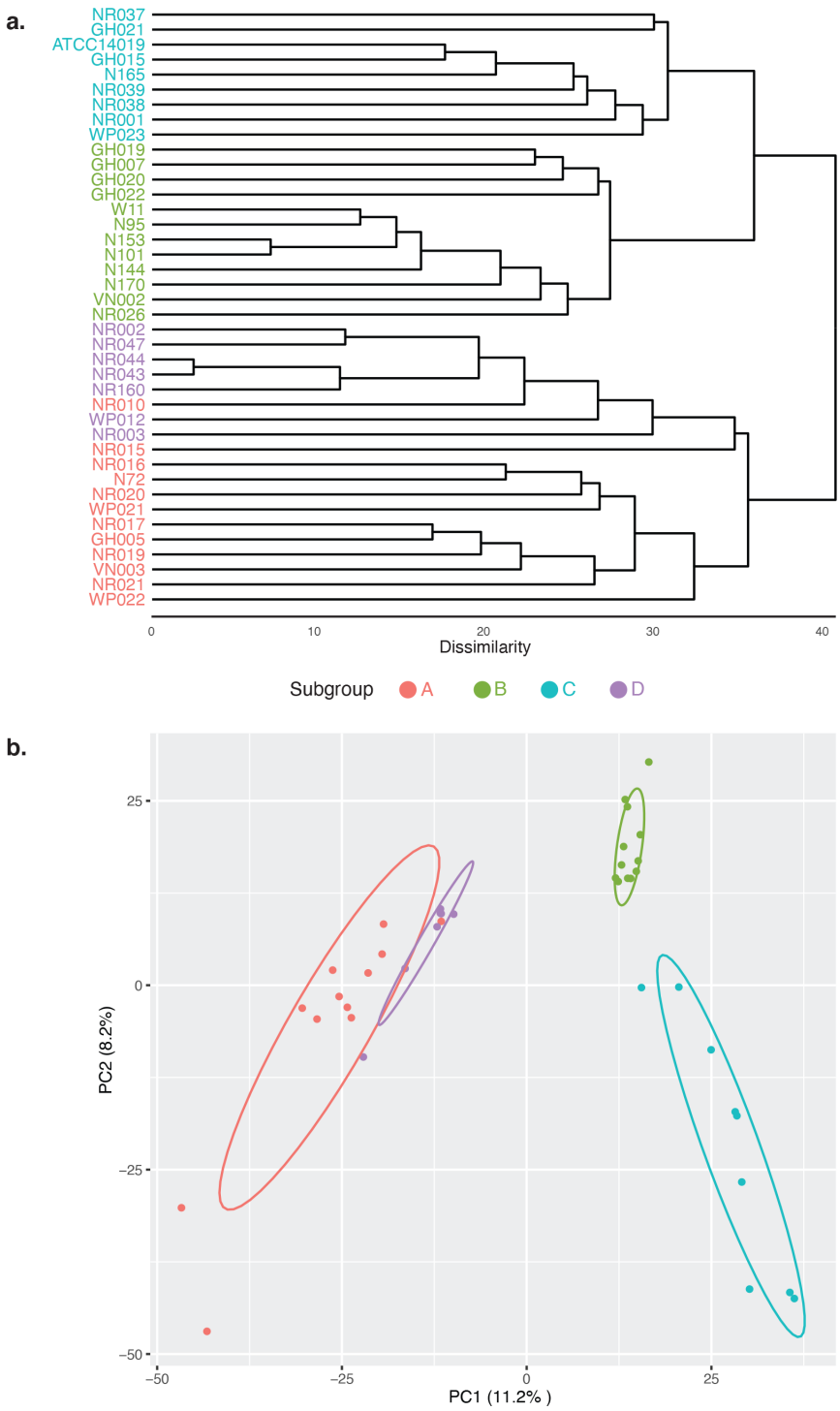

Fig. 1 Comparison of predicted proteomes of study isolates. **(a)** UPGMA dendrogram based on presence/absence of protein clusters in the predicted proteomes of *Gardnerella* isolates. **(b)** Principle components analysis (PCA) of Bray-Curtis dissimilarity matrices calculated from protein cluster distributions. Dissimilarity between the four subgroups is significant (pairwise ADONIS $p < 0.05$, Bonferroni adjusted). Subgroup affiliations of isolates are indicated by colour as shown in the legend between the panels.

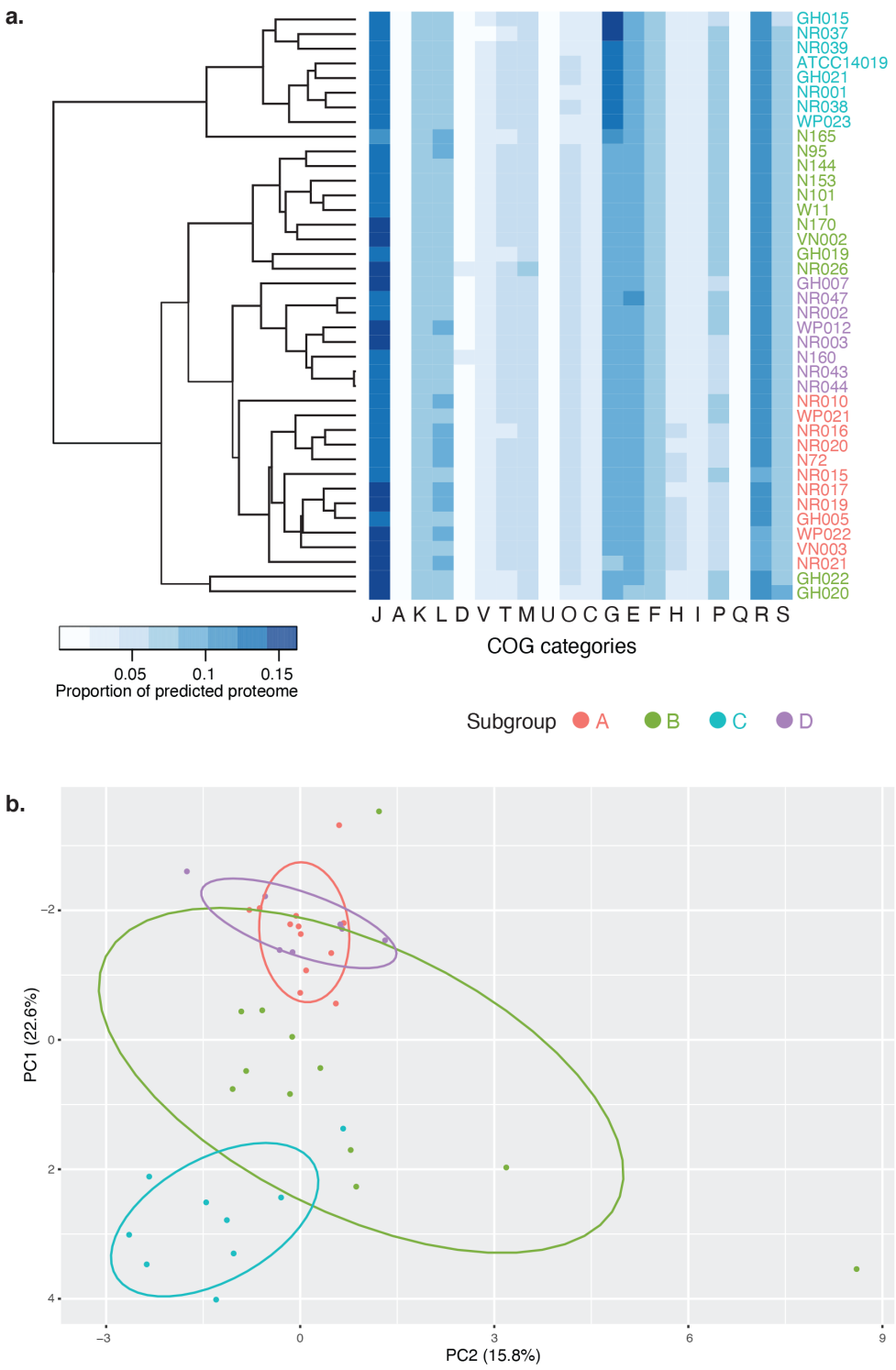

Fig. 2 COG analysis of predicted proteomes. **(a)** Hierarchical clustering of study isolates based on proportional abundance of COG categories in their predicted proteomes. Abundance is indicated by blue colour intensity in the heat map. **(b)** PCA of Bray-Curtis dissimilarity matrices calculated from the proportional abundance data. Dissimilarity between the four subgroups is significant (pairwise ADONIS, $p < 0.05$, Bonferroni adjusted). Subgroup affiliations of isolates are indicated by colour as shown in the legend between the panels.

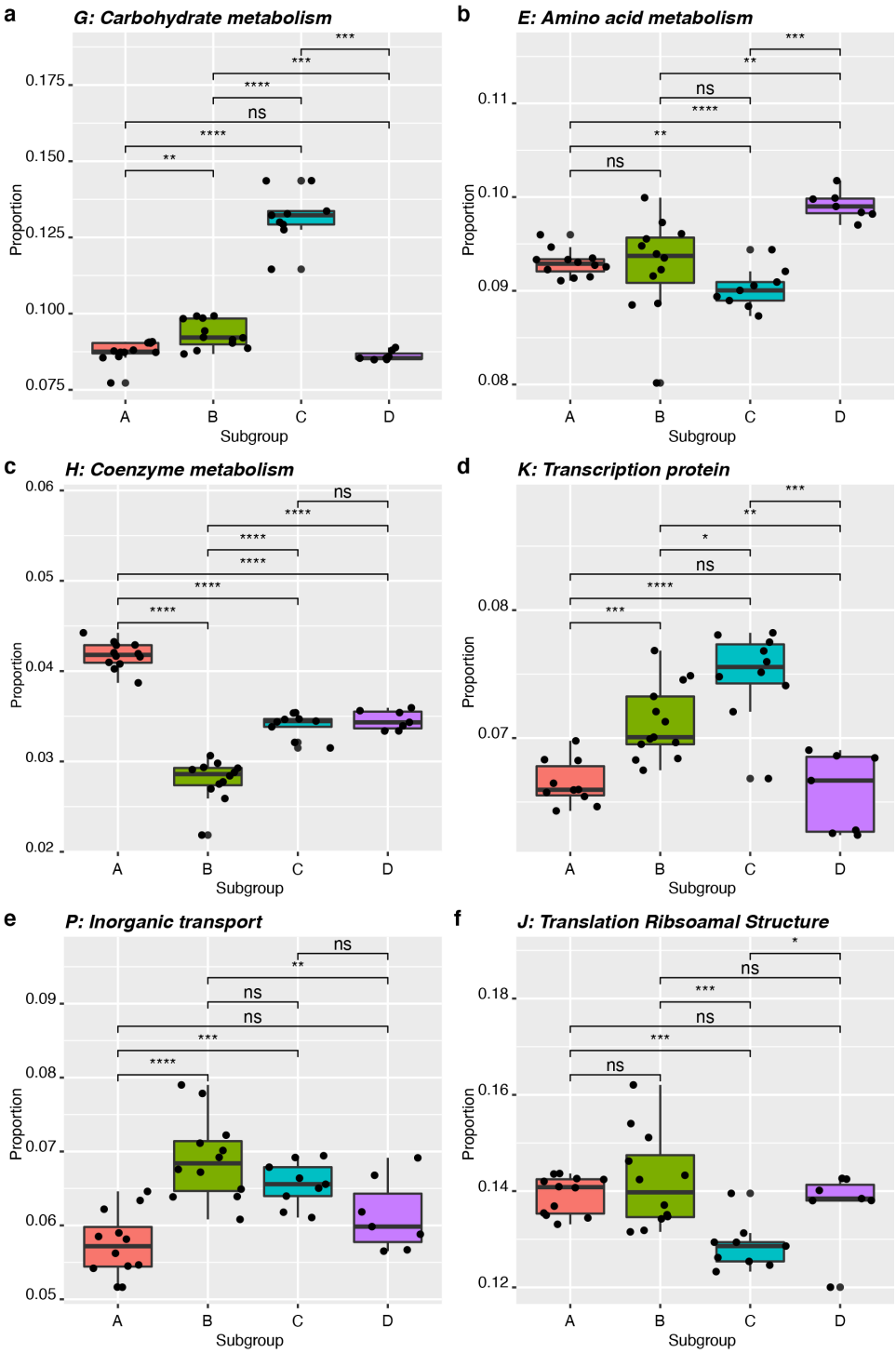

Fig. 3 Differential abundance of six COG categories that were identified by SIMPER analysis as main drivers of subgroup separation. **(a)** carbohydrate metabolism and transport proteins (Category G), **(b)** amino acid transport and metabolism proteins (Category E), **(c)** co-enzyme transport and metabolism proteins (Category H), **(d)** transcription proteins (Category K), **(e)** inorganic ion transport and metabolism proteins (Category P), and **(f)** translation, ribosomal structure and biogenesis proteins (Category J). Results of unpaired t-tests are indicated where * is $p \leq 0.05$, ** is $p \leq 0.01$, *** is $p \leq 0.001$, **** is $p \leq 0.0001$ and ns is not significant.

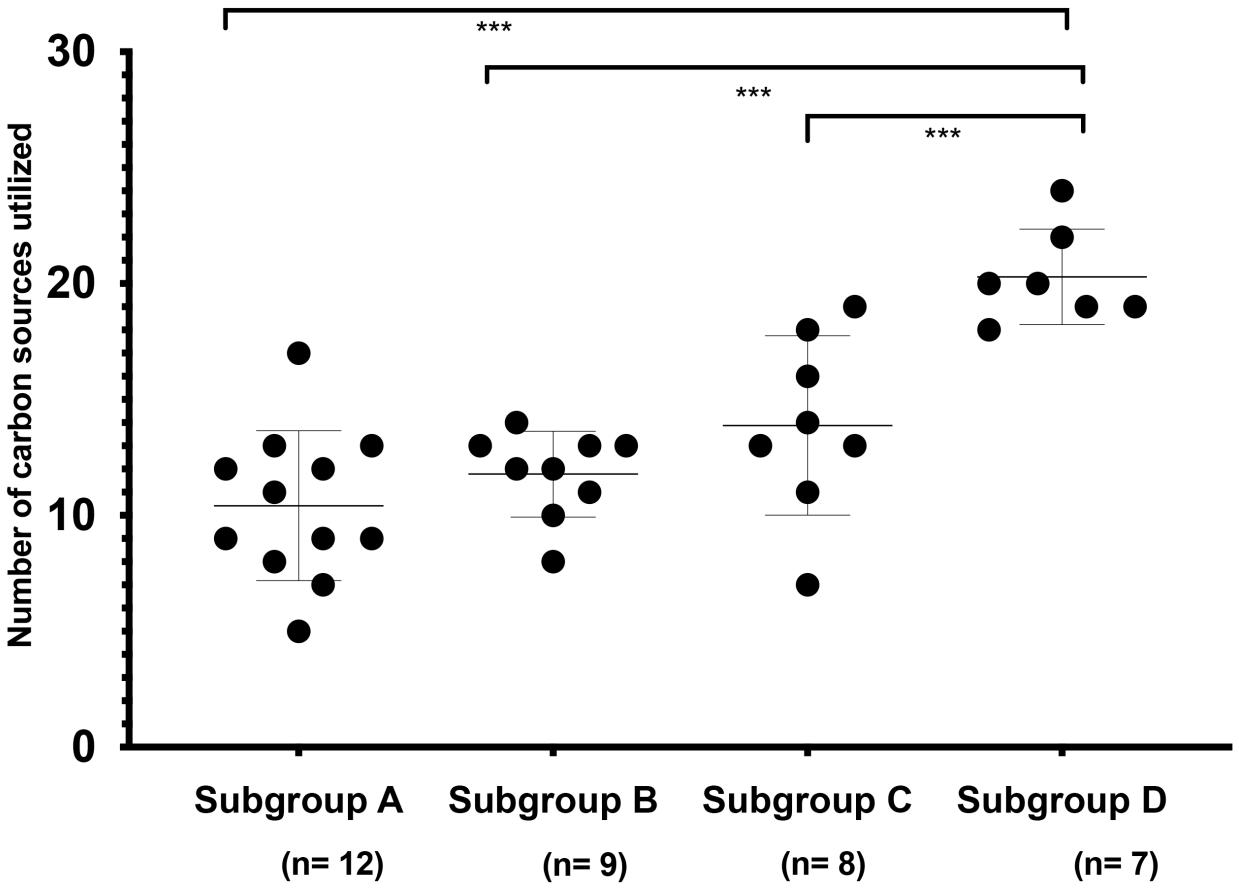

Fig. 4 Comparison of numbers of carbon sources utilized by isolates in each subgroup. The number of carbon sources utilized by isolates in subgroup D was significantly higher than those in subgroups A, B and C (TukeyHSD, $p \leq 0.001$). Only one fourth (9/36) of the tested isolates utilized more than 17 carbon sources, including all seven tested isolates of subgroup D.

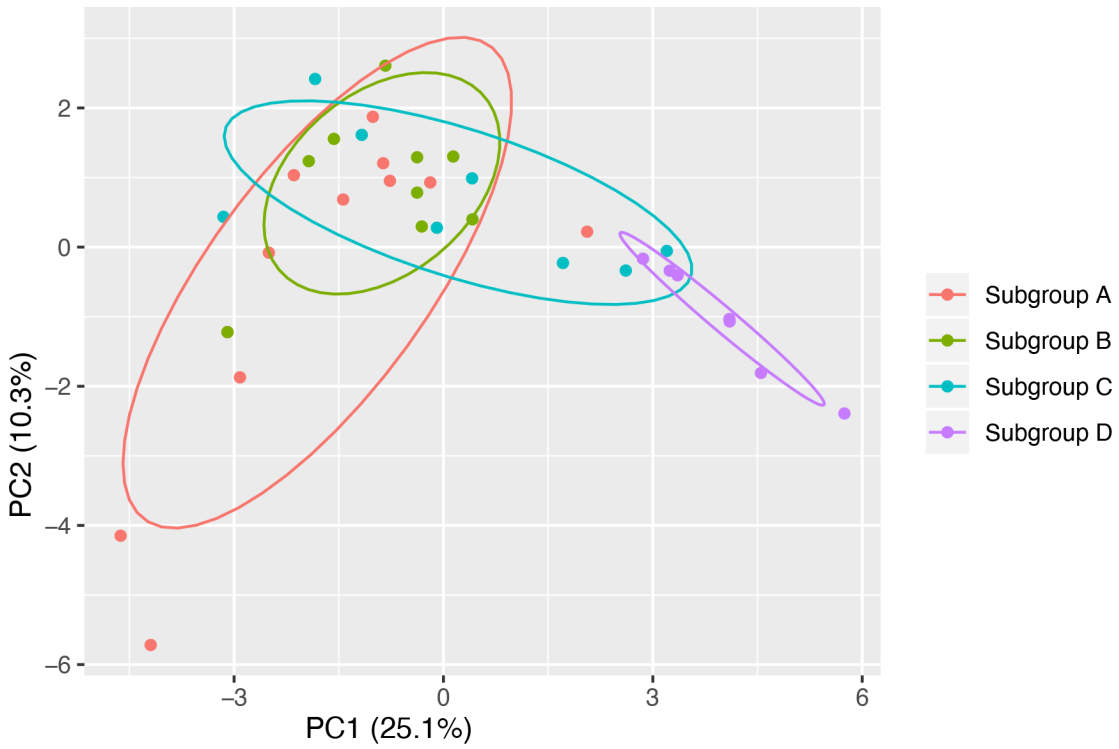

Fig. 5 Subgroup D has minimal overlap with the other subgroups in carbon source utilization. The degree of variation based on carbon source utilization between subgroup D and subgroups A and B was significant (pairwise ADONIS, $p < 0.05$, after Bonferroni adjustment). The variation in carbon sources utilization between subgroup C and D can be explained by subgroup affiliation in 39% of cases. Overall, 42% of differences in carbon sources utilization between subgroups can be explained by their subgroup affiliation (Adonis, $R^2 = 0.42$, $p < 0.05$).

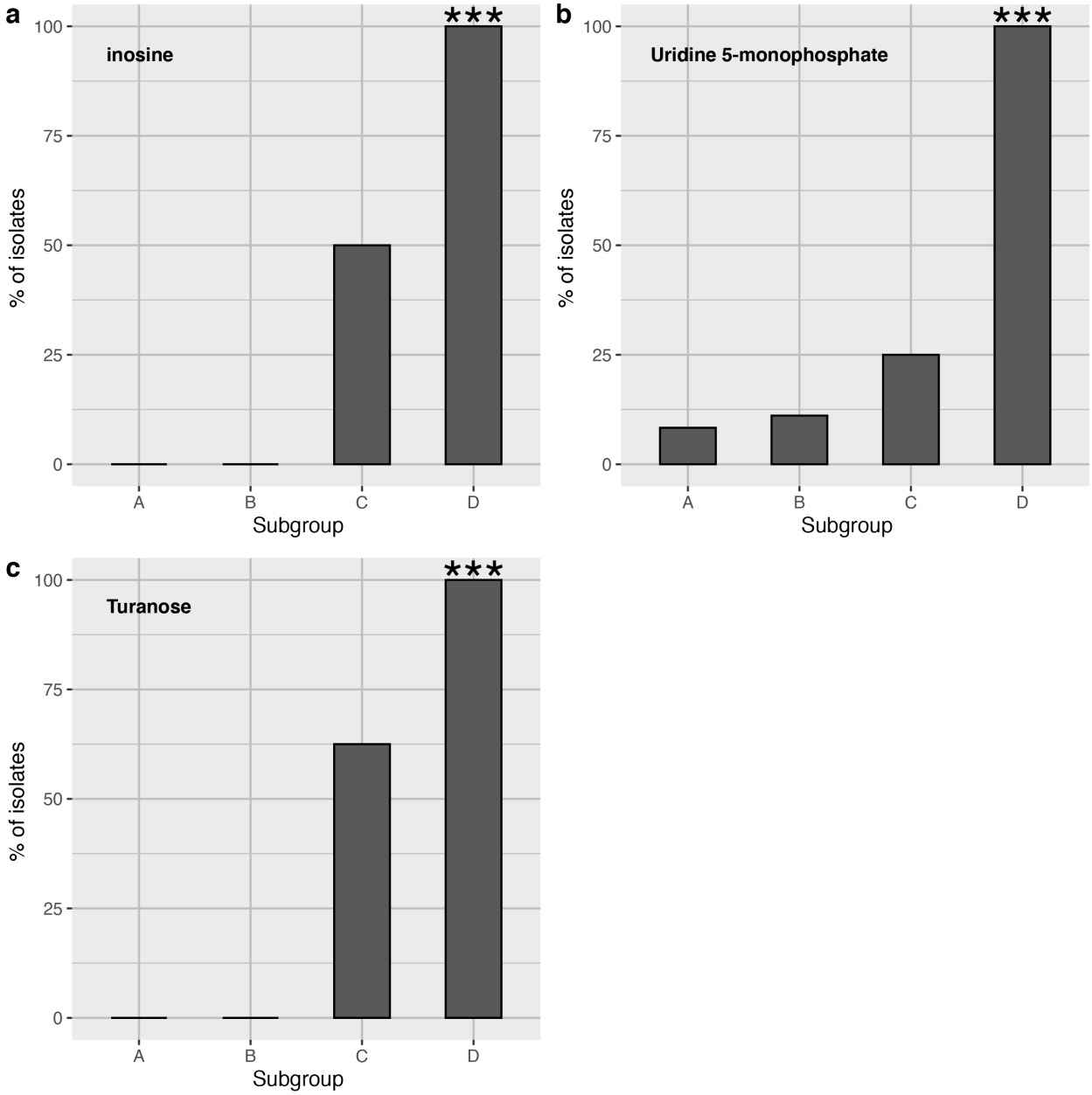

Fig. 6 Subgroup D can be differentiated from the other subgroups based on its capacity to utilize inosine **(a)**, uridine 5-monophosphate **(b)**, and turanose **(c)**. The percentage of isolates in each subgroup that utilize the indicated carbon source is shown. The utilization of these three carbon sources is significantly associated with subgroup D (Chi-square test, $p \leq 0.001$).

643 **Table 1**. Carbon source utilization by co-cultured isolates

Isolate 1 (Subgroup)	Isolate 2 (Subgroup)	No. carbon sources		
		Isolate 1	Isolate 2	Co-culture
VN003 (A)	VN002 (B)	11	13	12
VN003 (A)	NR001 (C)	11	19	13
VN003 (A)	WP012 (D)	11	24	22
VN002 (B)	NR001 (C)	13	19	15
VN002 (B)	WP012 (D)	13	24	20
NR001 (C)	WP012 (D)	19	24	19

644
645

646

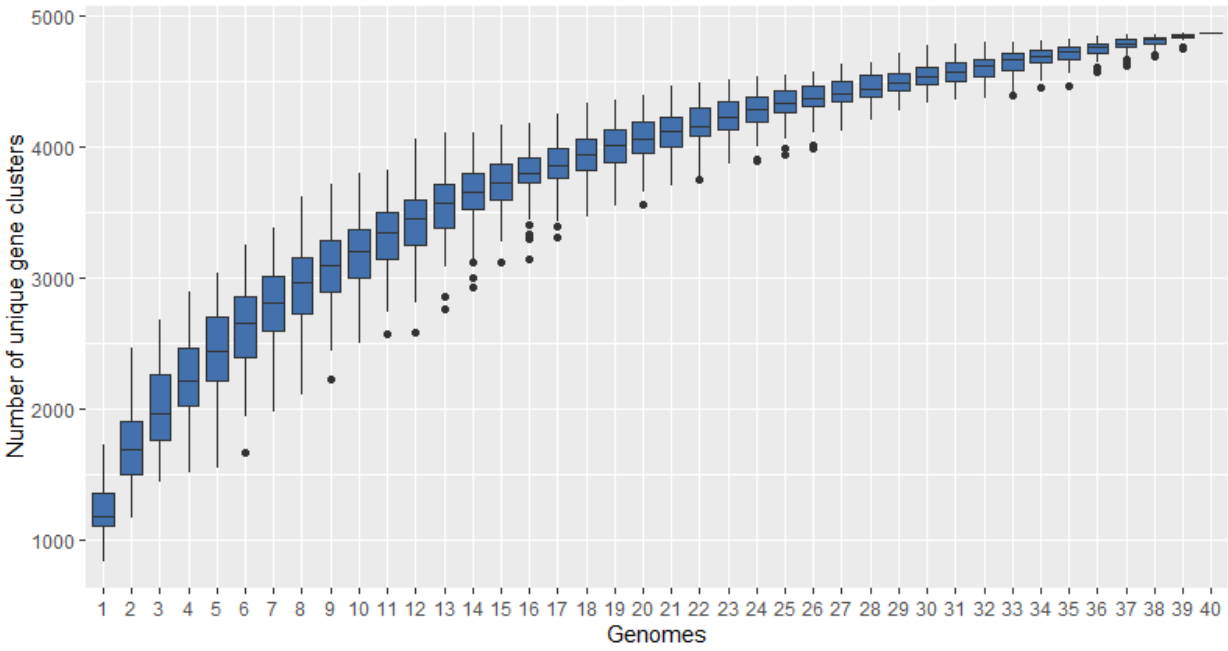

Fig. S1 Rarefaction curve of the pangenome of *Gardnerella* spp..

Table S1 Bacterial isolates and whole-genome sequences used for pangenome analysis and carbon source utilization profiling assay.

Isolate	cpn60 subgroup	Species	Included in pangenome and COG analyses?	Included in Biolog assay?	BioProject accession number
NR020	A	*G. swidsinskii*	Yes	Yes	PRJNA394757
NR015	A	Unknown	Yes	Yes	PRJNA394757
NR016	A	*G. swidsinskii*	Yes	Yes	PRJNA394757
GH005	A	*G. leopoldii*	Yes	Yes	PRJNA394757
NR020	A	*G. swidsinskii*	Yes	Yes	PRJNA394757
NR021	A	*G. swidsinskii*	Yes	Yes	PRJNA394757
NR017	A	*G. leopoldii*	Yes	Yes	PRJNA394757
NR019	A	*G. leopoldii*	Yes	Yes	PRJNA394757
NR010	A	*G. leopoldii*	Yes	Yes	PRJNA310104
N72	A	*G. swidsinskii*	Yes	Yes	PRJNA310104
VN003	A	*G. leopoldii*	Yes	Yes	PRJNA394757
WP021	A	*G. swidsinskii*	Yes	Yes	PRJNA394757
WP022	A	*G. swidsinskii*	Yes	Yes	PRJNA394757
N170	B	Genome species 3	Yes	Yes	PRJNA394757
N144	B	Genome species 3	Yes	Yes	PRJNA310104
N153	B	Genome species 3	Yes	No	PRJNA265102
W11	B	Genome species 3	Yes	Yes	PRJNA265103
NR026	B	Genome species 3	Yes	No	PRJNA394757
VN002	B	*G. piotii*	Yes	Yes	PRJNA394757
N101	B	Genome species 3	Yes	Yes	PRJNA265097
N95	B	Genome species 3	Yes	Yes	PRJNA265092
GH019	B	*G. piotii*	Yes	Yes	PRJNA394757
GH020	B	*G. piotii*	Yes	Yes	PRJNA394757
GH022	B	*G. piotii*	Yes	Yes	PRJNA394757
ATCC14018	C	*G. vaginalis*	No	Yes	PRJNA524873
ATCC14019	C	*G. vaginalis*	Yes	No	PRJNA55487
N165	C	*G. vaginalis*	Yes	Yes	PRJNA310104
NR037	C	*G. vaginalis*	Yes	No	PRJNA394757
NR038	C	*G. vaginalis*	Yes	Yes	PRJNA394757
NR039	C	*G. vaginalis*	Yes	Yes	PRJNA394757
GH015	C	*G. vaginalis*	Yes	Yes	PRJNA310104
GH021	C	*G. vaginalis*	Yes	Yes	PRJNA394757
NR001	C	*G. vaginalis*	Yes	Yes	PRJNA394757
WP023	C	*G. vaginalis*	Yes	Yes	PRJNA394757
NR003	D	Genome species 8	Yes	Yes	PRJNA394757
WP012	D	Genome species 9	Yes	Yes	PRJNA394757
N160	D	Genome species 10	Yes	Yes	PRJNA310104
NR047	D	Unknown	Yes	Yes	PRJNA394757
NR002	D	Genome species 8	Yes	Yes	PRJNA394757
NR043	D	Genome species 10	Yes	Yes	PRJNA394757
NR044	D	Genome species 10	Yes	Yes	PRJNA394757