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 genomic 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].

Interestingly, some species of *Gardnerella* co-occur together more frequently than others [4]. One of the factors contributing to the co-occurrence patterns of *Gardnerella* species observed in the vaginal microbiome could be ecological interactions between them. Recently, we

70 demonstrated that an indirect, exploitative competition between subgroups of *Gardnerella* is
71 prevalent in co-cultures *in vitro*. While the growth rates of isolates in subgroups A, B, and C,
72 were negatively affected by competition, growth rates of *Gardnerella* subgroup D isolates
73 increased with the increasing number of competing subgroups in co-culture communities [11].

74 The strength of microbial interactions between bacterial species can be affected by niche overlap
75 [12, 13], and species with similar nutritional requirements will naturally compete over the same
76 resources [14]. In addition to competition for nutritional resources, bacteria may also compete
77 for resources essential for colonizing a specific site. Since isolates from *Gardnerella* subgroups
78 A, B and C are negatively affected by competition, and subgroup D isolates experienced a boost
79 in growth rate, the degree of niche overlap between subgroup A, B, and C is presumably higher
80 than between subgroup D and any of the others.

81 Although the growth rate of subgroup D increases in co-cultures, it does not have an intrinsically
82 high growth rate. In fact, the *in vitro* growth rate of subgroup D is half of that of subgroup C,
83 which may contribute to its low abundance in the vaginal microbiome [11]. Low abundance
84 species are often favoured by negative-frequency-dependent selection [15, 16], which can be
85 governed by nutritional requirements [17]. Bacteria capable of utilizing relatively few,
86 abundantly available nutrients in a particular environment are nutritional specialists in the
87 context of that environment. Generalists, on the contrary, are bacteria capable of utilizing more
88 nutrient sources than their specialist counterparts. In negative frequency-dependent selection, the
89 resources accessible to rapidly growing specialists will dwindle, reducing the fitness of the
90 specialists as their population increases. As a result, the population of more generalist bacteria
91 capable of utilizing a wider range of nutrient sources will expand in a density-dependent manner

92 [17, 18]. Generalists can also negatively affect the growth of specialists by competing for the
93 resources that can be utilized by both of them [13].

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

101 **Methods**

102 **Bacterial isolates**

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

109 **Whole-genome sequencing**

110 Whole genome sequences for 10 of the study isolates had been published previously, and the
111 remaining 29 were sequenced as part of the current study (Table S1). DNA was extracted from
112 isolates using a modified salting out protocol [19] and was stored at -20°C. DNA was quantified
113 using Qubit dsDNA BR assay kit (Invitrogen, Burlington, Ontario) and the quality of the extracts

114 was assessed by the A260/A280 ratio. Isolate identity was confirmed by cpn60 barcode
115 sequencing as follows. cpn60 barcode sequences were amplified from extracted DNA with the
116 primers JH0665 (CGC CAG GGT TTT CCC AGT CAC GAC GAY GTT GCA GGY GAY
117 GGH CHA CAA C) and JH0667 (AGC GGA TAA CAA TTT CAC ACA GGA GGR CGA
118 TCR CCR AAK CCT GGA GCY TT). The reaction contained 2 μL template DNA in 1× PCR
119 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
120 of each primer, 2 U AccuStart Taq DNA polymerase, and water to bring to a final volume of 50
121 μL. PCR was carried out with incubation at 94°C for 30 seconds, 40 cycles of 94°C 30 sec, 60°C
122 for 1 min, 72°C for 1min, and final extension at 72°C for 10 min. PCR products were purified
123 and sequenced by Sanger sequencing and compared with the chaperonin sequence database
124 cpnDB [20] to confirm identity.

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

130 Raw sequences were trimmed using Trimmomatic [21] with a minimum quality score of 20 over
131 a sliding window of 4, and minimum read length of 40. Trimmed sequences were assembled
132 using SOAPdenovo2 [22] or SPAdes (NR002, NR043, NR044) [23]. Assembled genomes were
133 annotated using the National Center for Biotechnology Information Prokaryotic Genome
134 Annotation Pipeline.

135 **Pangenome analysis**

136 Pangenome analysis of the 39 study isolates and the published genome of *G. vaginalis* strain
137 ATCC 14019 (Accession number: PRJNA55487) was performed using the micropan R package
138 [24]. We used "complete" linkage for clustering, and the cut-off value for the generation of
139 clusters was set to 0.75. Jaccard Index was used to determine the relationships of patterns of
140 presence and absence of proteins among all isolates.

141 **COG analysis**

142 Predicted protein sequences from individual genomes were classified into Clusters of
143 Orthologous Groups (COG) categories using WebMEGA
144 (http://weizhonglab.ucsd.edu/webMEGA).

145 **Carbon source utilization assay**

146 Bacterial isolates from freezer stocks were streaked on 5% sheep blood agar plates and were
147 grown for 48 h anaerobically, prior to inoculation of AN Microplates (Biolog Inc, Hayward,
148 CA). Each plate contained 95 carbon sources and one blank well. Colonies of *Gardnerella*
149 isolates were harvested using a sterile swab and suspended in 14 mL of inoculating fluid
150 supplied by the manufacturer. The cell density was adjusted to 55%T (OD_{595} approximately
151 0.25) using a turbidimeter. Each well was filled with 100 µl of culture suspension and was
152 incubated at 35˚ C anaerobically for 48 h. All inoculations and incubations were performed in an
153 anaerobic chamber containing 10 % CO_2, 5% hydrogen, and 85% nitrogen. All plates were read
154 visually after 48h of incubation. If there was no carbon source utilization, the wells remained
155 colourless. A visual change from colourless to purple indicated carbon source utilization. To
156 avoid bias in interpretation, a subset of the plates was read by a second observer who was
157 blinded to the identity of the isolates. There was no disagreement between independent
158 observers. The entire experiment was performed in two biological replicates.

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

160 Representative isolates from the four subgroups were co-cultured in the Biolog AN Microplate in
161 a pairwise fashion (n =6, AB, AC, AD, BC, BD, CD), by combining 50 µL of each isolate
162 suspended in inoculation fluid in each well. The co-cultured AN Microplates were incubated at
163 35˚C for 48h before being assessed visually for colour change. The experiment was repeated on
164 separate days.

165 **Statistical analysis**

166 Statistical analyses included one-way ANOVA, student's t-test and chi-square tests to determine
167 if utilization of particular carbon sources was associated with specific subgroups. ADONIS and
168 pairwise-ADONIS were performed to assess the degree of similarity between the isolates in
169 terms of presence/absence of clusters generated in the pangenome analysis, proportional
170 abundance of proteins in various COG categories, and carbon source utilization patterns.
171 SIMPER analysis was performed to identify variables that drive the differences between the
172 subgroups in ADONIS. PCA was used to cluster the isolates based on the presence/absence of
173 clusters generated by pangenome analysis, proportional distribution of proteins among COG
174 categories, and positive/negative results from carbon source profiling assay in Biolog plates. All
175 figures were generated using GraphPad Prism 8.0 and RStudio (version 3.5.2). All statistical
176 analyses were performed in RStudio (version 3.5.2).

177

178 **Results**

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

180 The purpose of our pangenome analysis was to estimate the degree of niche overlap between
181 *Gardnerella* subgroups based on comparisons of their predicted proteomes. Hierarchical

182 clustering using complete linkage produced 4,868 clusters or predicted proteins in the
183 pangenome of the 40 isolates included (Fig. S1). The strict core (defined as the protein clusters
184 present in all isolates) included 176 clusters. Most of these core proteins were related to
185 metabolism, transcriptional control, DNA replication, and protein synthesis. Based on the
186 presence/absence pattern of clusters among isolates, a PCA was performed to determine the
187 extent of overlap between the subgroups. The amount of variance explained by the two principal
188 components was 19.4%, based on which, the four subgroups were separable (Fig. 1). The
189 dissimilarity between the four subgroups was significant (pairwise-ADONIS, Bonferroni
190 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, $R^2 = 0.34$; B
191 vs D, $R^2 = 0.45$; and C vs D, $R^2 = 0.55$).

192 Following the identification of core and accessory proteins, we investigated the distribution of
193 functional classifications of proteins encoded by isolates in the four subgroups. COG analysis
194 resulted in assignment of predicted proteins into 23 functional categories. Based on the output
195 from this process, the proportion of proteins in each of the COG categories was calculated for
196 each genome. The distributions of proportional abundances of each category were then used to
197 assess the relationships of the four subgroups in terms of COG category representation (Fig. 2).
198 The differences between all subgroups were significant (pairwise ADONIS, Bonferroni adjusted,
199 $p <0.05$, 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, $R^2 =$
200 0.20; and C vs D, $R^2 = 0.74$).

201 We identified the variables which caused the four subgroups to diverge in terms of abundance of
202 different COG categories in a multivariate analysis using SIMPER [25]. SIMPER calculates the
203 contribution of each variable to the dissimilarity observed between two groups and relies on
204 Bray-Curtis dissimilarity matrix for calculating the proportion of contribution of each variable

205 being tested. Thirty-six percent of the differences between subgroup A and subgroup B were
206 accounted for by amino acid transport and metabolism (COG category E), inorganic ion transport
207 metabolism (category P), translation, ribosomal structure and biogenesis proteins (category J).
208 The proportion of proteins with functions related to carbohydrate transport and metabolism
209 (category G) was the major factor that differentiated subgroup A from C, contributing to 34% of
210 the dissimilarity observed. Carbohydrate transport and metabolism also accounted for 31% of the
211 dissimilarity observed between subgroups B and C and 36% of the dissimilarity between
212 subgroups C and D. The major contributing factors that differentiated subgroup A and D were
213 proportional abundance of proteins assigned to functional categories H (co-enzyme transport and
214 metabolism) and E (amino acids transport and metabolism) (23%). Subgroups B from D were
215 separated primarily based on functional categories P (inorganic ion transport and metabolism), J
216 (translation, ribosomal structure and biogenesis), G (carbohydrate transport and metabolism
217 proteins), and E (amino acid transport and metabolism proteins), which together accounted for
218 37% of the dissimilarity observed.

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

220 We tested if the proportions of individual functional categories of proteins that drive the overall
221 separation of the four subgroups in multivariate analysis were significantly different between
222 pairs of *Gardnerella* subgroups. This analysis revealed that subgroup C has a significantly higher
223 proportion of its encoded proteins associated with carbohydrate transport and metabolism and
224 transport, and transcriptional regulation than the other subgroups (paired-t-test, $p < 0.05$,
225 Bonferroni adjusted, Fig. 3a, 3d). The proportion of proteins associated with amino acid
226 transport and metabolism is significantly higher in subgroup D than subgroups A, B, and C
227 (paired t-test, $p < 0.05$, Bonferroni adjusted, Fig. 3b). Proteins involved in co-enzyme transport

228 and metabolism were found in significantly higher proportional abundance in subgroup A than in
229 subgroup B, C and D (paired t-test, $p < 0.05$, Bonferroni adjusted, Fig. 3c). Subgroup B has a
230 significantly higher abundance of proteins associated with inorganic ion transport and
231 metabolism than subgroup A and D (paired t-test, $p < 0.05$, Fig. 3e), but the difference between
232 subgroup B and C was not significant. Subgroup B also has a significantly higher proportion of
233 translation, ribosomal structure and biogenesis proteins (paired t-test, $p<0.05$, Fig. 3f) compared
234 to subgroup C.

235 **Carbon source utilization phenotypes**

236 We hypothesized that subgroup D, a slow-growing, rarely detected *Gardnerella* subgroup is
237 maintained in the vaginal microbiome at a low level and avoids competitive exclusion through
238 negative-frequency-dependent selection, made possible by being a nutritional generalist. We
239 performed carbon source utilization profiling of thirty-six representative isolates (n= 12,
240 subgroup A; n= 9, subgroup B; n=8, subgroup C (including type strain *G. vaginalis* ATCC
241 14018); and n=7, subgroup D). The number of carbon sources utilized by any *Gardnerella* strain
242 ranged from 5 to 24. Only 25% (9/36) of the isolates utilized more than 17 carbon sources,
243 including two subgroup C (NR001, NR038) and all subgroup D isolates. Twenty isolates utilized
244 at least 13 carbon sources, including three subgroup A (3/12, 25%), four subgroup B (4/8, 50%),
245 six subgroup C (6/7, 86%), and all seven isolates of subgroup D (100%). The average number of
246 carbon sources utilized by isolates in subgroups A, B, C, and D was 10.4±3.1, 11.8±1.75,
247 13.9±3.6, and 20.3±1.9, respectively (Fig. 4). A one-way ANOVA was performed to compare
248 the overall difference in carbon sources utilization among the four subgroups showed significant
249 difference among the subgroups ($F(3,32) = 18.15$, $p < 0.05$). A posthoc comparison between the
250 subgroups revealed that the number of carbon sources utilized by subgroup D was significantly

251 higher than subgroup A, B, and C (Tukey HSD, $p < 0.05$). All of the tested *Gardnerella* isolates
252 were able to utilize pyruvic acid, palatinose, and L-rhamnose. The next most frequently utilized
253 carbon sources were D-fructose (32/36, 97%) and L-fucose (32/36, 97%).

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

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

261 To determine if subgroups could be distinguished based on carbon source utilization profiles, a
262 principal component analysis was performed (Fig. 5). The overlap between the representative
263 isolates of subgroups A, B, and C was significant. Subgroup D was significantly dissimilar to
264 subgroups A and B (Fig. 5, pairwise-ADONIS, A vs D, $R^2 = 0.55$; B vs D, $R^2 = 0.55$, $p<0.05$).
265 Although the dissimilarity between subgroup C and D was not statistically significant after
266 Bonferroni adjustment, 39% (pairwise ADONIS, C vs D, $R^2= 0.39$) of the variation in carbon
267 source utilization could be explained by subgroup affiliation of the tested isolates, which was
268 higher than between subgroups A, B and C (A vs B: 13%, A vs. C: 21%, and B vs. C: 8%).

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

270 To identify carbon sources that differentiate the subgroups, we selected twelve substrates that
271 were utilized by more than five isolates but fewer than thirty isolates. Chi-square tests were
272 performed to determine if the subgroups significantly differ in the utilization of those twelve

273 carbon sources. The four Gardnerella subgroups differed in their use of 3 of the 12 carbon
274 sources: turanose, inosine, and uridine 5-monophosphate (Chi-square test, $p < 0.05$, Bonferroni
275 adjusted) (Fig. 6). For each of these three carbon sources, subgroups A and B had low frequency
276 of use (9.5% = 2/21, 0.0% = 0/21, 0.0% = 0/21; subgroups A and B combined), subgroup C had
277 low or intermediate frequency of use (25.0% = 2/8, 50.0% = 4/8, and 62.5% = 5/8), whereas
278 subgroup D had high frequency of use (100.0% = 7/7, 100.0% = 7/7, 100.0% = 7/7).

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

280 Since the four subgroups co-exist in the same ecosystem, it is possible that mixing them might
281 facilitate the utilization of certain carbon sources. To detect any such facilitation in carbon
282 sources utilization, we co-cultured isolates from all four subgroups in six pairwise combinations
283 (A+B, A+C, A+D, B+C, B+D, and C+D). No combination resulted in the utilization of more
284 carbon sources than the best utilizer in any combination.

285

286 **Discussion**

287 Rarely abundant species can be maintained in the human microbiome through a variety of
288 mechanisms, which include but are not limited to sequestration of essential nutrients from
289 competing species, diversification of phenotype [26], social cheating [27], and negative
290 frequency dependent selection [16]. Differences in nutrient utilization among community
291 members can be a key factor that sets the stage for negative frequency dependent selection [28].
292 The reproductive fitness of nutritional specialist species will remain high as long as the supply of
293 nutrients usable by the specialists is abundant. As soon as the supply of these nutrients drops,
294 slower-growing generalists, by virtue of their greater utilization capacity, will have increased
295 fitness, which if unchecked, will eventually lead to their dominance.

296 Among the four subgroups of *Gardnerella* spp. that colonize the vaginal microbiome of
297 reproductive-aged women, subgroup D is the rarest [4]. Subgroup D is also relatively slow-
298 growing, yet shows an increased growth rate when the number of competitors in an *in vitro*
299 community increases [11]. We have reported previously that resource-based competition is
300 common among *Gardnerella* spp. and no evidence for contact-dependent interaction was
301 observed. Therefore, we set out to investigate if negative-frequency dependent selection is
302 responsible for persistence of subgroup D, which would require relatively small niche overlap
303 and a more generalist lifestyle than the other *Gardnerella* spp. in the vaginal microbiome.

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

305 Niche overlap may lead to competition for nutrients and space [28–31] and it has been reported
306 that competition is prevalent among metabolically similar bacterial species [13]. Occupying
307 distinct niches can therefore help bacterial species avoid competition for space, growth factors,
308 and nutrients, resulting in increased reproductive fitness. Since subgroup D isolates have higher
309 growth rates in the presence of competitors compared to when grown alone, these isolates
310 presumably occupy a distinct niche. The pangenome analysis showed that the four subgroups
311 differ significantly based on the composition of their predicted proteomes (Fig. 1), with only 176
312 proteins comprising the strict core of proteins found in all isolates. This finding is not surprising
313 since the genetic diversity among *Gardnerella* is well established, and genome sequence
314 comparisons formed the basis for the recent reclassification of *Gardnerella* into 13 genome
315 species [1–3].

316 Comparisons of the entire proteomes do not, however, focus on the key factors for a resource-
317 based competition: nutrient utilization potential. Proteins involved in nutrient uptake and
318 metabolism account for only a fraction of the 4,868 protein clusters comprising the pangenome.

319 Analysis of the distribution of various functional (COG) categories of proteins revealed
320 significant differences among subgroups in their predicted capacity to utilize carbohydrates and
321 amino acids (Fig. 3a, 3b), with subgroup D having significantly more of its proteome dedicated
322 to amino acid transport and metabolism than any of the other subgroups. Since a resource-based
323 competition encapsulates competition for space, growth factors and nutrients, our findings from
324 the pangenome and COG analyses suggest that the competition among the four subgroups is not
325 spatial but may be primarily for nutrients; a speculation supported by the previous observation
326 that *Gardnerella* spp. form multi-subgroup biofilms [11].

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

328 The diversity of nutrients available to microbiota in the vaginal microbiome is less than in the
329 gastrointestinal microbiome, where food intake provides a constant source of diverse nutrients
330 that affect the assembly of gut microbiota [32, 33]. Vaginal microbiota, on the contrary, are
331 largely dependent upon host-derived nutrients, the most abundant of which is glycogen.
332 Glycogen is deposited in the vaginal lumen by epithelial cells under the influence of estrogen
333 [34], and is digested into maltooligosaccharides, maltodextrins and glucose by the combined
334 activities of host and microbial enzymes prior to uptake and metabolism by the microbiota [35–
335 37]. Given the relatively narrow range of nutrients available in the vaginal microbiome, it is
336 expected that the resident microbiota, including the four subgroups of *Gardnerella,* overlap to a
337 considerable extent in their nutrient utilization capacity, resulting in some level of competition
338 among them [29, 33, 38]. As discussed earlier, subgroup D is an exception since the growth of
339 these isolates was actually facilitated in co-cultures, suggesting that while it may compete with
340 other *Gardnerella* spp. over common nutrients like the breakdown products of glycogen, it may
341 be able to utilize a greater overall diversity of nutrients (i.e. it is a generalist).

342 The AN microplate assay results showed that subgroup D isolates utilized more of the provided
343 carbon sources than isolates in the three other subgroups (Fig. 4). Furthermore, when the patterns
344 of substrate use were considered, subgroup A, B and C were not separable from each other, but
345 subgroup D was significantly different (Fig. 5). The distinct pattern observed in subgroup D was
346 partially driven by utilization of three particular substrates: turanose, inosine, and uridine 5-
347 monophosphate (Fig. 6). Turanose is an isomer of sucrose, known as a non-accumulative
348 osmoprotectant, aiding bacterial growth at high osmolarity [39]. The importance of turanose
349 utilization in the vaginal environment is not known yet, but our observation is an indication that
350 subgroup D isolates can metabolize sucrose-like sugars. The two other carbon sources: inosine
351 and uridine 5-monophosphate are probably used in purine and pyrimidine biosynthesis in
352 *Gardnerella* spp..

353 Some findings of the pangenome and COG analyses could not be reconciled with the phenotypic
354 carbon source utilization assay. For example, although subgroup C and D have higher
355 proportions of their proteomes predicted to be involved in transport and metabolism of
356 carbohydrates (Category G) and amino acids (Category E), respectively, than the other
357 subgroups, subgroup C isolates did not utilize the greatest number of available sugar substrates
358 in the AN microplate and subgroup D isolates did not utilize any of the amino acid substrates
359 available. It is, however, important to consider that the carbon source utilization assay was
360 performed in a plastic environment and included only 95 substrates, many of which are not
361 relevant to the vaginal microbiome. More relevant amino acid sources available in the vagina,
362 including those whose abundance is altered in bacterial vaginosis, such as isoleucine, leucine,
363 proline, and tryptophan, are not included [40–42]. Ideally, this study would have involved a
364 vaginal-microbiome specific nutrient panel, but such reagents were not available. Even with this

365 limitation, our results support the conclusion that subgroup D is a nutritional generalist relative
366 to other *Gardnerella* spp.. Most of the ecological studies that have been performed to date to
367 elucidate mechanisms shaping the assembly of bacterial communities have included either
368 environmental bacterial species or well-characterized model organisms [13, 15, 26, 33, 38, 43–
369 47]. There are understandably fewer studies that focus on interactions among host-associated
370 microbiota [48, 49].

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

372 The genomic and phenotypic differences we observed between subgroup D and the three other
373 subgroups, including the potential to utilize more amino acids, use of a greater number of carbon
374 sources, and a distinct pattern of substrate utilization, suggest that subgroup D is a candidate for
375 negative frequency dependent selection. Why then are these *Gardnerella* spp. only observed
376 rarely, and in low abundance in reproductive aged women? Among 413 vaginal samples from
377 reproductive aged Canadian women, genome species comprising subgroup D of *Gardnerella*
378 were detected in <10% of samples and never accounted for more than 5% of the microbiota [4].
379 Vaginal environmental dynamics and related host factors, such as menstruation, sloughing of
380 epithelial cells, and fluctuating pH contribute to the turnover of bacterial species, shifting the
381 bacterial population density and changing the nutrients available [50]. A decline in population
382 density would reshuffle the vaginal ecosystem, increasing the supply of abundant nutrients
383 accessible to faster growing, specialists, and checking the growth of slower growing generalist
384 subgroup D.

385 Although subgroup D is likely rare due the factors described above, it could still be a major
386 player in ecological succession and transition of vaginal microbiota between a *Lactobacillus*
387 dominated community and the overgrowth of anaerobes characteristic of bacterial vaginosis.

388 These organisms may also play a particular role in biofilm formation or competition for
389 occupancy of the vaginal mucosa. Rarely abundant species often act as keystone species helping
390 colonization by other bacterial species, which are also essential to maintain homeostasis of an
391 ecosystem [51–54]. Resolution of the role of low abundance *Gardnerella* spp. will depend on the
392 development and application of experimental systems that more closely model the human
393 vaginal microbiome. Rodent models have shown some promise, especially for studies of specific
394 combinations of organisms [55], but there is also potential in bioreactors [56], and cell and tissue
395 culture systems that attempt to recapitulate many of the environmental and physiological aspects
396 of the vaginal microbiome [57]. Further study of rare *Gardnerella* spp. will likely also result in
397 the definition of additional species within this diverse genus.

398 **Acknowledgements**

399 The authors are grateful to Champika Fernando for excellent technical support and to all
400 members of the Hill Lab for their review of the manuscript. Special thanks to Prairie Diagnostic
401 Services for access to their anaerobic chamber. No thanks to COVID-19. The research was
402 supported by an Natural Sciences and Engineering Research Council of Canada Discovery Grant
403 to JEH.

404 **References**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

459 18. Ross-Gillespie A, Gardner A, Buckling A, et al (2009) Density dependence and cooperation:
460 theory and a test with bacteria. Evolution 63:2315–2325. https://doi.org/10.1111/j.1558-
461 5646.2009.00723.x

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

525 43. Stubbendieck RM, Straight PD (2015) Escape from lethal bacterial competition through
526 coupled activation of antibiotic resistance and a mobilized subpopulation. PLOS Genet
527 11:e1005722. https://doi.org/10.1371/journal.pgen.1005722

528 44. Gonzalez DJ, Haste NM, Hollands A, et al (2011) Microbial competition between *Bacillus*
529 *subtilis* and *Staphylococcus aureus* monitored by imaging mass spectrometry. Microbiology
530 157:2485–2492. https://doi.org/10.1099/mic.0.048736-0

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

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

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

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

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

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

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

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

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

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

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

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

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

568

569

570

571

572

573

574

575

576

577

578

579

580

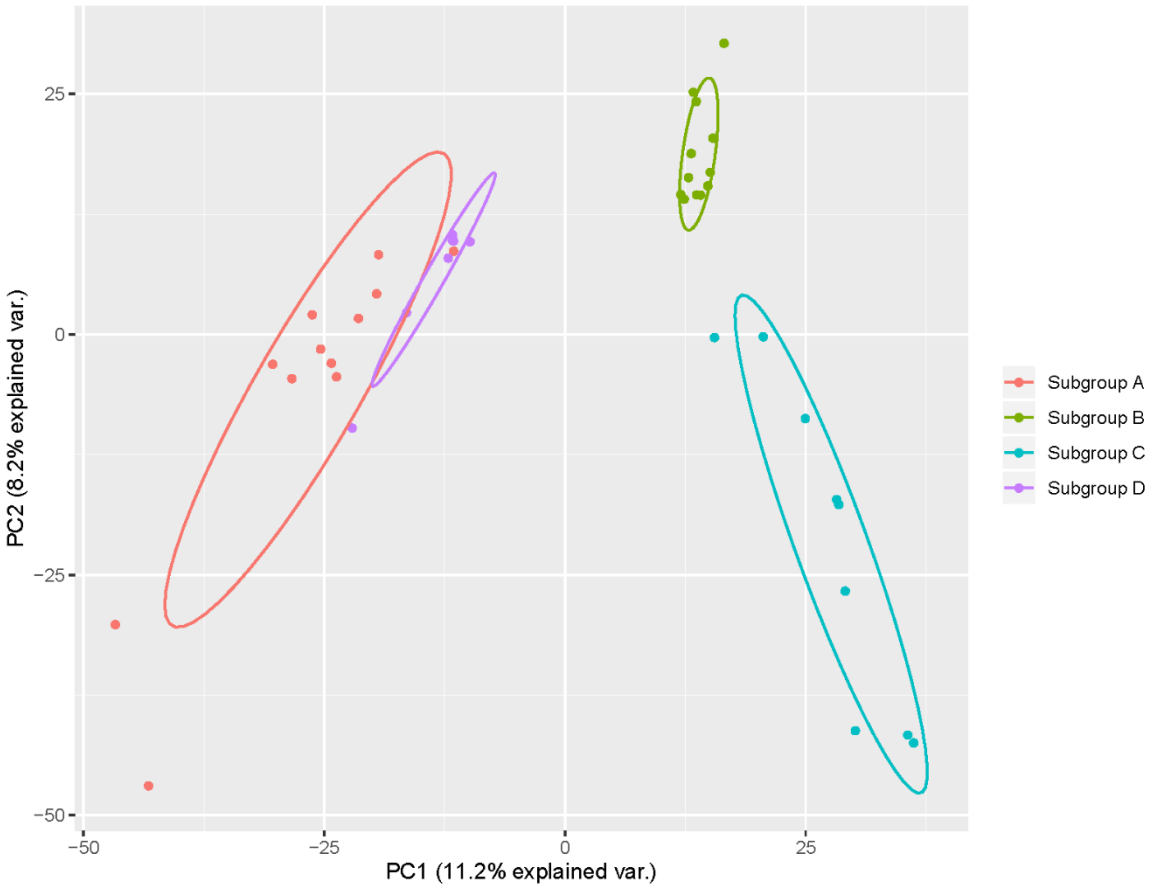
581

582

583

584

585



586

587 **Fig. 1** Four cpn60 defined subgroups of *Gardnerella* spp. are separable based on predicted

588 protein sequences. A presence/absence matrix was created for all protein clusters identified in the

589 pangenome analysis. The matrix was then used to perform PCA to determine the overlap

590 between the four subgroups. A pairwise ADONIS performed between the four groups showed

591 that the dissimilarity between the four subgroups is significant ($p < .05$, Bonferroni adjusted). The

592 dissimilarity between subgroups was 45%, 48%, 26%, 34%, 45%, and 55% for A vs B, A vs C,

593 A vs D, B vs C, B vs D, and C vs D, respectively.

594

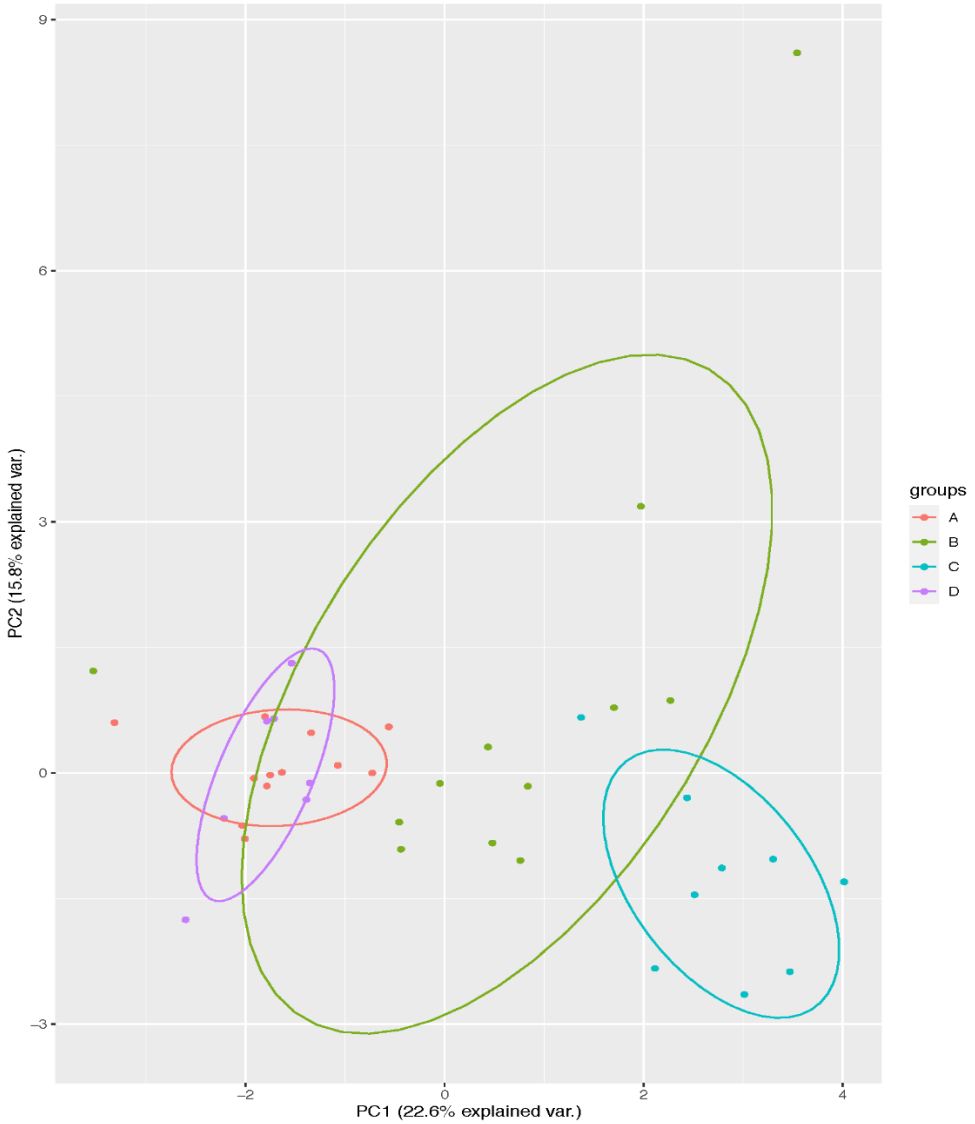


Fig. 2 Four cpn60 defined subgroups of *Gardnerella* spp. are separable based on the proportion of different categories of orthologous protein. The distribution of proportional abundance of twenty-three COG categories was compared among the four subgroups using PCA. The separability between the four subgroups based on the distributions of COG categories was significant (pairwise ADONIS, $p < .05$, Bonferroni adjusted). The dissimilarities between subgroups were 31% (A vs. B), 71% (A vs. C), 26% (A vs. D), 48% (B vs. C), 20% (B vs. D), and 74% (C vs. D).

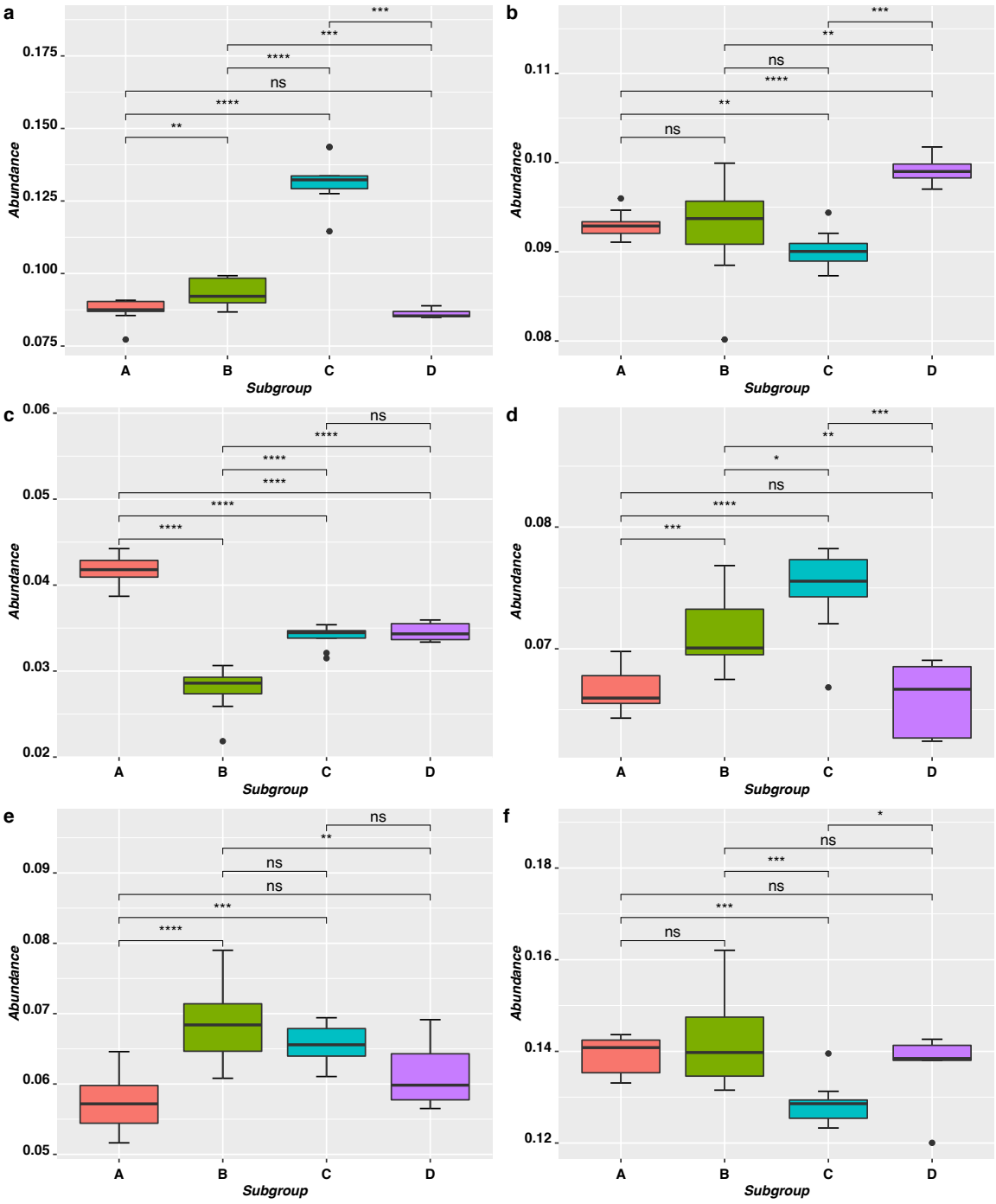
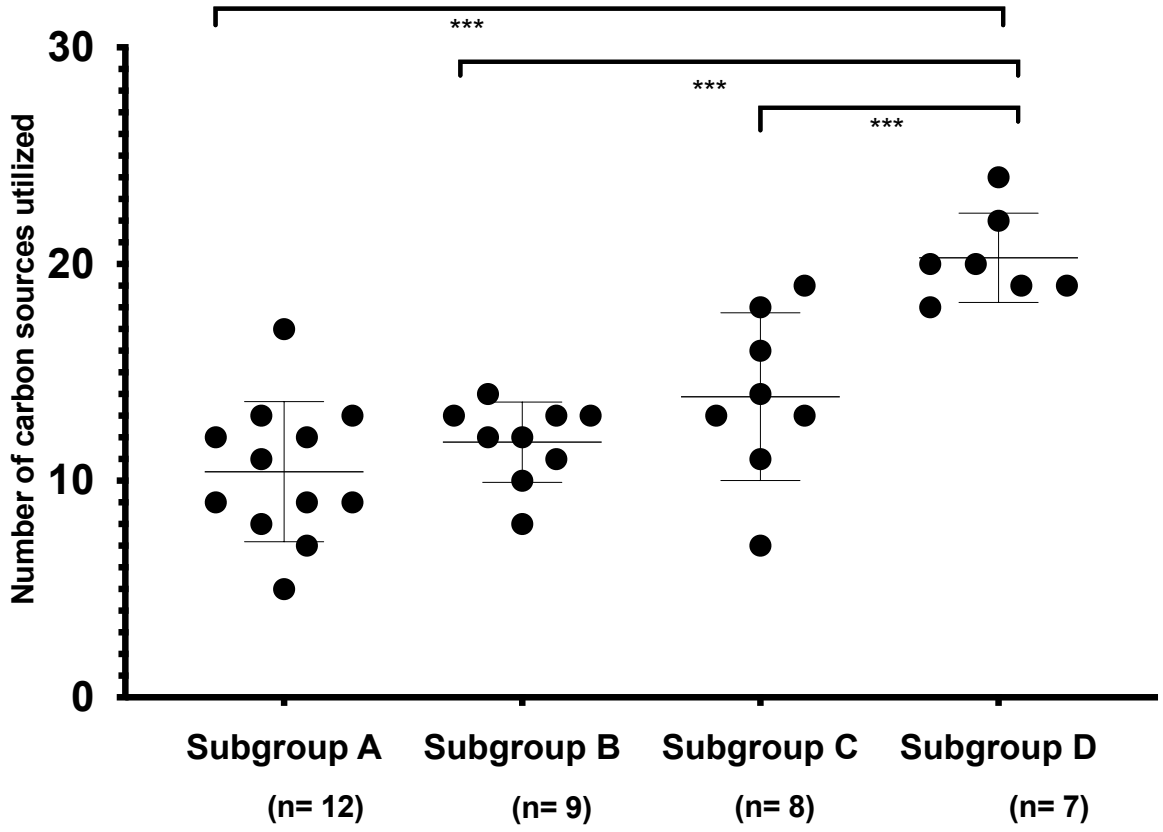


Fig. 3 The separation of the four subgroups are driven mainly by differential abundance of six COG categories: (a) carbohydrate metabolism and transport proteins (Category G), (b) amino acid transport and metabolism proteins (Category E), (c) co-enzyme transport and metabolism proteins (Category H), (d) transcription proteins (Category K), (e) inorganic ion transport and

609 metabolism proteins (Category P), and (f) translation, ribosomal structure and biogenesis
610 proteins (Category J). Subgroup C has significantly higher proportion of its proteome classified
611 in Category G (carbohydrate metabolism proteins) than the other subgroups, whereas D and A
612 had significantly higher proportions of their proteomes classified into Category E (amino acid
613 and transport metabolism) and co-enzyme transport and metabolism (Category H), respectively.
614

615



616

617

618 **Fig. 4** Subgroup D isolates utilize more carbon sources than the three other subgroups. The

619 number of carbon sources utilized by isolates in subgroup D was significantly higher than those

620 in subgroups A, B and C (TukeyHSD, $p < 0.05$). Only one fourth (9/36) of the tested isolates

621 utilized more than 17 carbon sources, including all seven tested isolates of subgroup D.

622

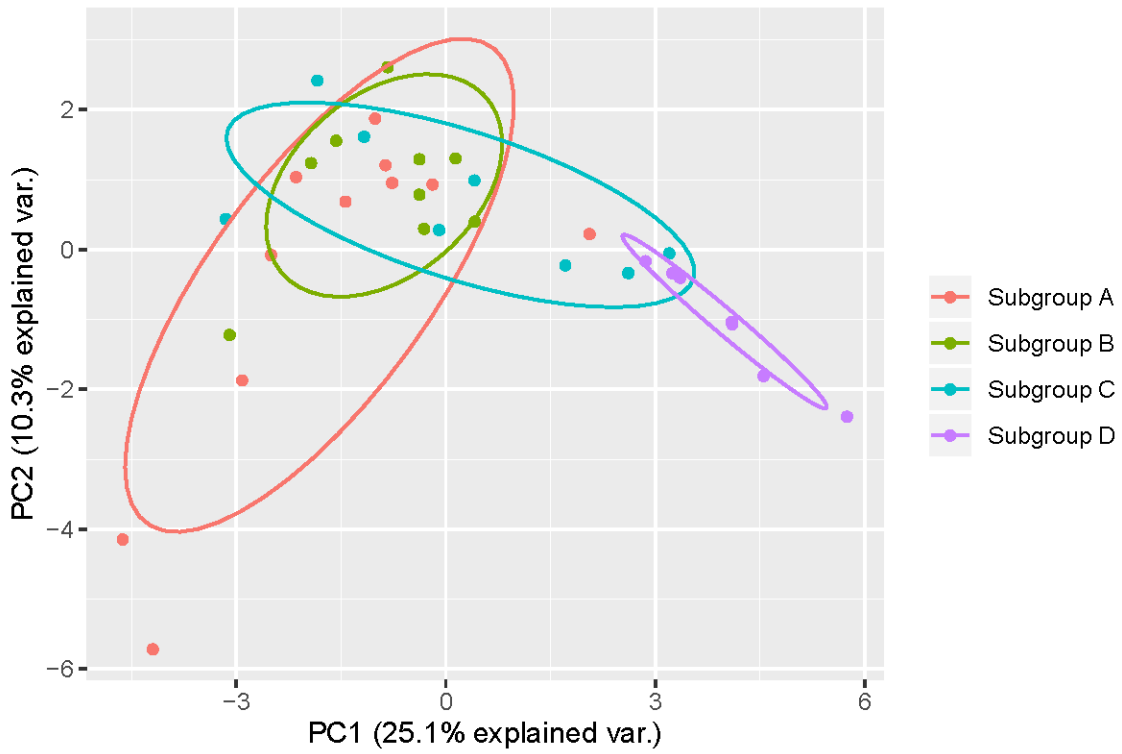
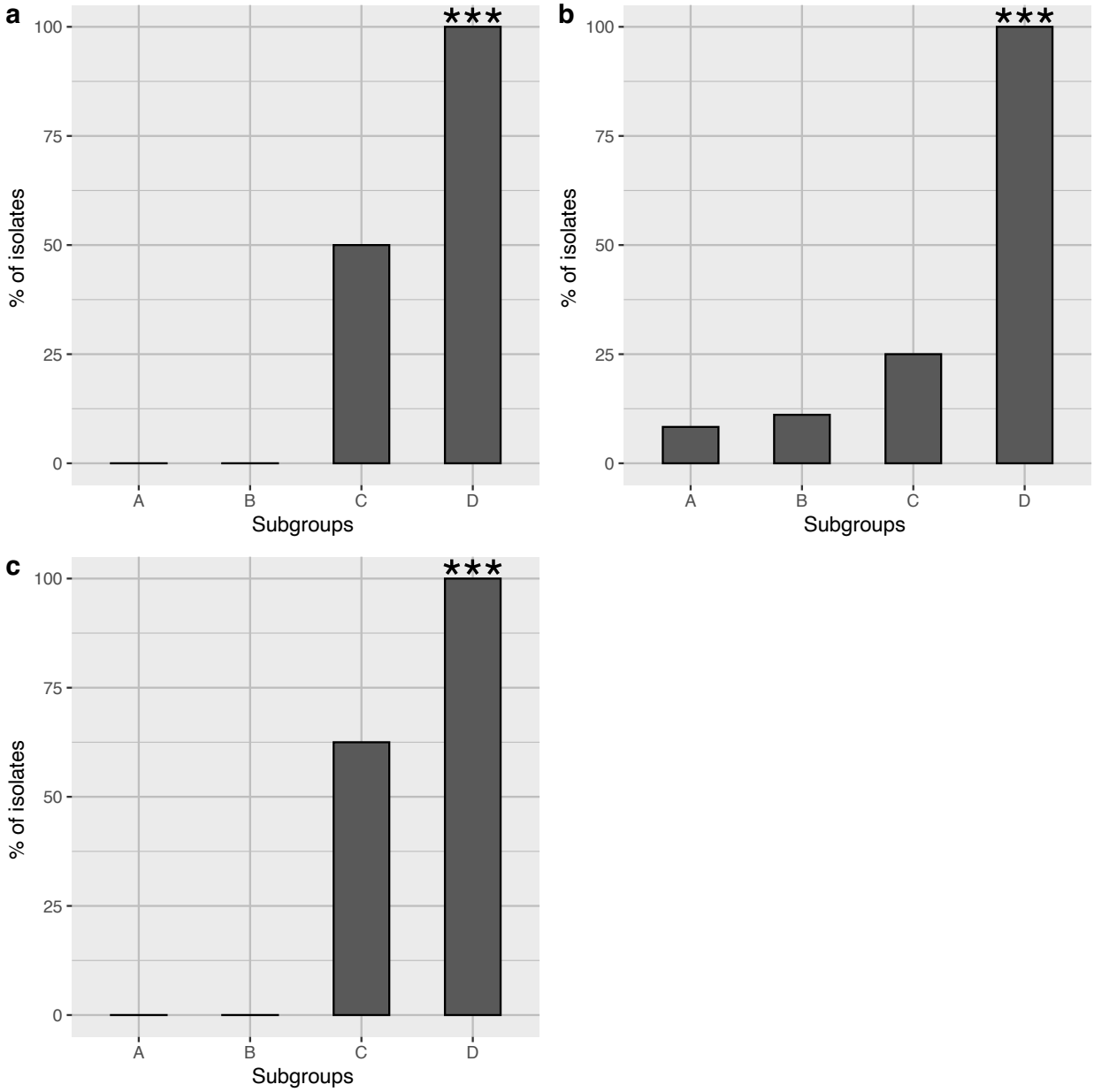


Fig. 5 Subgroup D has minimal overlap with the other subgroups in carbon source utilization. The degree of variation based on carbon sources utilization between subgroup D and subgroups A and B was significant (Pairwise ADONIS, $P < 0.05$, after Bonferroni adjustment). The variation in carbon sources 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, $R^2 = 0.42$, $p < 0.05$).

633

634

635



636

637 **Fig. 6** Subgroup D could be differentiated from the other subgroups based on its capacity to

638 utilize inosine (a), uridine 5-monophosphate (b), and turanose (c). The utilization of these three

639 carbon sources is significantly (Chi-square test, $p < 0.05$) associated with subgroup D.

640 **Table S1** Bacterial isolates and whole-genome sequences used for pangenome analysis and
641 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

642

643

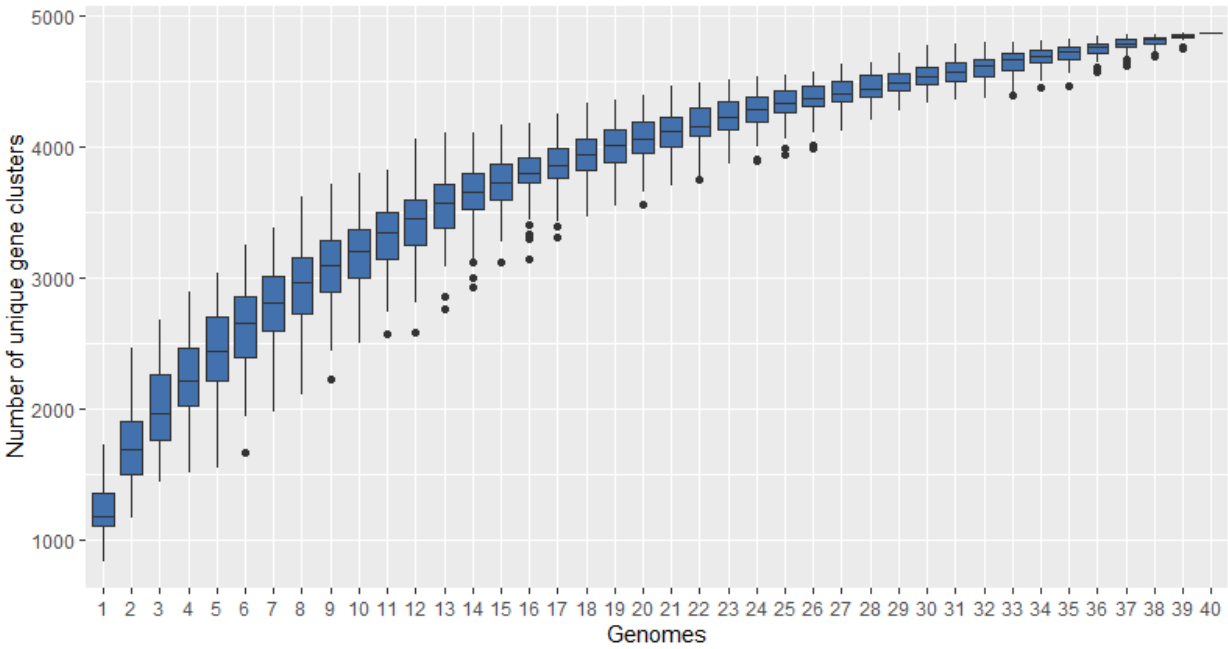


Fig. S1 Rarefaction curve of the pangenome of *Gardnerella* spp..