A Syntenic Pangenome of *Gardnerella* Reveals Novel Plasmids and Phage, Taxonomic Boundaries, and Species-Level Stratification of Metabolic and Virulence Potential

Heather K. Bouzek*[1,2], Martha A. Zepeda-Rivera*[1,2], Sujatha Srinivasan[2], Elliot M. Lee[2], Susan M. Strenk[2], Dakota S. Jones[2], Elsa F. McMahon[2], Tina L. Fiedler[2], Marko Kostovski[3], Michael T. France[4,5], Jacques Ravel[4,5], David N. Fredricks [‡2], and Christopher D. Johnston[‡1,2]

[1] Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[2] Vaccine and Infectious Disease Division, Fred Hutch Cancer Center, Seattle, Washington, USA

[3] Institute of Microbiology and Parasitology, Faculty of Medicine, University Ss Cyril and Methodius in Skopje, Skopje, Macedonia

[4] Center for Advanced Microbiome Research and Innovation, Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA

[5] Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

[‡]Address correspondence to:

David N. Fredricks, MD, dfredric@fredhutch.org

Christopher D. Johnston, PhD, cdjohnston@mdanderson.org

* These authors contributed equally to this work.

Abstract

Gardnerella species are central to bacterial vaginosis (BV), a condition affecting nearly one in three women of reproductive age and associated with preterm birth and increased susceptibility to sexually transmitted infections. Despite decades of study, progress in defining *Gardnerella* diversity has been hindered by inconsistent taxonomy and poor-quality genomic resources. Here, we sequenced 392 new *Gardnerella* isolates from asymptomatic and BV-associated microbiota and integrated this collection with all publicly available genomes. After stringent curation, we generated a high-quality reference set of 313 distinct genomes that underpins a comprehensive taxonomic framework. Using average nucleotide identity (ANI), digital DNA–DNA hybridization (dDDH), and phylogenomics, we resolved 21 genomic lineages encompassing 11 species and 15 subspecies, each assigned a provisional formal name. Integration of complete long-read assemblies enabled construction of the first syntenic *Gardnerella* pangenome, revealing lineage-specific repertoires of virulence, metabolic, and defense systems - including variable sialidases (NanH), vaginolysin, and amino-acid biosynthetic pathways - and defining conserved genomic architecture across species. Comparative methylome profiling further highlighted restriction-modification system diversity that may influence genetic exchange. Finally, we identified the first native cryptic plasmids in *Gardnerella*, overturning the assumption that the genus lacks plasmids, and demonstrated their use in generating a replicative *E. coli*-*Gardnerella* shuttle vector. Together, these results establish a complete genomic and functional framework for *Gardnerella*, providing a reproducible foundation for mechanistic and translational studies of BV and a model for resolving taxonomy and functional stratification in other urogenital-associated bacteria.

22 **Main**

23 Bacterial vaginosis (BV) is a common and global condition characterized by an imbalance in the
24 vaginal microbiota[1-3], associated with significant health risks, including increased susceptibility to
25 sexually transmitted infections[4,5] and adverse pregnancy outcomes such as preterm birth and low
26 birth weight[6-8]. While molecular methods have associated many bacterial species with BV,
27 *Gardnerella vaginalis* was the first bacterium isolated from individuals affected with BV[9], and was
28 long thought to be the sole causative agent of this condition[1,10,11] due to its high virulence
29 potential[3,12-20], its presence in 95–100% of BV cases[11,21,22] and its ability to thrive in the vaginal
30 environment[10,19,20,23-25]. Our previous research revealed that while *Gardnerella vaginalis* is
31 correlated with a greater risk ratio for BV than other validly named *Gardnerella* species, no single
32 *Gardnerella* species is specific for BV based on a *cpn60* qPCR. In fact, the presence of a high
33 *Gardnerella* species diversity was identified as a risk factor for the development of BV in our
34 longitudinal study[22]. The presence of *Gardnerella* species in both BV-affected and asymptomatic
35 individuals suggests the potential for both commensal and pathogenic lineages or states[3]. However,
36 the genetic capabilities of individual *Gardnerella* lineages remain unclear, partially due to
37 inconsistent taxonomic assignments and overly complex *Gardnerella* nomenclature[19], despite
38 laudable efforts from multiple research groups[26-35]. Further, many of the public *Gardnerella*
39 genome assemblies are duplicated, contaminated, or incompletely assembled and did not pass our
40 quality metrics for pangenome analysis. Thus, we sequenced over 392 *Gardnerella* isolates
41 genomes from both asymptomatic individuals or those with BV and combined these with all
42 available public genomes to generate a highly curated and deduplicated collection for analyses. We
43 applied comprehensive comparative genomic analysis of this curated collection to taxonomically
44 identify *Gardnerella* lineages and propose a single naming scheme that is reflective of their

45 phylogenetic placement and historical context. Syntenic pangenome partitioning revealed lineage-
46 specific repertoires of metabolic pathways, virulence-associated genes, and defense systems.
47 Collectively, this study integrates pangenome analyses, genome-based taxonomy, and experimental
48 validation to define species boundaries within *Gardnerella*, characterize lineage-specific metabolic
49 and virulence-associated potential, and identify mobile genetic elements, including a cryptic
50 plasmid that enables the development of a shuttle vector for the genetic engineering of *Gardnerella*.

51 **Results**

52 ***Gardnerella* and *Bifidobacterium* are distinct genera**

53 We obtained genomes for 392 *Gardnerella* isolates (Fredricks and Ravel bacterial isolate
54 collections) from asymptomatic individuals and those with BV. We curated these and additionally
55 available public *Gardnerella* genomes (n=271 within the NCBI database as of May 2025) using
56 stringent quality metrics to remove duplicate and low-quality genomes (Methods, Supplementary
57 Table 1). This resulted in a final curated set of 312 genomes for comparative genomic analyses
58 (Supplementary Table 2). Analysis of the *cpn60* gene (which encodes a 60 kDa chaperonin)
59 currently provides the most reliable method for *Gardnerella* species typing, as these species are
60 indiscernible by traditional 16S rRNA gene sequence typing[36,37]. BLASTn analysis of the *cpn60*
61 gene of these isolates supported their identification as *Gardnerella* species (Supplementary Table
62 2).

63 Despite the classification of these genomes as *Gardnerella* against and within the NCBI
64 database, initial speciation against the Genome Taxonomy Database (GTDB) assigned many of
65 these genomes to *Bifidobacterium* (Supplementary Table 2). However, *Bifidobacterium* is known
66 to be a polyphyletic group, consisting of at least six distinct clades[38,39], whose members exhibit

67 substantial genetic diversity. The separation of *Gardnerella* from *Bifidobacterium* has previously
68 been demonstrated through comparisons of genome length, G+C percentage, and overall genome
69 composition[31,39]. Consistent with theses prior observations, comparisons of genome length and
70 G+C percentage between our *Gardnerella* genome set, *Bifidobacterium,* and closely related
71 scardovial genera (e.g., *Alloscardovia*[40], *Neoscardovia*[41], *Aeriscardovia*[42], *Scardovia*[43]) (n=229
72 genomes; Supplementary Table 3) indicate that *Gardnerella* genomes are significantly shorter
73 ($p<0.001$, Wilcoxon *post hoc* test) and have lower G+C content ($p<0.001$, Wilcoxon *post hoc* test)
74 than other *Bifidobacterium* members (Extended Data Fig. 1a-b). Notably, *Gardnerella* shares a
75 human-focused and pathogenic niche with *Alloscardovia, Aeriscardovia* and *Scardovia* and clusters
76 more closely with the scardovial genera across these genomic features than with *Bifidobacterium.*
77 Additionally, average amino acid identity (AAI) values between *Gardnerella* and *Bifidobacterium*
78 fell below genus thresholds of 65%[44] (Extended Data Fig. 1c). Together, these results confirmed
79 that our final curated genome set comprises only *Gardnerella* species, further supporting previous
80 evidence that *Gardnerella* and *Bifidobacterium* represent distinct genera within the
81 *Bifidobacteriaceae* family.

82 **Identification and proposed revised nomenclature for *Gardnerella* lineages**

83 Given the importance of *Gardnerella* species diversity in BV[21], we first assessed the number
84 of lineages present in our genome collection. *Gardnerella* has been classified using various
85 approaches[30-33,45,46], but consistent naming has been a challenge for *Gardnerella* despite extensive
86 efforts[26-28,32] (Fig. 1). Early subtyping by Piot *et al.*[26] relied on sugar hydrolysis phenotypes and
87 later expanded to 17 biotypes[27]. Genomic methods such as restriction fragment length
88 polymorphism (RFLP)[28] and amplified ribosomal DNA restriction analysis (ARDRA)[29] identified

89 several distinct genotypes. PCR and whole-genome sequence comparisons introduced two new
90 naming schemes including designations "A - D"[37] or "1 - 4"[30] which led to hybrid labels ("C/1",
91 "B/2", "A/4", and "D/3")[47]. These species were then reframed as reassorted ecotypes[31]
92 incorporating the genomic context of the distinct types. With widespread whole-genome
93 sequencing, additional standards based on ANI emerged, describing thirteen genomic species
94 (Gsp)[32], nine genomospecies (GS)[33], or nine different genotypes (GGtypes)[35]. Subsequent formal
95 naming of a subset of these species included *G. vaginalis*[9], *G. pickettii*[48], *G. piotii*[32], *G. leopoldii*[32],
96 *G. swidsinskii*[32], *G. greenwoodii*[48], *G. lacydonensis*[49], *G. bretellae*[49], *G. massiliensis*[49], and *G.*
97 *phocaeensis*[49] (Fig. 1). Despite the use of diverse classification schemes, phylogenetic analyses
98 across studies consistently recover a deep bifurcation within *Gardnerella*, with the same two major
99 lineages forming the basal split of the genus. While this division has been explicitly designated as
100 "Set A" and "Set B" only by Ahmed *et al*., analogous groupings confirm the primary branching
101 structure in multiple independent phylogenetic frameworks[30-35,37,46,50,51].

102 Consistent with this, *cpn60* gene sequences and reference-free whole-genome phylogenies
103 separated all genomes in our collection into two deeply divergent lineages (Fig. 2, Extended Data
104 Fig. 2, Supplementary Table 4) here maintaining the Set A and Set B designation. To determine the
105 number of *Gardnerella* species represented, we calculated pairwise genome-wide average
106 nucleotide identity (ANI) values across all genomes and examined their distribution (Extended
107 Data Fig. 3a). In accordance with prior large-scale analyses demonstrating a pronounced
108 discontinuity in ANI values around the canonical prokaryotic boundary, most inter-lineage
109 comparisons fell well below 95%, whereas closely related genomes clustered above 95% with an
110 exceedingly small fraction of the sequence comparisons falling above 94% (Extended Data Fig.
111 3a). Because ANI and digital DNA–DNA hybridization (dDDH) values were strongly colinear

112 across the dataset, dDDH did not provide additional resolution but independently supported the
113 same groupings: all genomes that clustered together by ANI also exceeded the 70% dDDH species
114 threshold (Extended Data Fig. 3b–c). Thus, species boundaries inferred from ANI were fully
115 congruent with the universally accepted 70% dDDH. Integrating whole-genome phylogenies with
116 ANI/dDDH clustering, we identified eleven *Gardnerella* species. Within five of these species,
117 additional phylogenomic structure supported sub-species level differentiation, resulting in twenty-
118 one distinct *Gardnerella* lineages (Fig. 2, Supplementary Table 4). These species-level assignments
119 were congruent with the classification scheme proposed by Potter[33]. The *Gardnerella* species
120 *bretellae, lacydonensis, leopoldii, phocaeensis*, *pickettii, piotii*, and *swidsinskii* were previously
121 described as distinct species when integrating phenotypic characterization with genome-based
122 analyses. In our dataset, these lineages likewise form well-supported monophyletic clades in whole-
123 genome phylogenies and remain clearly differentiated from other *Gardnerella* taxa. However, when
124 evaluated across the expanded genome collection as analyzed here, pairwise ANI and dDDH
125 comparisons reveal that these lineages consistently fall within the genomic coherence of broader
126 species-level clusters defined by ≥95% ANI and ≥70% dDDH thresholds. While they are
127 phylogenetically structured and phenotypically distinguishable, they do not cross the currently
128 accepted genomic discontinuity thresholds that would necessitate recognition as separate species
129 under genome-based standards.

130 Given the extensive history of overlapping and partially incompatible classification
131 schemes for *Gardnerella* (Fig. 1), we next harmonized existing species and subspecies names and
132 established a new nomenclature across all identified lineages (Fig. 1, Fig. 2, Supplementary Tables
133 2 and 4). Details on the rationale for each proposed nomenclatural revision, including species-

specific functional attributes and ecological isolation sources, are provided in the Protologues. Our aim was to integrate previously used species and subspecies names to maintain a consistent framework. The historically recognized "*G. vaginalis*" corresponds to a subspecies-level clade within a broader species and is designated here as *G. vaginalis* subspecies *vaginalis*, preserving historical continuity while reflecting its phylogenetic placement (Fig. 1, Supplementary Table 1, Supplementary Table 4). Similarly, we propose distinct species names for lineages with well-supported genomic and phenotypic divergence, including *G. bivia*, *G. hutchinsoni*, and *G. washingtoni*, among others (Extended Fig. 3g). To minimize ambiguity in referring to *Gardnerella* species in this paper and future efforts, we intentionally propose species names that can be abbreviated with unique two letter abbreviations (*Gb*, *Gg, Gh, Gi*, *Gk*, *Gl*, *Gm, Gp, Gr*, *Gv* and *Gw*), while individual subspecies were designated with discernable three-letter abbreviations (except for *Gardnerella piotii* subsp. *pickettii* and *Gardnerella piotii* subsp. *piotii,* which each require four letters, *Gppk* and *Gppt*) (Fig. 1, Supplementary Table 4).

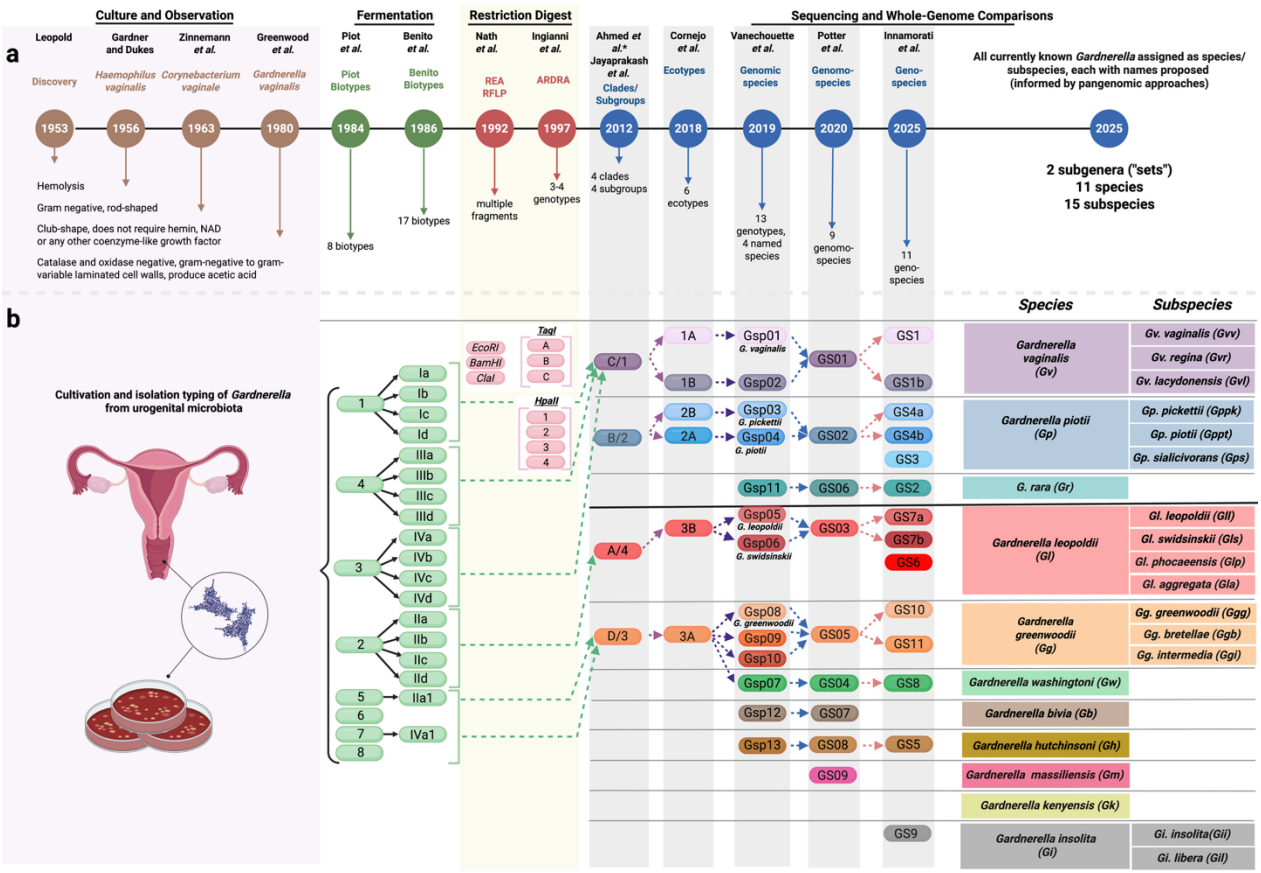
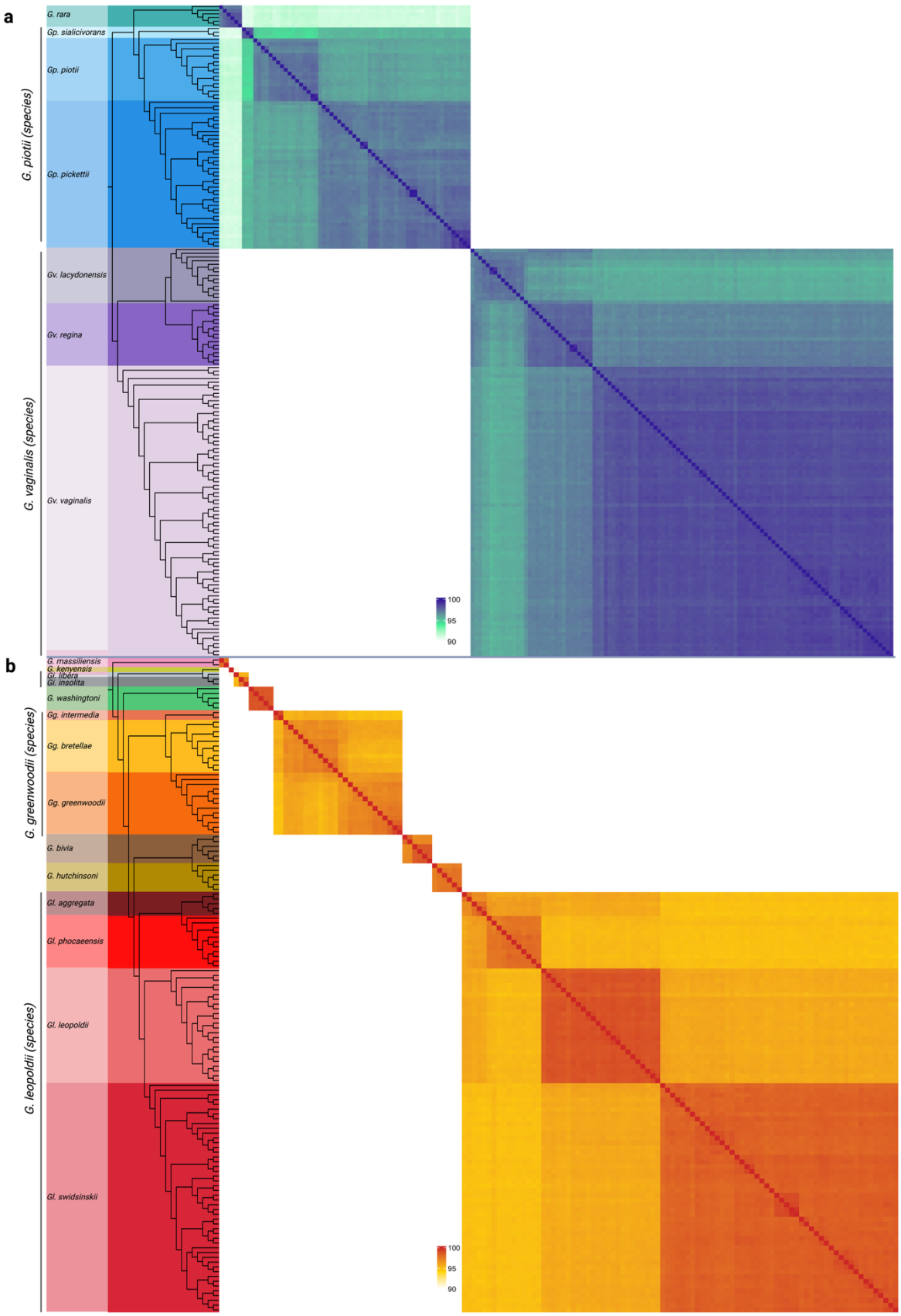


Figure 1: Historical context of *Gardnerella* nomenclature. a, From top to bottom, the schematic shows the identification method by type, publication first author, names given to *Gardnerella* or naming framework, date of publication, and resulting conclusions from naming schemes. **b**, *Gardnerella* subtyping resulting from the publication in (**a**), where each column represents a subset of *Gardnerella*. Arrows indicate consistent lineages between studies without implying a definitive link. The two right-most columns contain updated lineage assignments with corresponding species and subspecies nomenclature proposed within this study.

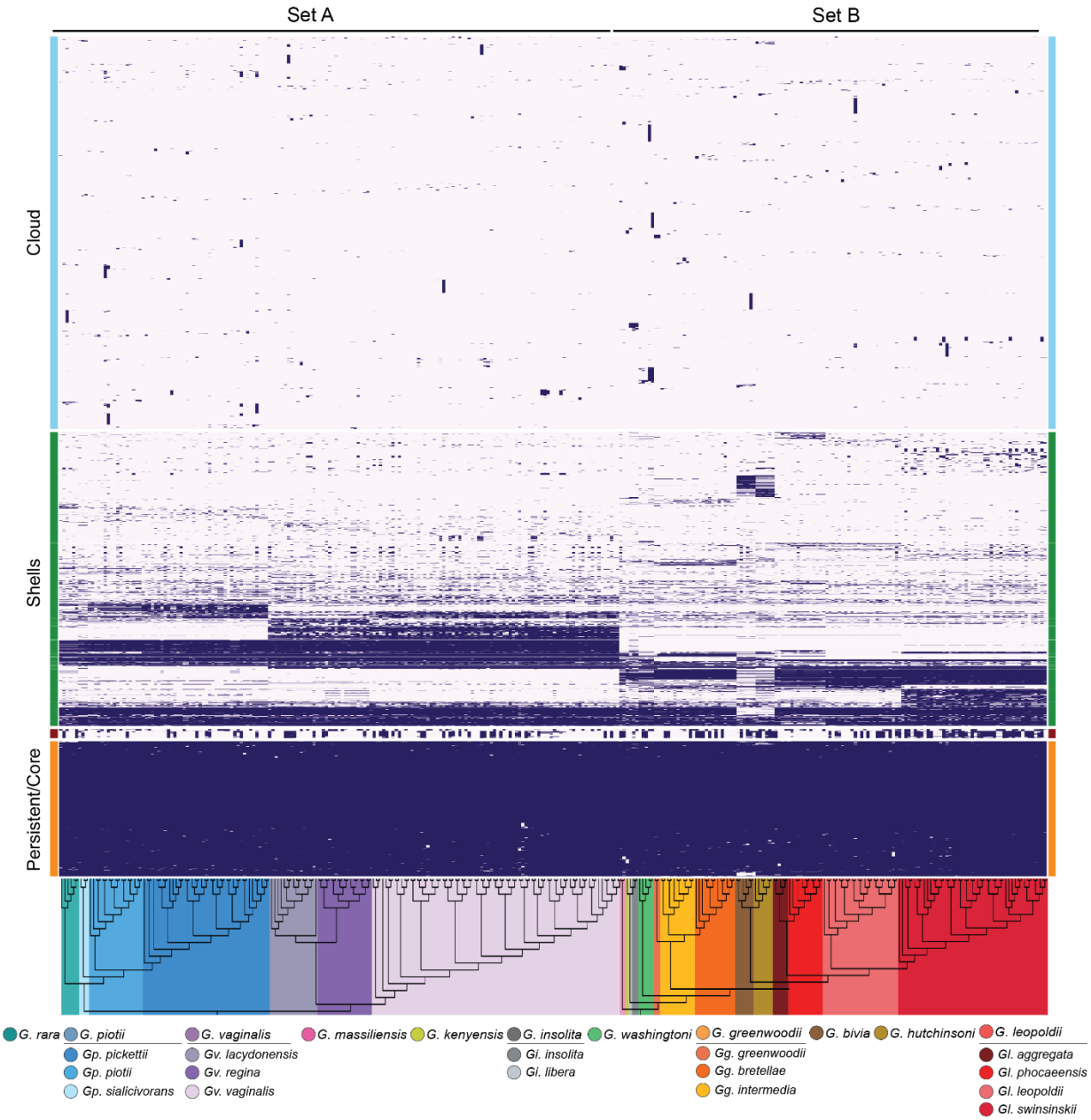

a
G. rara
Gp. sialicivorans
Gp. piotii
Gp. pickettii
G. piotii (species)
Gv. lacydonensis
Gv. regina
Gv. vaginalis
G. vaginalis (species)
100
95
90
b
G. massiliensis
G. kenyensis
Gl. libera
Gl. insolita
G. washingtoni
Gg. intermedia
Gg. bretellae
Gg. greenwoodii
G. greenwoodii (species)
G. bivia
G. hutchinsoni
Gl. aggregata
Gl. phocaeensis
Gl. leopoldii
Gl. swidsinskii
G. leopoldii (species)
100
95
90

154 **Figure 2: Average nucleotide identity values between *Gardnerella* genomes.** Average nucleotide
155 identity values between *Gardnerella* genomes in **a**, Set A and **b**, Set B. Key for each panel indicates
156 corresponding values to each color. White represents values below 90%. Genomes are organized
157 based on a kSNP4-generated whole-genome reference-free phylogeny. Figure is a composite of the
158 kSNP4 tree created in iTOL and the map sorted by the values in the kSNP4 tree as created using
159 ComplexHeatmap in R.

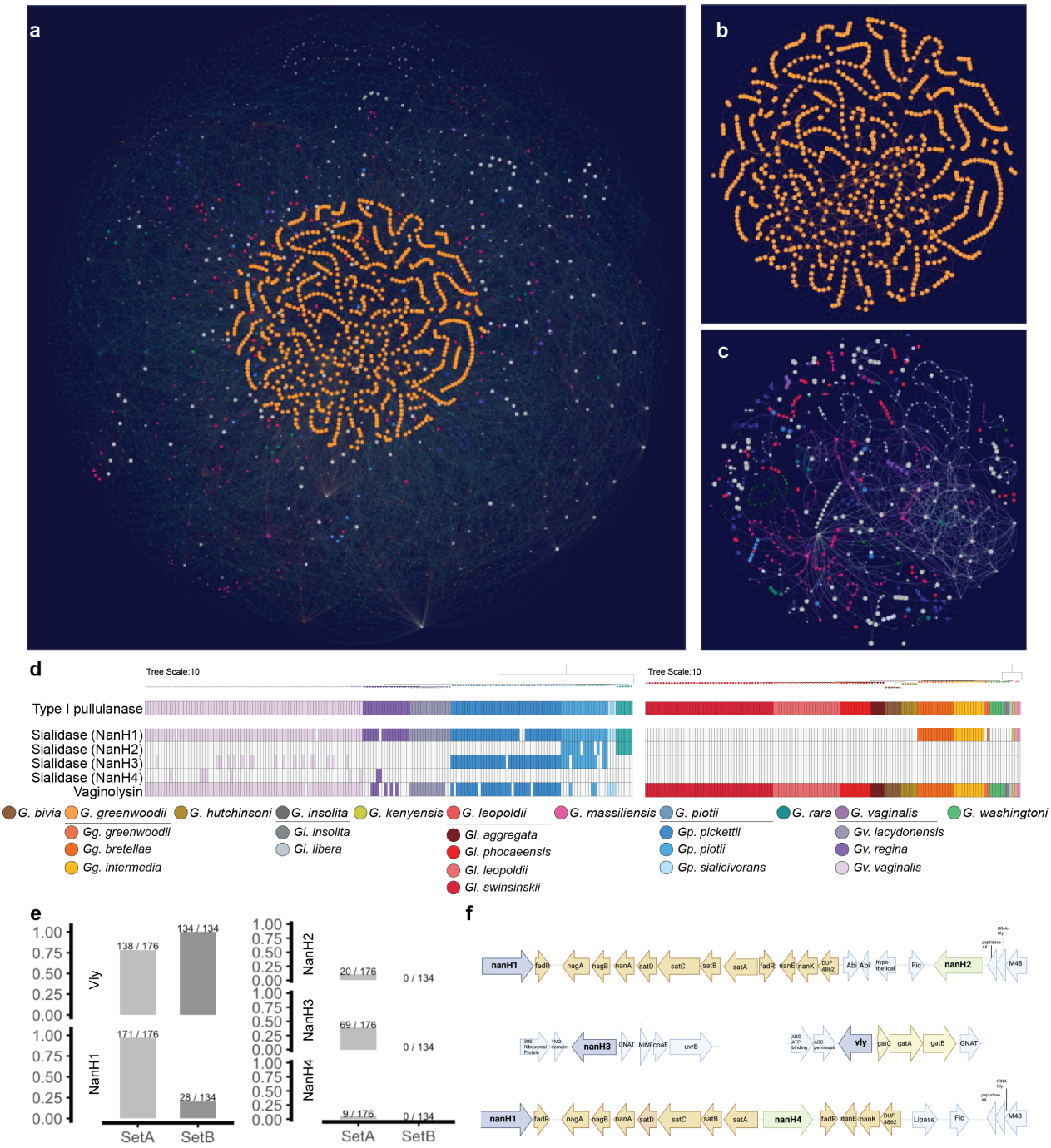
160 **A syntenic pangenome for *Gardnerella***

161 Having stringently identified *Gardnerella* species and subspecies lineages in our genome
162 set and assigning names that adhered as closely to precedent, where possible, we were next able to
163 investigate the genetic features associated with each lineage. Pangenome analysis provides a
164 comprehensive view of the genetic repertoire within a group of bacteria by identifying genetic
165 features that are well-conserved across most (>95%) genomes ("persistent" or "core" genome)
166 versus those that are only present in a subset of genomes ("cloud" or "accessory" genome). While
167 the core genome often contains genes involved in essential cellular processes[52,53], analysis of the
168 accessory genomes can provide key insights about genetic diversity, adaptation, and evolutionary
169 relationships. For example, previous pangenome efforts in *Gardnerella* have expanded taxonomic
170 groupings, identified metabolically essential genes, and uncovered virulence and niche adaptation
171 features[34,50,54]. However, bacterial genes often function in co-regulated units or clusters involved
172 in specific biological processes (e.g., operons). To identify putative operons within the *Gardnerella*
173 pangenome, we applied a syntenic approach that accounts for gene order by using the Partitioned
174 Pangenome Graph of Linked Neighbors (PPanGGOLiN) tool[52] (Figs. 3-4, Supplementary Table
175 5). Briefly, in these analyses, the coding sequences of genes are grouped into "nodes" based on

176 amino acid similarity and visualized as circular nodes, with node size reflecting the relative
177 presence of that gene group across the genome set (Fig. 4a-c). Nodes representing neighboring
178 genes are linked into syntenic units called "modules" (Extended Data Fig. 4). This analysis clearly
179 distinguished between the core ("persistent" partition) and accessory genomes, and enabled
180 identification of accessory genetic elements present at intermediate frequency ("shell" partitions)
181 or low frequency ("cloud" partition) (Fig. 3). By leveraging modules and pangenome partitions,
182 conserved arrangements of genes can be traced across multiple genomes and enables insight into
183 the functional and evolutionary dynamics driving diversity at the genus, species, and strain levels.
184 In our collection, the core genome of *Gardnerella* reflected approximately 16% of identified
185 nodes, enriched in essential cellular functions including translation and ribosomal structure,
186 transcription, RNA processing, signal transduction, energy production, and metabolism of
187 nucleotides, amino acids, inorganic ions, and lipids (Fig. 3, Supplementary Table 5). Periodically,
188 these conserved nodes are interrupted by shell-associated nodes (Fig. 4a; nodes in distinct colors,
189 branching from orange nodes), which mark points of evolutionary divergence. These shell nodes
190 often represent horizontal gene transfer (HGT) events and are commonly associated with alternate
191 evolutionary pathways. We noted node differences between shell partitions that were consistent
192 with virulence, metabolism, and defense systems, warranting further investigation.



193 **Figure 3: *Gardnerella* pangenome partitions.** The tile plot displays the presence and absence of
194 nodes in each partition where each row represents a node, and each column corresponds to a
195 *Gardnerella* genome. Colors on the side correspond to core/persistent (orange), phage (red), shells
196 (green), and cloud (blue) partitions. Genomes are ordered by kSNP4 phylogenies of Set A and Set
197 B, with colors on dendrogram indicating lineage.



198 **Figure 4: Syntenic maps of the *Gardnerella* pangenome. a–c,** Visualizations of the
199 PPanGGOLiN-generated *Gardnerella* pangenome. Each node represents a gene group, with edges
200 connecting syntenic neighboring genes. Node size reflects the relative frequency of the gene group
201 across *Gardnerella* genomes. Colors indicate core (orange) or shell (other colors) partitions. **b** and

202 **c**, The core and shell partitions are shown individually. **d**, map of the presence and absence of the
203 sialidase genes (*nanH1-nanH4*), type II α-amylase-pullulanase and vaginolysin as colored by the
204 species and subspecies of *Gardnerella*. **e**, Presence–absence bar chart showing the percentage of
205 conserved *Gardnerella* vaginolysin and the sialidase genes across each *Gardnerella* set (gray). Raw
206 presence/absence counts are shown above each bar. **f**, Gene context maps for the sialidase loci.

207 **Virulence factors and predicted pathogenicity across *Gardnerella* lineages**

208 Colonization of the vaginal niche by *Gardnerella*, and its transition to a pathogenic state,
209 depend on mucin degradation, adherence to the vaginal epithelium, and the formation of
210 structurally stable biofilms[13,55]. Canonical *Gardnerella* virulence factors implicated in these
211 processes include exported sialidases and vaginolysin[56], a cholesterol-dependent cytolysin. Among
212 *Gardnerella*-associated sialidase isoforms, NanH2 and NanH3 account for the majority of
213 extracellular sialidase activity, whereas NanH1 lacks a signal peptide and is therefore predicted to
214 function intracellularly for sugar catabolism[57]. Consistent with this role, *nanH1*[58] is embedded
215 within a conserved operon across genomes (Extended Figure 5e). In contrast, *nanH2* and *nanH4*[59]
216 are frequently located in proximity to this operon but are associated with mobile genetic elements,
217 suggesting more recent paralogous diversification and horizontal acquisition. Notably, *nanH3* is
218 uniquely not operon-associated; instead, it is found as a solitary gene within defense-associated
219 genomic regions, often adjacent to prophage elements, consistent with a distinct evolutionary
220 trajectory and potential phage-linked mobility. Within our pangenome dataset, we identified
221 lineage-specific nodes encoding vaginolysin, which contributes to host interaction by binding to
222 and lysing epithelial cells of the vaginal mucosa[55], and multiple enzymes involved in the stepwise

223 degradation of mucin-associated glycans (Fig. 4d-e, Extended Data Fig. 5, Supplementary Table
224 5). These include secreted sialidases and additional carbohydrate-active enzymes, such as L-
225 fucosidase[60-62] and galactosidase[27] that facilitate mucin catabolism when exported[63].

226 A Type II α-amylase-pullulanase protein was well conserved (present in 100% of
227 *Gardnerella* genomes), in agreement with previous reports[64] (Fig. 4d). Comparisons between
228 *Gardnerella* Set A and Set B showed that the four sialidase genes (*nanH1-nanH4*) are significantly
229 enriched in Set A genomes, while vaginolysin (*vly*) is significantly enriched in Set B genomes (Fig.
230 4e, Extended Data Fig. 5), highlighting lineage-dependent differences in metabolic capacity and
231 the distribution of putative virulence-associated genes across *Gardnerella*.

232 **Predicted metabolic capabilities across *Gardnerella* lineages**

233 Given the observed difference in metabolism-related nodes, we bioinformatically predicted
234 the biosynthesis of amino acids and the catabolism of small carbon sources across our *Gardnerella*
235 genome dataset (Fig. 5a, Fig. 5f, Supplementary Table 6). The biosynthesis of glutamine, glycine,
236 and proline was well conserved (>99% of genomes) while a high proportion of genomes (89%)
237 harbored the capacity for cysteine biosynthesis (Fig. 5a). Asparagine biosynthesis was enriched for
238 overall for Set B versus Set A, but within Set A, 100% of the *Gr* strains carried the genetic pathway.
239 Within Set B, asparagine biosynthesis was enriched for *Gg* (although depleted for *Ggi*) and the *Gll*
240 lineage alone amongst other *Gl* (Fig. 5a, b). Methionine biosynthesis is also enriched throughout
241 Set B, but absent in all of Set A. Within Set B methionine biosynthesis is notably lacking for strains
242 of *Gb* and *Ggg* (Fig. 5a, Fig. 5d). Threonine biosynthesis is much more sporadically throughout Set
243 A as compared to Set B (Fig. 5a, Fig. 5e). Additionally, threonine synthesis is more likely to be
244 found in *Gv* compared to other Set A (Fig. 5a, Fig. 5e). Although cysteine biosynthesis is well

245 conserved this pathway is entirely absent in *Gh* genomes (Fig. 5a) and depleted in *Gps* as
246 compared to other *Gp* genomes (Fig. 5a, Fig. 5e) .

247 Genes for the catabolism of ethanol appear to be a well-conserved feature of *Gardnerella*
248 genomes (Fig. 5f). Genomes across Set A were additionally significantly enriched for the
249 catabolism of alanine, gluconate, and serine (Fig. 5f, Fig. 5h, Fig. 5j, Fig. 5k). Notably, lineage-
250 specific associations within Set A were observed for the catabolism of each of these small carbon
251 sources. While gluconate catabolism is significantly enriched in *Gv*, alanine catabolism is
252 significantly depleted in *Gvl* as compared to other Set A members (Fig. 5f, Fig. 5h, Fig. 5j). Serine
253 catabolism was completely absent in *Gr* and (Fig. 5f) and was significantly depleted in *Ggb* and
254 *Gls* compared to other lineages within their respective species (Fig. 5f, Fig. 5k). Interestingly, the
255 catabolism of acetate and deoxyribose was lineage-specific and unique to *Gb* and *Gv*, respectively
256 (Fig. 5f). Deoxyribose catabolism was enriched in the *Gvv* compared to other *Gv* subspecies (Fig.
257 5f, Fig. 5i). Our results indicate that Set A and Set B genomes are distinguished by their relative
258 enrichment of pathways for catabolism of small carbon sources or synthesis of distinct amino acids
259 with lineage-specific associations.

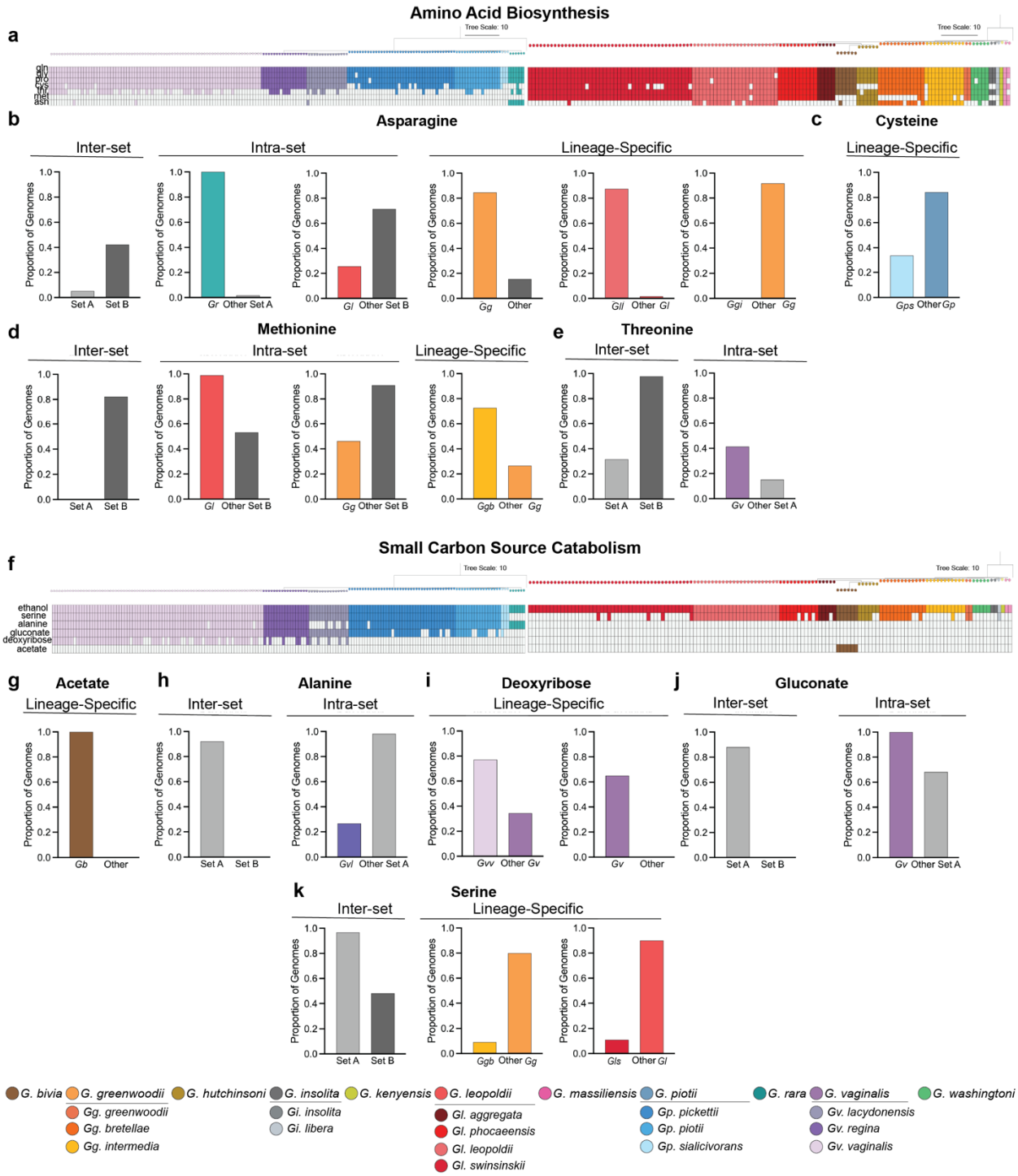


Figure 5: Predictions for metabolic capabilities across *Gardnerella* lineages. Presence-absence map for GapMind-predicted pathways to **a,** synthesize amino acids (AA) or **f,** catabolize small carbon sources (SCS) where each row represents a *Gardnerella* genome. Genomes are ordered by

263 kSNP4 phylogenies of Set A and Set B. Presence by proportion comparison of genomes for the **b-**
264 **e**, biosynthesis of a subset of amino acids or **g-k**, catabolism of a subset of small carbon sources.
265 Colors represent each lineage.

266 **Phage and antiphage defense mechanisms across *Gardnerella* genomes**

267 In our pangenome data set, we noted that multiple nodes were related to phage-associated
268 proteins that encompassed two distinct *Siphoviridae* family *Gardnerella* prophage, including the
269 previously described vB_Gva_AB1[51,65] and a novel phage we have tentatively designated as
270 "pGamma" (Extended Data Fig. 5, Supplementary Table 7). This newly recognized phage is from
271 a distinct family from vB_Gva_AB1 within the Class *Caudoviricetes* in whole genome
272 comparisons[66]. vB_Gva_AB1 is sometimes reported as several different prophages[46,51], the discrete
273 synteny within our PPanGGOLiN pangenome suggests that these variants represent at least one
274 discrete taxon (Extended Data Fig. 5). It is notable that some strains appear to have remnant phage
275 genes, although these appear to be non-syntenic and scattered throughout the genome
276 (Supplementary Table 5). While both vB_Gva_AB1 and pGamma were identified in Set A and Set
277 B genomes, vB_Gva_AB1 was significantly enriched in the Set B species *G. leopoldii* (Bonferroni
278 adj. $p < 0.01$, Fisher's Exact test) and *G. greenwoodii* (Bonferroni adj. $p = 0.4$, Fisher's Exact test)
279 (Extended Data Fig. 6).

280 To counteract phages, bacteria evolved an array of defense systems[67], with the most
281 widespread and well-characterized being restriction-modification (RM)[68,69] and CRISPR-Cas
282 mechanisms[70]. Across our *Gardnerella* genome dataset we identified the presence of sixty different
283 systems, with an average of twelve defenses per genome (Fig. 6a-b, Extended Data Fig. 7,

284 Supplementary Table 8). Of the defense systems with described mechanisms, dXTPase systems[71]
285 were well conserved (99% genomes). Type I RM systems were prevalent (56% of genomes) and
286 did not differ significantly between *Gardnerella* sets. Set A genomes were significantly enriched
287 for some Abi proteins (AbiD[72] and AbiL[73]), CBASS defenses[74,75], and DRT mechanisms[76,77], while
288 Set B showed enrichment for a larger number of defenses, including those with Abi-mechanisms
289 (AbiE[69], Gabija[75], and Lamassu[75,78]), CRISPR-Cas (CRISPR arrays and Cas Type IE proteins), and
290 Type II RM defenses (Extended Data Fig. 8). Across *Gardnerella* lineages, *Gls* showed an
291 enrichment for DNA-modification system (DMS) proteins, while *Gvv* showed an enrichment for
292 AbiL and DRT defenses (Extended Data Fig. 8).

293 Although defense mechanisms evolved primarily to protect against bacteriophages, their
294 activity extends against other horizontally transferred genetic elements, as well as plasmid vectors
295 used during *in vitro* transformation. RM defenses, which are dependent upon detection and
296 destruction of inappropriately methylated foreign DNA, are one of the most prevalent across
297 bacteria. Further, it is notable that approaches to overcome these RM systems during genetic
298 engineering have been shown to significantly improve transformation efficiency across multiple
299 bacterial species[79-81]. Thus, we sought to characterize RM defenses across a subset of our
300 *Gardnerella* genome dataset. We obtained and analyzed methylomes for 42 *Gardnerella* isolates
301 across thirteen lineages within the Fredricks isolate collection (Supplementary Table 2). A total of
302 39 methylated motifs were detected and assigned to Type I (25 motifs), Type II (11 motifs), or Type
303 III (3 motifs) RM systems (Fig. 6c, Supplementary Table 9).

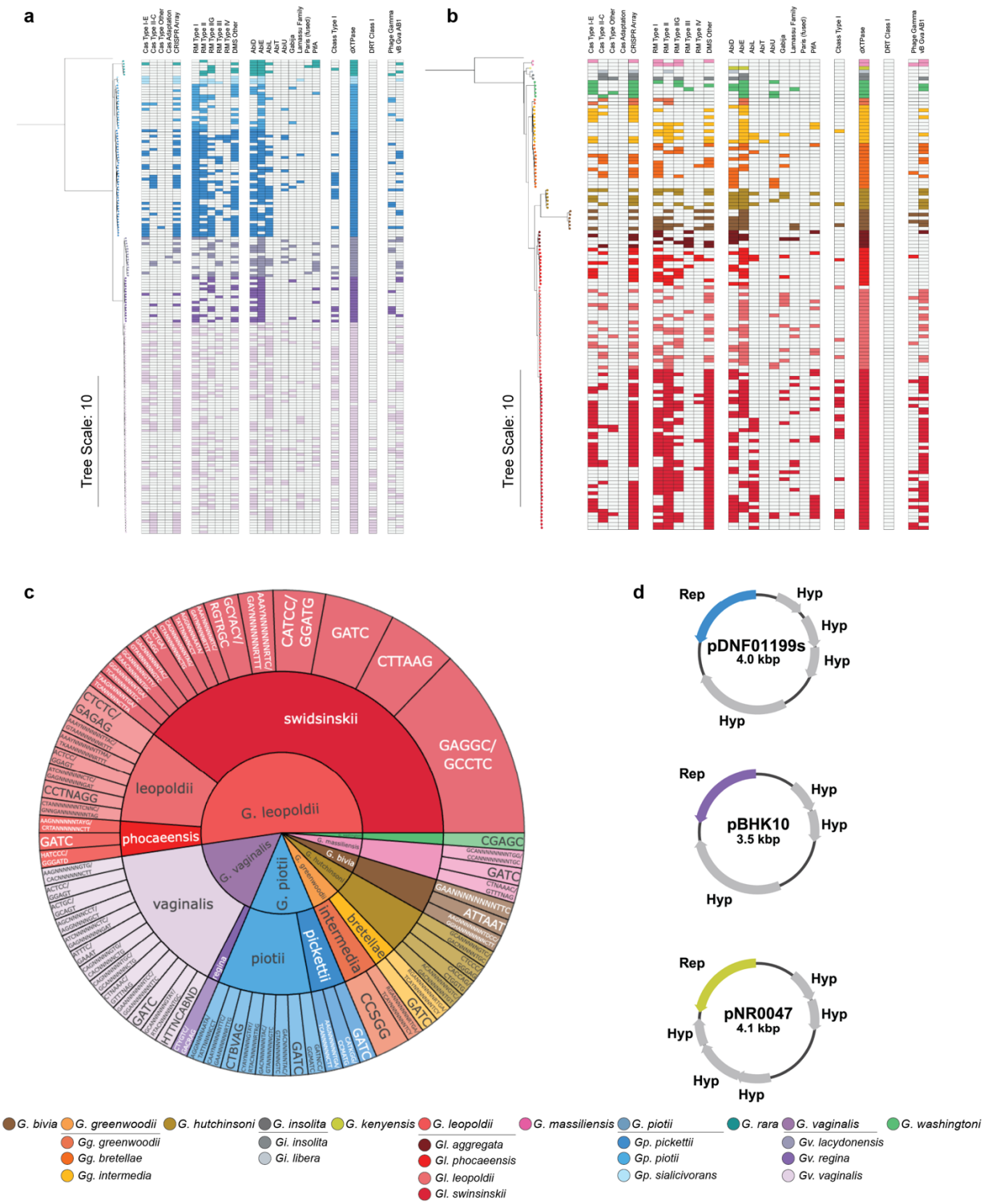


Figure 6: Defense systems, methylation motifs, and plasmid architectures across *Gardnerella* lineages

306 **a-d,** Presence-absence maps for PADLOC-predicted defense mechanisms. Genomes are ordered
307 by kSNP4 phylogenies of **a,** Set A and **b,** Set B. **c,** Sunburst plot shows the sequence of each
308 methyl-modified motif detected within each *Gardnerella* lineage. Size of section represents
309 prevalence within each lineage. **d,** Schematics depict plasmid annotations for pDNF01199S-01,
310 p0756_546_1_1_BHK10-01, and pNR047-01. Full sequences for the plasmids are available in
311 Supplementary Table 10.

312 **Identification of a native *Gardnerella* cryptic plasmid**

313 There is renewed interest in developing genetic tools for engineering *Gardnerell*a
314 species[82,83]. Because *Gardnerella* was long thought to lack mobile genetic elements, such as native
315 plasmids, most efforts have focused on adapting non-*Gardnerella* genetic tools for use in this
316 genus. Recent studies have been extensive and show promising results, achieving targeted
317 chromosomal integrations and deletions using *E. coli*-based suicide vectors[82], as well as
318 reproducible transformation with a *Bifidobacterium-E. coli* shuttle vector[83]. During genome
319 sequencing of the Fredricks clinical isolate collection, we observed an independent circular contig,
320 with high coverage, forming a second replicon within the otherwise closed genome sequence of
321 *Gppk* strain DNF01199S (Fig. 6d, Supplementary Table 10). Annotation of this contig showed
322 three coding sequences for hypothetical proteins and a single putative replication initiation protein
323 (Supplementary Table 11). Subsequent plasmid isolation and validation confirmed the episomal
324 nature of this contig, marking pDNF01199S-01 as the first cryptic plasmid identified within the
325 *Gardnerella* genus.

326 Motivated by this discovery, we assessed whether additional cryptic plasmids were present
327 in our *Gardnerella* pangenome. We reasoned that the current limited knowledge of such plasmids
328 reflects their rarity. As such, we focused on the cloud partition of the accessory genome, which
329 highlights strain-specific genetic factors and collectively comprises 47.23% of the *Gardnerella*
330 pangenome. Using structural sequence insights from pDNF01199S-01, we identified two
331 additional putative plasmid sequences within the cloud partition: one in *Gvv* strain
332 0756_546_1_1_BHK10[84] (p0756_546_1_1_BHK10-01), which was isolated from cervicovaginal
333 mucus in South Africa, and another in *Gk* strain NR047[23] (pNR047-01), isolated from a vaginal
334 swab in Nairobi, Kenya. All three plasmids are approximately 4kb in size, contain coding
335 sequences for putative hypothetical proteins, and share a highly similar replication initiation
336 protein (84.48%-95.26% similar), suggesting a potentially conserved replication mechanism for
337 cryptic plasmids within the *Gardnerella* genus (Fig. 6d, Supplementary Tables 10-11). Given that
338 pDNF01199S-01 replicates stably in *Gardnerella*, we next sought to engineer a *Gardnerella-E.*
339 *coli* shuttle vector suitable for transformation of additional *Gardnerella* isolates. Analysis of our
340 pangenome revealed that 22% of *Gardnerella* genomes encode *tetM*, including *Gppk* DNF01199S,
341 consistent with previous reports on variable tetracycline resistance phenotypes across *Gardnerella*
342 strains[85]. We therefore assessed the growth of both *tetM*+ and *tetM*- isolates in the presence of
343 supplemented tetracycline. As predicted, *tetM*+ isolates were resistant to tetracycline, whereas
344 *tetM*- isolates were susceptible, confirming that the *tetM* cassette confers tetracycline resistance *in*
345 *vitro*. This supports the use of *tetM* as a selectable marker for this vector. To enable the propagation
346 in *E. coli* strains, we combined pDNF01199S-01, the *tetM* cloned from DNF01199S, and an *E. coli*
347 minicircle plasmid backbone (pMC) into a single parental plasmid (herein designated pG01-pp)
348 (Extended Data Fig. 9a-c). As previously reported[86], we induced and purified the pDNF01199S-

349 tetM-MC minicircle vector (pG01-MC). We initially attempted to transform this minicircle into
350 the *tetM*-strain, ATCC 14018^{T} (*Gvv*), but no transformants were recovered (Extended Data Fig.
351 9d).

352 The strain ATCC 14018^{T} encodes a restriction-modification (RM) system which modifies
353 the "GGCC" motif[84] and has been shown to limit transformation efficiency[82]. Thus, we used
354 commercially available *HaeIII* methyltransferase (MTase) to methyl-modify these sites in pG01-
355 MC prior to transformation. Electroporation with methylated pG01-MC successfully transformed
356 *Gvv* ATCC 14018^{T}, but at a very low efficiency (two transformants per µg of DNA). Further
357 optimization of the protocol increased efficiency to 685 transformants per µg of DNA by
358 transforming with 0.5 µg total *HaeIII*-treated vector on competent cells grown with 100 mM NaCl
359 to an OD_{600} of 0.32 and washed with 10% glycerol (Extended Data Fig. 9d). A detailed, step-by-
360 step version of the optimized protocol is available online at
361 (https://www.protocols.io/private/1A47BAD2F1AC11F09AA30A58A9FEAC02).

362 Collectively, these results show that pG01-pp is the first replicative shuttle vector derived
363 from a native *Gardnerella* cryptic plasmid. Together with recent work by the Dillard and Hill
364 groups[82,83], these efforts should establish a robust platform for more mechanistic investigations of
365 *Gardnerella* biology and pathogenesis.

366 **Discussion**

367 Taxonomic inconsistencies and fluctuating nomenclature have made it inherently
368 challenging to delineate the roles of *Gardnerella* lineages in health and disease. In this study, we

369 integrate whole-genome analyses to establish a refined and internally consistent taxonomic
370 framework for the *Gardnerella* genus. Using a curated set of 313 distinct, high-quality genomes,
371 we define clear lineage boundaries and propose an updated *Gardnerella* nomenclature that reflects
372 genome-scale divergence while preserving continuity with existing naming conventions. Although
373 nomenclatural revisions, particularly for well-established organisms such as '*G. vaginalis,'*
374 (updated here to *G. vaginalis* subsp. *vaginalis*), can often be contentious, applying established
375 genomic criteria for species and subspecies delineation is warranted to prevent further taxonomic
376 ambiguity and this misinterpretation of functional studies.

377 While single-locus markers such as *cpn60*, which are widely used for *Gardnerella* typing,
378 can distinguish between *Gardnerella* species, they lack sufficient resolution for subspecies
379 assignments. Accurate subspecies assignments require whole-genome approaches including
380 phylogenomic inference and average nucleotide identity or digital DNA-DNA comparisons. As
381 additional *Gardnerella* lineages are isolated, whole-genome sequencing efforts should allow for
382 continuous integration into this framework, supporting a stable and extensible taxonomy.

383 Our genome-based analyses also intersect with a broader, long-standing debate regarding
384 the placement of *Gardnerella* within the family *Bifidobacteriaceae*. Although, marker gene
385 frameworks such as the GTDB place *Gardnerella* within *Bifidobacterium*, the two groups differ
386 markedly in genome size, G+C content, metabolic repertoire, virulence-associated gene content,
387 mobile genetic element composition, and ecological specialization. Notably, *Gardnerella*
388 genomes are substantially smaller and lack many biosynthetic and metabolic pathways retained in
389 larger *Bifidobacterium* genomes. Such genome reduction is widely interpreted as a hallmark of

390 specialization, reflecting adaptation to host-associated niches where broad metabolic versatility is
391 unnecessary. In this context, reduced genome size represents ecological specialization rather than
392 diminished biological significance.

393 Together, these genome-scale differences suggest that conserved marker phylogeny alone
394 may be insufficient to define biologically meaningful genus boundaries within this family. A
395 parsimonious interpretation is that *Bifidobacterium*, as currently circumscribed, encompasses
396 multiple genomically and ecologically distinct genera. While a formal genus-level revision of
397 *Bifidobacteriaceae* lies beyond the scope of this study, our data provide additional genome-scale
398 evidence supporting the continued recognition of *Gardnerella* as a distinct genus and contribute to
399 an evidence base that may inform future ICNP-compliant taxonomic restructuring. With this
400 refined taxonomic and nomenclature framework, we probed for genetic features and putative
401 operons associated with each lineage by applying a syntenic pangenome approach. These analyses
402 linked *Gardnerella* lineages to distinct repertoires of virulence-associated factors, metabolic
403 pathways, and genetic defense systems, highlighting functional diversity across the genus. Notably,
404 most *Gardnerella* gene content (84%) is species- or strain-variable with many differences related
405 to virulence-associated functions, metabolism, and genetic defense.

406 *Gardnerella* Set A genomes (*Gr, Gp,* and *Gv*) were enriched for sialidases and exhibited
407 increased potential for the catabolism of small carbon sources. We further noted lineage-specific
408 enrichment of established virulence-associated factors, including exported sialidases (*nanH2* in
409 *Gr* and *Gppt*; *nanH3* in *Gppk* and *Gppt*) and vaginolysin (enriched in *Gvv, Gvl, Gps* and *Gppk*).
410 The restricted distribution of these factors suggests functional specialization among lineages that

411 may contribute to mucin degradation and epithelial barrier disruption within the vaginal
412 microbiome. In contrast, Set B genomes (*Gb, Gc, Gg, Gi, Gl, Gk, Gm,* and *Gw*) were enriched in
413 putative virulence-associated proteins (e.g. predicted surface-associated proteins, secretion system
414 components and factors implicated in host interaction), exhibited greater capacity for amino acid
415 biosynthesis, showed a higher propensity to harbor the vG_gva_AB1 prophage, and encoded a
416 broader repertoire of genetic defense mechanisms. Collectively, these features are consistent with
417 distinct colonization strategies among Set B *Gardnerella* lineages, including enhanced persistence
418 on epithelial surfaces, flexibility under nutrient-limited or variable resource conditions, and
419 adaptation to competitive, phage-rich, polymicrobial niches.

420 Mechanistic studies to assess the roles of *Gardnerella* genetic factors, including those
421 identified in this study, have been hindered by limited genetic tools available for *Gardnerella*. In
422 fact, it has previously been reported that *Gardnerella* species lack mobile genetic elements, such
423 as plasmids[30]. However, our extensive genome sequencing efforts reveal the first *Gardnerella*
424 cryptic plasmid naturally occurring within a single cultivated strain, along with additional
425 homologous sequences in publicly available genomes. Through the identification of this cryptic
426 plasmid (pDNF01199S), we developed a replicative *Gardnerella-E. coli* shuttle vector (pG01-pp).
427 Coupled with recent advances in *Gardnerella* species engineering[82,83], this discovery of
428 extrachromosomal and self-replicative genetic elements within the *Gardnerella* genus should
429 facilitate the development of advanced genetic tools. Combined, these foundational studies seek to
430 advance genetic tools for mechanistic studies of *Gardnerella* species. Collectively, overcoming
431 genetic intractability within *Gardnerella* will greatly advance our ability to investigate the role of
432 this genus in health and disease states.

433 Although these interpretations are based on comparative genomic potential and do not
434 directly infer *in vivo* activity or clinical outcomes, taken together they support the view of
435 *Gardnerella* as an ecologically diverse genus composed of lineages with distinct metabolic
436 capacities, host-interaction mechanisms, and mobile genetic repertoires. Different *Gardnerella*
437 seem to occupy overlapping but not identical niches within their human hosts. Some lineages may
438 be better adapted to persistence and competitive growth in the vaginal microbiome and urinary
439 tract, whereas others may be more consistent with transient or context-dependent colonization.

440 **Methods**

441 Study design and participants

442 Vaginal bacterial isolates were collected from human study participants both with and without BV
443 in Seattle, Washington, USA[22,87] or from the Ravel isolate collection, harboring strains from
444 multiple geographic locations, including USA and Africa (Supplementary Table 2). Bacterial
445 isolates consistent with *Gardnerella* morphotypes on Gram stain were subjected to 16S rRNA
446 and/or *cpn60* gene amplicon sequencing to confirm phylogenetic placement in the *Gardnerella*
447 genus[24]. Phylogenetically diverse *Gardnerella* species were selected for whole-genome
448 sequencing. High molecular weight DNA was extracted from each isolate using the MasterPure
449 DNA purification kit (Lucigen; Middleton, WI), with two phenol/chloroform cleanups prior to
450 DNA precipitation where needed. DNA extraction was quantified on a TapeStation 2200
451 instrument run with a Genomic DNA tape (Agilent Technologies; Santa Clara, CA).

452 Genome sequencing and assembly

453 *Sequencing on the Pacific Biosciences SeqI/II Platforms*

454 A total of forty-two strains from the Fredricks collection, and seventeen from the Ravel collection,
455 were selected for single-molecule real-time (SMRT) long-read sequencing using PacBio (SeqI or
456 SeqII, Pacific Biosciences; Menlo Park, CA) at the Fred Hutch Cancer Center or the Genomic
457 Resource Center of the University Maryland School of Medicine. Sequencing libraries were
458 prepared using the SMRTBell Template Prep Kit 1.0/2.0 and were selected for a 3-5kb size cutoff
459 using a BluePippin (Sage Science; Beverly, MA). Sequencing reads were processed via Microbial
460 Assembly software within the Pacific Biosciences' SMRTAnalysis pipeline (v.9.0.0.92188).

461 *Sequencing on the Illumina HiSeq 4000 Platform*

462 332 isolates were selected for Illumina sequencing on the HiSeq4000 platform. Short read libraries
463 were prepared using the KAPA HyperPlus Kit (Kapa Biosystems/Roche; Willmington, MA) with
464 KAPA Single-Indexed Adapter Kit Set B. 35µL of genomic DNA was used as input, and libraries
465 were prepared following the manufacturer's protocol with modifications based on concentration.
466 For samples with 0.5 or 0.2 ng input DNA, the fragmentation enzyme was diluted 1:2 or 1:5 with
467 water. All samples were fragmented at 37°C for 5 minutes. Adapter concentrations varied
468 according to the input, and the adapter ligation was carried out overnight at 4°C for all samples.
469 The post-ligation cleanup was performed with 0.8x Ampure XP beads (Beckman Coulter;
470 Indianapolis IN) and 20 µL of sample was used in library amplification. Post-amplification cleanup
471 was performed with 1x Ampure XP beads; libraries with remaining adapter dimer peaks were
472 cleaned a second time. The final elution was in 25 µL of nuclease-free water. Libraries were run
473 on a TapeStation instrument with a D1000 tape (Agilent Technologies; Santa Clara, CA) to assess
474 quality and concentration. Libraries were sequenced on an Illumina HiSeq 4000 instrument using
475 the 150 bp paired-end protocol. Reads were assembled using Unicycler.

476 Publicly available genomes from the National Center for Biotechnology Information (NCBI)

477 We used publicly available genomes both to provide a direct reference for previous studies and to
478 supplement our own sequencing efforts. We searched GenBank[88] assembly databases for the genus
479 *Gardnerella* and family *Bifidobacteraceae* (without genus assignment). Using the Type (Strain)
480 Genome Server (TYGS)[89] to confirm the genus of these genomes, we discovered that several of
481 the unassigned *Bifidobacteraceae* bacteria in NCBI were *Gardnerella* and included them in our
482 analysis. *Bifidobacteraceae* (without genus assignment) were included if they were collected as
483 part of the same NCBI BioProject which added 20 *Gardnerella* that were not typed further than
484 the *Bifidobacteraceae* family. Additionally, we included three sequences (350_GVAG,
485 476_GVAG, and 842_GVAG) that were initially assigned to *Gardnerella* but were later renamed
486 to *Alloscardovia* to verify this placement. The initial number of potential genomes was n = 649.
487 We removed genomes that were assigned outside of *Gardnerella*. Draft genomes with more than
488 250 contigs, modified in the laboratory, and those with truncated or missing cpn60 sequences were
489 removed due to inherent unreliability.

490 *Bifidobacterium* strains were collected and curated in much the same manner as the *Gardnerella*
491 strains. A total of 98 strains from all known species of the *Bifidobacterium* genus and the closely
492 related scardovial genera (e.g., *Alloscardovia, Aeriscardovia*, *Bombiscardovia*, *Pseudoscardovia*,
493 and *Scardovia*) were downloaded from NCBI. Whenever possible, the type strains were collected.
494 Sequences were assessed using CheckM, TYGS, and ANI via cutoffs similar to those used for the
495 *Gardnerella* strains. *Bifidobacteriaceae* groups were assigned according to Alessandri *et al*[90].

496 Genome quality and comparison assessments

497 Genome completeness and contamination for all newly sequenced and publicly available
498 *Gardnerella* genomes retrieved from NCBI were assessed using CheckM (v 1.0.18)[91]. All genomes
499 were annotated via RAST (v 1.073)[92]. Genomes with contamination greater than 1.5% or
500 completeness below 90% were excluded from downstream analyses.
501 Redundant genome sequences were identified based on pairwise average nucleotide identity (ANI).
502 Genomes sharing > 99.5% ANI with other genomes in the collection were considered duplicates
503 and a single representative genome was retained for each duplicate group (Supplementary Table
504 1). Pairwise average nucleotide identity (ANI) metrics were calculated separately for Set A and Set
505 B genomes using FastANI (v1.34)[93] and visualized using the R packages ComplexHeatmap (v
506 2.27.0)[94]. Digital DNA-DNA hybridization (dDDH)[95] values were calculated via the Genome-to-
507 Genome Distance Calculator (v 3.0)[96] and visualized in R (v 4.5.1 -- 2025-06-13 -- "Great Square
508 Root"). Amino Acid Identity (AAI) values for *Bifidobacteriaceae* type strains were determined
509 using pairwise comparisons of FastAAI (version 2)[44].

510 PPanGGOLiN *Gardnerella* pangenome

511 We used Partitioned PanGenome Graph of Linked Neighbors (PPanGGOLiN v 2.2.0)[52] on our
512 curated genome set to graph and cluster the genes from all strains, find regions of genomic
513 plasticity and synteny[75], and separate the pangenome gene clusters into core, shell, and cloud
514 partitions, with clustering set to 50% coverage and 50% identity. EggNOG[96], InterPro (accessed
515 October 2025)[97], and KEGG Mapper (v 5.2)[98], were run separately on each genome to gather and
516 confirm the CAZyme[99], Pfam[100], and KEGG (all accessed October 2025), designations of the gene
517 clusters. COG categories[101] were assigned based on the consensus of these designations. We added

518 the annotation "Unclassified Metabolism" to capture gene clusters that are clearly related to
519 metabolism, but the exact metabolite is not defined or could be of multiple types.

520 Phylogenomic analyses

521 Barrnap was used to find and extract 16S rRNA sequences for submission to NCBI[98]. The
522 sequences in the PPanGGOLiN for the persistent *cpn60* node were MUSCLE[99]-aligned to the same
523 gene from *Bifidobacterium bifidum* (ATCC 29521) to generate maximum-likelihood distances, and
524 a neighbor-joined dendrogram using the R package phangorn[100]. The resulting tree was visualized
525 and annotated using iTOL (v 7.3)[101]. kSNP4 (v 4.1)[102] was used to construct a reference-free whole-
526 genome phylogenetic tree for *Gardnerella* Set A (kmer=17, FCK=0.167) and *Gardnerella* Set B
527 (kmer=17, FCK=0.797) genomes, separately. *Bombiscardovia bombi* strain DSM19703 was
528 included to root both dendrograms. The resulting parsimony trees were visualized in
529 iTOL[101].GapMind (accessed September 2025)[103-105] metabolism predictions and PADLOC (v
530 2.0.0)[106,107] defense mechanism predictions were annotated where appropriate. Bacteriophage
531 detection was performed by VIBRANT (v. 1.2.1)[108], PHageR (accessed Feb 2026)[109], and the use
532 of phage-annotated nodes to find syntenic modules of phages in PPanGGOLiN. These methods
533 were combined as no one method found all bacteriophage. Phage sequences are mostly as found
534 by VIBRANT, when VIBRANT failed to find the phage sequence they were determined from
535 PPanGGOLiN. Phages were typed by aligning the large terminase sequences from each intact with
536 MAFFT (v7.526) and then creating a maximum likelihood tree with IQTree2 (v. 2.3.6)[110].
537 Assessment of greater taxonomic grouping was determined using VICTOR
538 (https://ggdc.dsmz.de/victor.php accessed March 2026)[66].

539 Isolation of pDNF01199S-01 plasmid

540 *Gardnerella piotii* subsp. *pickettii* strain DNF01199S was grown anaerobically in New York City
541 III (NYCIII) liquid media in an AS-580 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA)
542 in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C. A QIAamp DNA Miniprep kit (Qiagen,
543 Hilden, Germany) was used to isolate pDNF01199S-01 from a 5 mL culture of the bacteria.
544 Isolated DNA was digested with *XhoI* (New England Biolabs, Ipswich, MA) and visualized by gel
545 electrophoresis to confirm plasmid presence.

546 *Gardnerella* tetracycline resistance testing

547 *Gardnerella* strains were grown anaerobically in an AS-580 anaerobic chamber (Anaerobe
548 Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 at 37°C. Bacteria
549 were recovered from frozen stocks on NYCIII agar plates and subcultured twice in 2 mL of NYCIII
550 liquid media. Culture purity was verified by Gram stain. The final cultures were normalized to 0.5
551 McFarland turbidity using fresh NYCIII broth and 25 µL was spotted onto an NYCIII agar plate
552 containing 16 µg/mL tetracycline and an antibiotic-free NYCIII agar plate. 25 µL of *Lactobacillus*
553 *reuteri* CF48-3A was spotted onto each tetracycline plate as a tetracycline-resistant control[102].
554 Tetracycline resistance was determined according to Clinical and Laboratory Standards Institute
555 for antimicrobial susceptibility testing protocols, and any growth on tetracycline NYCIII agar
556 plates was considered resistant to tetracycline.

557 pMC-pDNF01199S-*tetM* construction and minicircle preparation

558 pDNF01199S-01 and *tetM* were both cloned from gDNA isolated from strain DNF01199S. Gene
559 fragments were assembled with pMC by Gibson assembly[103]. The parental construct was

560 transformed into *E. coli* ZYCY10P3S2T (System Biosciences, Palo Alto, CA) by heat shocking
561 according to the SBI Minicircle DNA Technology user manual. Transformant colonies were
562 selected by growing on Luria-Burtani (LB) agar plates containing 50 µg/mL Kanamycin. A
563 successfully transformed colony was chosen and stocked, then induction of minicircles was
564 performed according to the SBI Minicircle DNA Technology user manual by growing the bacteria
565 in induction media containing arabinose. pDNF01199S-*tetM* minicircle DNA was then purified
566 from the induced *E. coli* using a Qiagen Plasmid Midi Kit (Qiagen, Hilden, Germany). The
567 sequence of purified plasmid was confirmed against the construct sequence by whole-plasmid
568 sequencing performed by Plasmidsaurus (Plasmidsaurus, Eugene, OR). "GGCC" sites on the
569 plasmid were protected by incubating 1mg of the DNA with 1µL of *HaeIII* MTase (New England
570 Biolabs, Ipswich, MA) at 37°C for 4hrs in the presence of 160 µM S-Adenosyl methionine (New
571 England Biolabs, Ipswich, MA). Following enzyme treatment, DNA was purified with Monarch
572 PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, MA).

573 Preparation of competent *Gardnerella* cells

574 *Gardnerella vaginalis* subsp. *vaginalis* ATCC 14018^T was grown anaerobically in an AS-580
575 anaerobic chamber (Anaerobe Systems, Morgan Hill, CA) in a gas mixture of 5% CO_2, 5% H_2,
576 and 90% N_2 at 37°C. Cells were recovered from frozen stocks on NYCIII agar plates and
577 subcultured into NYCIII liquid media. 500 µL of culture was subcultured into 5 mL NYCIII media
578 and grown overnight. 4mL of this culture was inoculated into 40mL of modified de Man-Rogosa-
579 Sharpe media[104] supplemented with 5 g/L glucose, 0.4 g/L L-cysteine HCl, 100 mL/L filter-
580 sterilized horse serum, and 100 mM NaCl when stated. Once the target OD_{600} was reached, cells
581 were harvested by centrifugation at 4°C at 4,052 x *g* for 10 minutes. Cells were washed in one of

582 the following three ways: 1) Twice in 25 mL ice-cold maltose citrate buffer (0.2625 g/L citric acid
583 and 0.5 M maltose at pH 5.8), 2) Twice in 25 mL ice-cold, autoclaved, 10% glycerol, or 3) Once
584 in 25 mL ice-cold autoclaved water from a Milli-Q IQ7000 pure water system (Millipore Sigma,
585 St. Louis, MO) followed by a second wash in 25 mL ice-cold maltose citrate buffer. Washed cells
586 were then resuspended in 200 µL of maltose citrate buffer, mixed with 200 µL 80% glycerol, and
587 stored at -80°C in aliquots.

588 *Gardnerella* electrotransformation and verification

589 Competent cells were mixed with plasmid DNA on ice and transferred into pre-chilled 0.1 cm gap
590 Gene Pulser/MicroPulser Electroporation Cuvettes (Bio-Rad, Hercules, CA). Electroporation was
591 then performed on a Gene Pulser Xcell (Bio-Rad, Hercules, CA) at 25 µF, 200 Ω, and 2,000 V.
592 Immediately following electroporation, 950 µL of 37°C pre-warmed NYCIII broth was added to
593 the cells and they were transferred to an AS-580 anaerobic chamber (Anaerobe Systems, Morgan
594 Hill, CA) in a gas mixture of 5% CO_2, 5% H_2, and 90% N_2 to recover at 37°C for 1 hour. 100 µL
595 – 200 µL of transformed cells were then plated on NYCIII agar plates with and without 16 µg/mL
596 tetracycline. Plates were incubated anaerobically at 37°C for 48 hours and transformed colonies
597 on the tetracycline plates were counted to determine transformation efficiency. After the first
598 successful transformation, transformant colonies were grown in 5 mL NYCIII media with 16
599 µg/mL tetracycline and pDNF01199S-*tetM* was recovered from them using a QIAamp DNA
600 Miniprep kit (Qiagen, Hilden, Germany). The sequence of the purified plasmid was confirmed by
601 sequencing performed by Plasmidsaurus (Plasmidsaurus, Eugene, OR). Following transformation
602 of *Gardnerella vaginalis* subspecies *regina* strain DNF00354 with pDNF01199S-*tetM*, the
603 transformation reaction was plated onto NYCIII agar plates with and without 16 µg/mL

604 tetracycline. A small amount of growth from a transformant colony on the tetracycline plate or
605 lawn growth from the control plate was resuspended in sterile water and colony PCR was
606 performed to amplify and detect the *tetM* gene.

607 One microliter of resuspended bacteria was added to a reaction mix of 12.5 µL GoTaq Green master
608 mix (Promega, Madison, WI), 0.5 µL of each primer, and 6.5 µL H_2O. PCR was carried out for 28
609 cycles with an annealing temperature of 60°C and an extension time of 2.5 min. A detailed version
610 of this protocol is available online at:
611 https://www.protocols.io/private/1A47BAD2F1AC11F09AA30A58A9FEAC02.
612 Primer sequences were as follows (5' to 3'):
613 Forward: CGCGACCAAATATTGGTACATTATTACAGCTATTTTGTAATCACGTACTC
614 Reverse: GAGACTGTCGTATTGATGTCACGGGCTTAATAGGTCTTATGTCGG
615

616 **Protologues**

617 **Description of *Gardnerella vaginalis* subsp. *vaginalis* Gardner and Dukes 1955 (Approved**
618 **Lists 1980) Bouzek *et al*. comb. nov.**

619 Basonym: *Haemophilus vaginalis* Gardner and Dukes 1955 (Approved Lists 1980)[9]. Basonym:
620 *Haemophilus vaginalis* Gardner and Dukes 1955 (Approved Lists 1980)[9].

621 Earlier homotypic synonym: *Gardnerella vaginalis* (Gardner and Dukes 1955) Greenwood and
622 Pickett 1980[114]. Earlier homotypic synonym: *Gardnerella vaginalis* (Gardner and Dukes 1955)
623 Greenwood and Pickett 1980[111].

Gardnerella vaginalis subsp. *vaginalis* (va.gi.na'lis. L. fem. n. *vagina*, sheath, vagina; L. masc. adj. suff. -*alis,* suffix denoting pertaining to; N.L. fem. adj. *vaginalis*, pertaining to vagina, of the vagina).

The description of this subspecies remains as given to *Gardnerella vaginalis* by Gardner and Dukes 1955[9] and Greenwood and Pickett 1980[111] with the following additional information. Curated genomes of 79 *Gardnerella vaginalis* subsp. *vaginalis* cultivated isolates were used in this analysis, isolated from the vagina, endometrium, cervix, catheter urine, and placenta of humans from Bangladesh, Belgium, China, France, Russia, the United Kingdom, the United States, and Zambia (Supplementary Table 2). The mean genome length of *Gardnerella vaginalis* subsp. *vaginalis* is 1.68 Mb [range1.58 – 1.78] and the median G+C content is 41.26% [range 40.99% – 41.82%]. The type strain for *Gardnerella vaginalis* subsp. *vaginalis* remains 594^{T} of Gardner and Dukes (=ATCC 14018^{T}=DSM 4944^{T}=CCUG 3717^{T}=JCM 11026^{T}=NCTC 10915^{T}=NCTC 10287^{T}).

Strains used in previous classifying publications: ATCC14019, ATCC49145, DNF01149, FDAARGOS296, FDAARGOS568, GH015, GH021, HMP9231, JCM11026, JCP7276, JCP7672, NR001, NR037, NR038, NR039, UGent09.01, UGent09.07, UGent25.49, UMB0032A, UMB0032B, UMB0061, UMB0233, UMB0298, UMB0386, UMB0768, UMB0770, UMB0775, WP023, 287, 377, 653, 1492, 3549624, 0284V, 0286E, 0288E, 1-010-6, 1-014-4, 104-V2-10, 1388E, 18-4, 203-8, 23-12, 302000480-V1-10, 315-A, 7571-2, 772-1, and 825-1.

642 **Description of *Gardnerella vaginalis* subsp. *lacydonensis* (Allini Ntiguemassa *et al.* 2026)**
643 **Bouzek *et al.* comb. nov.**

644 Basonym: *Gardnerella lacydonensis* Allini Ntiguemassa *et al*. 2026[49] Basonym: *Gardnerella*
645 *lacydonensis* Allini Ntiguemassa *et al*. 2026[49]

646 *Gardnerella vaginalis* subsp. *lacydonensis* (la.cy.do.nen'sis. N.L. fem. adj. *lacydonensis*, from
647 Lacydon, the name of the ancient port of Marseille, the French city where the strain was first
648 described).

649 The description of *Gardnerella vaginalis* subsp. *lacydonensis* remains as given to *Gardnerella*
650 *lacydonensis* by Allini Ntiguemassa *et al*.[49] with the following additional information. Our
651 collection of curated genomes contains 15 *Gardnerella vaginalis* subsp. *lacydonensis* cultivated
652 strains, isolated from the vagina, endometrium, and ectocervical mucosa of humans from France,
653 Kenya, South Africa, the United States, and Zambia (Supplementary Table 2). The median genome
654 length of these isolates is 1.66 Mb [range 1.62 – 1.72] and the median G+C content is 41.29%
655 [range 41.11% – 41.42%]. The type strain of *Gardnerella vaginalis* subsp. *lacydonensis* remains
656 Marseille-Q9181^T (=CSUR Q9181^T=CECT 31121^T).

657 **Description of *Gardnerella vaginalis* subsp. *regina* subsp. nov.**

658 *Gardnerella vaginalis* subsp. *regina* (re.gi'na. L. fem. n. *regina*, the queen, referring to its
659 placement in the first validly published species of the *Gardnerella* genus).

660 We analyzed 17 curated genomes from cultivated strains of *Gardnerella vaginalis* subsp. *regina*,
661 isolated from the vagina of humans from Bangladesh, France, the United States, and Zambia
662 (Supplementary Table 2). The median genome length of the *Gardnerella vaginalis* subsp. *regina*
663 is 1.63 Mb [range 1.57 – 1.67], and the median G+C content is 41.41% [range 41.25% – 41.62%].
664 The isolate is a facultative anaerobe and has gram variable cells. Growth develops on Brucella agar
665 in anaerobic conditions at 37°C after one to two days. In a genome comparison of *Gardnerella*
666 *vaginalis* subsp. *regina* strain DNF00622A^{T} with validly published *Gardnerella* type strains and
667 the type strains proposed in this study, the highest ANI values correspond with its sister subspecies
668 *Gardnerella vaginalis* subsp. *vaginalis* ATCC14018^{T} (96.82%) and *Gardnerella vaginalis* subsp.
669 *lacydonensis* Marseille-Q9181^{T} (96.17%), followed by *Gardnerella piotii* subsp. *piotii* UGent
670 18.01^{T} (89.24%) (Figure 2). Within the genome, 83 CAZymes were identified, including α-
671 mannosidase (GH125), β-galactosidase (GH2), β-mannosidase (GH130), Class II α-mannosidase
672 (GH38), β-N-acetylglucosaminidase (GH20), α-L-fucosidase (GH29), and sialidase (GH33)[112].
673 KEGG analysis identified a total of 97 transporters, 13 secretion genes, and 365 enzymes[112].
674 Metabolism predictions using GapMind[103-105] include biosynthesis of cysteine, glutamine, glycine,
675 and proline and catabolism of ethanol, gluconate, and serine (Supplementary Table 6). The type
676 strain for *Gardnerella vaginalis* subsp. *regina* is DNF00622A^{T}, isolated from the vaginal fluid of
677 a woman positive for bacterial vaginosis in Seattle, Washington, USA. Accessions for the 16S
678 rRNA gene sequence and whole genome sequence for DNF00622A^{T} are PZ013896 and CP058395.

679 Strains used in previous classifying publications: 1-004-3, 819-1, and 819-3.

680 **Description of *Gardnerella piotii* subsp. *piotii* (Vaneechoutte *et al*. 2019) Bouzek *et al*. comb.**
681 **nov.**

682 Basonym: *Gardnerella piotii* Vaneechoutte *et al*. 2019[32]

683 *Gardnerella piotii* (pi.o'ti.i. N.L. gen. masc. n. *piotii*, of Piot, named after Peter Piot, a Belgian
684 microbiologist who first attempted to distinguish different groups within *G. vaginalis*).

685 The description of *Gardnerella piotii* subsp. *piotii* remains as given to *Gardnerella piotii* by
686 Vaneechoutte *et al*.[32] with the following additional information. There are 17 genomes of cultivated
687 *Gardnerella piotii* subsp. *piotii* in our curated collection, isolated from the vagina and endometrium
688 of humans from Bangladesh, Belgium, China, France, and the United States (Supplementary Table
689 2). The median genome length of these isolates is 1.54 Mb [range 1.50 – 1.59] and the median
690 G+C content is 42.43% [range 42.22% – 42.61%]. The *Gardnerella piotii* subsp. *piotii* type strain
691 remains UGent 18.01^{T} (= LMG 30818^{T}=CCUG 72427^{T}).

692 **Description of *Gardnerella piotii* subsp. *pickettii* (Sousa *et al*. 2023) Bouzek *et al*. comb. nov.**

693 Basonym: *Gardnerella pickettii* Sousa *et al*. 2023[48]

694 *Gardnerella piotii* subsp. *pickettii* (pic.ket'ti.i. N.L. gen. masc. n. *pickettii*, of Pickett, named after
695 M. John Pickett, an American microbiologist who proposed the transfer of *Haemophilus vaginalis*
696 to the new genus *Gardnerella*).

697 The description of *Gardnerella piotii* subsp. *pickettii* remains as given to *Gardnerella pickettii* by
698 Sousa *et al*.[48] with the following additional information. We included the curated genomes of 40
699 cultivated isolates of *Gardnerella piotii* subsp. *pickettii* in our analysis, isolated from the vagina,
700 endometrium, ectocervical mucosa, urine, bladder and rectum of humans from Bangladesh,
701 Canada, China, Kenya, Portugal, South Africa, the United States, and Zambia (Supplementary
702 Table 2). The median genome length of *Gardnerella piotii* subsp. *pickettii* is 1.56 Mb [range 1.51
703 – 1.68] and the median G+C content is 42.32% [range 42.00% - 42.74%]. The type strain of
704 *Gardnerella piotii* subsp. *pickettii* remains c17Ua_112^T (=DSM 113414^T=CCP 71^T).

705 **Description of *Gardnerella piotii* subsp. *sialicivorans* subsp. nov.**

706 *Gardnerella piotii* subsp. *sialicivorans* (si.a.li.ci.vo'rans. N.L. neut. n. *acidum sialicum*, sialic acid;
707 L. pres. part. *vorans*, devouring; N.L. fem. part. adj. *sialicivorans*, sialic acid devouring).

708 The subspecies *Gardnerella piotii* subsp. *sialicivorans* is named for the intact sialidase catabolism
709 operon and one or more sialidase genes present in the genomes of *Gardnerella piotii*. Our
710 collection of curated genomes includes three *Gardnerella piotii* subsp. *sialicivorans* cultivated
711 strains, all isolated from the vagina of humans from the United States (Supplementary Table 2).
712 The median genome length is 1.51 Mb [range 1.50 – 1.53], and the median G+C content is 42.48%
713 [range 42.39% - 42.48%]. This isolate is a gram-positive coccus and produces pinpoint colonies

714 (≤ 1 mm) on Brucella agar after three days of anaerobic incubation at 37˚C. Colonies are grey,
715 pleomorphic, and alpha-hemolytic. In a genome comparison of *Gardnerella piotii* subsp.
716 *sialicivorans* strain C0013B2^T with validly published *Gardnerella* type strains and the type strains
717 proposed in this study, the highest ANI values correspond with members of its sister subspecies
718 *Gardnerella piotii* subsp. *pickettii* c17Ua_112^T (95.68%) and *Gardnerella piotii* subsp. *piotii*
719 UGent 18.01^T (94.66%), followed by *Gardnerella rara* C0101A1^T (90.88%) (Figure 2). Within the
720 genome, 76 CAZymes were identified, including sialidase (GH33) and Class II α-mannosidase
721 (GH38)[112]. KEGG analysis identified a total of 79 transporters, 13 secretion genes, and 355
722 enzymes[112]. Metabolism predictions using GapMind[103-105] include biosynthesis of glutamine,
723 glycine, and proline, and catabolism of alanine, ethanol, and serine (Supplementary Table 6). The
724 *Gardnerella piotii* subsp. *sialicivorans* type strain is C0013B2^T, isolated from the vagina of an
725 asymptomatic menstruating woman in Birmingham, Alabama, USA. Accession for the 16S rRNA
726 gene sequence for C0013B2^T is PZ016854.

727 Strain used in previous classifying publications: 582.

728 **Description of *Gardnerella rara* sp. nov.**

729 *Gardnerella rara* (ra'ra. L. fem. adj. *rara*, rare, uncommon, referring to its rare genitourinary
730 colonization of humans).

731 Our curated collection includes six genomes of *Gardnerella rara* from isolates cultivated from the
732 vagina and endometrium of humans in Bangladesh, the United States, and Zambia (Supplementary
733 Table 2). The median genome length of *Gardnerella rara* is 1.52 Mb [range 1.47 – 1.58] and the
734 median G+C content is 43.23% [range 43.01% – 43.29%]. This isolate is a gram-positive bacillus
735 and produces small colonies on Brucella agar after two days of anaerobic incubation at 37˚C.
736 Colonies are grey, round, and alpha-hemolytic. In a genome comparison of *Gardnerella rara* strain
737 C0101A1^{T} with validly published *Gardnerella* type strains and the type strains proposed in this
738 study, the highest ANI values correspond with *Gardnerella piotii* subsp. *piotii* UGent 18.01^{T} at
739 91.38% (Figure 2). Within the genome, 71 CAZymes were identified, including β-mannosidase
740 (GH130), Class II α-mannosidase (GH38), and sialidase (GH33)[112]. KEGG analysis identified a
741 total of 78 transporters, 13 secretion genes, and 344 enzymes[112]. Metabolism predictions using
742 GapMind[103-105] include biosynthesis of glutamine, glycine, and proline, and catabolism of alanine
743 and ethanol (Supplementary Table 6). The type strain of *Gardnerella rara* type is C0101A1^{T},
744 isolated from the vagina of a woman positive for *Chlamydia trachomatis* in Baltimore, Maryland,
745 USA. Accession for the 16S rRNA gene sequence for C0101A1^{T} is PZ016855.

746 Strains used in previous classifying publications: 1401E, GED7760B, and HUZ601048-2.

747 **Description of *Gardnerella leopoldii* subsp. *leopoldii* (Vaneechoutte *et al*. 2019) Bouzek *et al*.**
748 **comb. nov.**

749 Basonym: *Gardnerella leopoldii* Vaneechoutte *et al*. 2019[32]

750 *Gardnerella leopoldii* (le.o.pol'di.i. N.L. gen. masc. n. *leopoldii,* of Leopold, named after Sidney
751 Leopold, who described the first strains of the species).

752 The description of this subspecies remains as given to *Gardnerella leopoldii* by Vaneechoutte *et*
753 *al*.[32] with the following additional information. There are 24 genomes of cultivated isolates of
754 *Gardnerella leopoldii* subsp. *leopoldii* in our curated collection, isolated from the vagina and
755 catheter urine of humans from Belgium, Canada, France, Kenya, and the United States
756 (Supplementary Table 2). The median genome length is 1.53 Mb [range 1.37 – 1.67], and the
757 median G+C content is 42.15% [range 42.07% - 42.34%]. The type strain of *Gardnerella leopoldii*
758 subsp. *leopoldii* remains UGent 06.41^{T} (=LMG 30814^{T}=CCUG 72425^{T}).

759 **Description of *Gardnerella leopoldii* subsp. *swidsinskii* (Vaneechoutte *et al*. 2019) Bouzek *et***
760 ***al*. comb. nov.**

761 Basonym: *Gardnerella swidsinskii* Vaneechoutte *et al*. 2019[32]

762 *Gardnerella swidsinskii* (swid.sins'ki.i. N.L. gen. masc. n. *swidsinskii*, of Swidsinski, named after
763 Alexander Swidsinski, a German microbiologist, who recognized that clue cells in Gram stains of
764 vaginal swabs of women with BV are indicative for biofilm formation by *Gardnerella*).

765 The description of this subspecies remains as provided by Vaneechoutte *et al*. for *Gardnerella*
766 *swidsinskii*[32] with the following additional information. We included 46 genomes of cultivated
767 *Gardnerella leopoldii* subsp. *swidsinskii* in our curated collection, isolated from the vagina,
768 endometrium, rectum, and catheter urine of humans from China, France, Russia, and the United
769 States (Supplementary Table 2). The median genome length is 1.63 Mb [range 1.52 – 1.72], and
770 the median G+C content is 41.93% [range 41.64% - 42.29%]. The type strain of *Gardnerella*
771 *leopoldii* subsp. *swidsinskii* remains GS 9838-1^T (=LMG 30812^T=CCUG 72429^T).

772 **Description of *Gardnerella leopoldii* subsp. *aggregata* subsp. nov.**

773 *Gardnerella leopoldii* subsp. *aggregata* (ag.gre.ga'ta. L. fem. part. adj. *aggregata*, joined together,
774 referring to the presence of genes related to cell aggregation).

775 The genomes of *Gardnerella leopoldii* subsp. *aggregata*, as well as those of the overarching
776 species, differ from other *Gardnerell*a species in the presence of genes consistent with adherence
777 and biofilm formation. There are five genomes of cultivated *Gardnerella leopoldii* subsp.
778 *aggregata* in our curated collection, isolated from the vagina and ectocervical mucosa of humans
779 from Bangladesh, South Africa the United States, and Zambia (Supplementary Table 2). The
780 median genome length is 1.48 Mb [range 1.48 – 1.53], and the median G+C content is 42.19%
781 [range 42.11% - 42.50%]. This isolate is a gram variable coccobacillus and produces pinpoint
782 colonies (≤ 1 mm) on Brucella agar after three days of anaerobic incubation at 37˚C. Colonies are
783 glossy white, round, convex, and alpha-hemolytic. In a genome comparison of *Gardnerella*

784 *leopoldii* subsp. *aggregata* strain C0026F5^T with validly published *Gardnerella* type strains and
785 the type strains proposed in this study, the highest ANI values correspond with its sister subspecies
786 *Gardnerella leopoldii* subsp. *phocaeensis* Marseille-Q9179^T (96.09%), *Gardnerella leopoldii*
787 subsp. *leopoldii* UGent 06.41^T (95.58%), and *Gardnerella leopoldii* subsp. *swidsinskii* GS 9838-
788 1^T (94.70%), followed by *Gardnerella greenwoodii* subsp. *bretellae* Marseille-QA0894^T (87.40%)
789 (Figure 2). Within the genome, 74 CAZymes were identified, including α-glucosidase (GH65),
790 endo-β-galactosidase (GH16), and β-glucosidase (GH3)[112]. KEGG analysis identified a total of 84
791 transporters, 14 secretion genes, and 353 enzymes[112]. Metabolism predictions using GapMind[103-
792 105] include biosynthesis of glutamine, glycine, methionine, proline, and threonine, and catabolism
793 of ethanol (Supplementary Table 6). The type strain of *Gardnerella leopoldii* subsp. *aggregata* is
794 C0026F5^T, isolated from the vagina of an asymptomatic menstruating woman from Birmingham,
795 Alabama, USA. Accession for the 16S rRNA gene sequence for C0026F5^T is PZ016853.

796 **Description of *Gardnerella leopoldii* subsp. *phocaeensis* (Allini Ntiguemassa *et al*. 2026)**
797 **Bouzek *et al*. comb. nov.**

798 Basonym: *Gardnerella phocaeensis* Allini Ntiguemassa *et al*. 2026[49]

799 *Gardnerella phocaeensis* (pho.cae.en'sis. L. fem. adj. *phocaeensis*, referring to Phocaea, the name
800 of the Ionian Greek city where the founders of Marseille came from. The strain was isolated in
801 Marseille).

802 The description of *Gardnerella leopoldii* subsp. *phocaeensis* remains as given to *Gardnerella*
803 *phocaeensis* by Allini Ntiguemassa *et al*.[49] with the following additional information. There are
804 eleven genomes of cultivated *Gardnerella leopoldii* subsp. *phocaeensis* in our curated collection,
805 isolated from the vagina and endometrium of humans from Bangladesh, France, the United States,
806 and Zambia (Supplementary Table 2). The median genome length of *Gardnerella leopoldii* subsp.
807 *phocaeensis* is 1.54 Mb [range 1.52 – 1.66] and the median G+C content is 42.26% [range 42.03%
808 - 42.43%]. The type strain of *Gardnerella leopoldii* subsp. *phocaeensis* remains Marseille-Q9179^T
809 (=CSUR Q9179^T=CECT 31120^T).

810 **Description of *Gardnerella greenwoodii* subsp. *greenwoodii* (Sousa *et al*. 2023) Bouzek *et al.***
811 **comb. nov.**

812 Basonym: *Gardnerella greenwoodii* Sousa *et al.* 2023[48]

813 *Gardnerella greenwoodii* (green.woo'di.i. N.L. gen. masc. n. *greenwoodii*, of Greenwood, named
814 after James Greenwood, an American microbiologist who proposed the transfer of *Haemophilus*
815 *vaginalis* to the new genus *Gardnerella*).

816 The description of this subspecies remains as given to *Gardnerella greenwoodii* by Sousa *et al*.[48]
817 with the following additional information. We included 13 genomes of *Gardnerella greenwoodii*
818 subsp. *greenwoodii* in our curated collection, cultivated from the vagina, endometrium, and urine
819 of humans from France, Portugal, the United Kingdom, and the United States (Supplementary

820 Table 2). The median genome length of *Gardnerella greenwoodii* subsp. *greenwoodii* is 1.54 Mb
821 [range 1.49 – 1.65] and the median G+C content is 43.40% [range 42.90% - 43.44%]. The type
822 strain of *Gardnerella greenwoodii* subsp. *greenwoodii* remains c31Ua_26^T (=DSM 113415^T=CCP
823 72^T).

824 **Description of *Gardnerella greenwoodii* subsp. *intermedia* subsp. nov.**

825 *Gardnerella greenwoodii* subsp. *intermedia* (in.ter.me'di.a. L. fem. adj. *intermedia*, intermediate,
826 between, referencing the shared possession of operons with both *Gardnerella* subgenera).

827 Pangenome analysis revealed that *Gardnerella greenwoodii* subsp. *intermedia*, as well as other
828 members of the species, shares biologically relevant operons with the Set A subgenus (*Gardnerella*
829 *vaginalis*, *Gardnerella piotii* and *Gardnerella rara*) while also falling within the Set B subgenus,
830 suggesting that the species is an intermediate between Set A and Set B. Our curated set of genomes
831 contains two strains of *Gardnerella greenwoodii* subsp. *intermedia*, both cultivated from the vagina
832 of humans from the United States (Supplementary Table 2). The lengths of these genomes are 1.53
833 and 1.55 Mb, and the G+C contents are 43.04% and 43.05%. Cells are gram variable. Optimal
834 growth occurs within two to three days on Brucella agar in an anerobic atmosphere at 37°C but
835 can occur in aerobic environments. In a genome comparison of *Gardnerella greenwoodii* subsp.
836 *intermedia* strain KA00603^T with validly published *Gardnerella* type strains and the type strains
837 proposed in this study, the highest ANI values correspond with sister subspecies *Gardnerella*
838 *greenwoodii* subsp. *bretellae* Marseille-QA0894^T (95.76%) and *Gardnerella greenwoodii* subsp.

839 *greenwoodii* c31Ua_26^{T} (95.04%), followed by *Gardnerella kenyensis* NR047^{T} (88.32%) (Figure
840 2).

841 Within the genome, 66 CAZymes were identified, including β-glucosidase (GH3), and KEGG
842 analysis identified a total of 85 transporters, 13 secretion genes, and 347 enzymes[112]. Metabolism
843 predictions using GapMind[103-105] include biosynthesis of cysteine, glutamine, glycine, methionine,
844 proline, and threonine, and catabolism of ethanol and serine (Supplementary Table 6). The type
845 strain of *Gardnerella greenwoodii* subsp. *intermedia* is KA00603^{T,} isolated from the vaginal fluid
846 of a woman positive for bacterial vaginosis in Seattle, WA, USA. Accessions for the 16S rRNA
847 gene sequence and whole genome sequence for KA00603^{T} are PZ013893 and CP058371.

848 Strain used in previous classifying publications: 1500E.

849 **Description of *Gardnerella greenwoodii* subsp. *bretellae* (Allini Ntiguemassa *et al*. 2026)**
850 **Bouzek *et al*. comb. nov.**

851 Basonym: *Gardnerella bretellae* Allini Ntiguemassa *et al*. 2026

852 *Gardnerella bretellae* (bre.tel'lae. N.L. gen. fem. n. *bretellae*, honoring Florence Bretelle for her
853 contribution to the description of vaginal microbiota).

854 The description of *Gardnerella greenwoodii* subsp. *bretellae* remains as given to *Gardnerella*
855 *bretellae* by Allini Ntiguemassa *et al*.[49] with the following additional information. Our curated

856 collection of genomes contains eleven strains of *Gardnerella greenwoodii* subsp. *bretellae*,
857 cultivated from the vagina and ectocervical mucosa of humans from Canada, France, Kenya, South
858 Africa, the United States, and Zambia (Supplementary Table 2). The median genome length is 1.52
859 Mb [range 1.50 – 1.64], and the median G+C content is 43.22% [range 43.05% - 43.50%]. The
860 *Gardnerella greenwoodii* subsp. *bretellae* type strain remains Marseille-QA0894^T (=CSUR
861 QA0894^T=CECT 31122^T).

862 **Description of *Gardnerella washingtoni* sp. nov.**

863 *Gardnerella washingtoni* (wa.shing.to'ni. N.L. gen. n. *washingtoni*, named in honor of Washington
864 University in St. Louis, MO, USA, where the first two genomes of this species originated, and the
865 University of Washington in Seattle, WA, USA, which is associated with the first complete genome
866 of this species).

867 There are five genomes of *Gardnerella washingtoni* in our curated collection. These isolates have
868 been cultivated from the vagina and urine of humans in the United States (Supplementary Table
869 2). The median genome length of *Gardnerella washingtoni* is 1.60 Mb [range 1.56 – 1.65] and the
870 median G+C content is 42.96% [range 42.89% – 43.11%]. Cells are Gram variable. The isolate is
871 a facultative anaerobe that grows on Brucella agar in anaerobic conditions at 37°C in one to two
872 days. In a genome comparison of *Gardnerella washingtoni* strain DNF00536^T with validly
873 published *Gardnerella* type strains and the type strains proposed in this study, *Gardnerella*
874 *greenwoodii* subsp. *bretellae* Marseille-QA0894^T has the highest ANI value (87.40%) (Figure 2).

875 Within the genome, 69 CAZymes were identified, including endo-β-galactosidase (GH16), α-
876 glucosidase (GH65), and β-glucosidase (GH3)[112]. KEGG analysis identified a total of 83
877 transporters, 13 secretion genes, and 348 enzymes[112]. Metabolism predictions using GapMind[103-
878 105] include biosynthesis of cysteine, glutamine, glycine, methionine, and threonine, and catabolism
879 of ethanol (Supplementary Table 6). The type strain of *Gardnerella washingtoni* is DNF00536^T,
880 isolated from the vaginal fluid of a woman positive for bacterial vaginosis in Seattle, WA, USA.
881 Accessions for the 16S rRNA gene sequence and whole genome sequence for DNF00536^T are
882 PZ013897 and CP058398.

883 Strains used in previous classifying publications: JCP8481A, JCP8481B, and PSS7772B.

884 **Description of *Gardnerella bivia* sp. nov.**

885 *Gardnerella bivia* (bi'vi.a. L. fem. adj. *bivia*, having two ways, in reference to the genome
886 containing genes for two major pathways for mucin degradation).

887 *Gardnerella bivia* has genes associated with two major pathways for mucin degradation. There are
888 six isolates of this species in our curated collection, cultivated from the vagina of humans in
889 Bangladesh, the United States, and Zambia (Supplementary Table 2). The median genome length
890 of *Gardnerella bivia* is 1.59 Mb [range 1.55 – 1.64] and the median G+C content is 38.08% [range
891 37.90% – 38.31%]. The isolate is a gram variable obligate anaerobe. Growth occurs on Brucella
892 agar within one to two days in anaerobic conditions at 37°C. In a genome comparison of

893 *Gardnerella bivia* strain DNF00502^T with validly published *Gardnerella* type strains and the type
894 strains proposed in this study, *Gardnerella hutchinsoni* KA00225^T has the highest ANI value
895 (82.91%) (Figure 2). Within the genome, 75 CAZymes were identified, including β-galactoside
896 phosphorylase (GH112), α-L-fucosidase (GH29), and β-glucosidase (GH3)[112]. KEGG analysis
897 identified a total of 74 transporters, 13 secretion genes, and 349 enzymes[112]. Metabolism
898 predictions using GapMind[103-105] include biosynthesis of cysteine, glutamine, glycine, proline, and
899 threonine, and catabolism of acetate, ethanol, and serine (Supplementary Table 6). The type strain
900 of *Gardnerella bivia* is DNF00502^T, isolated from the vaginal fluid of a woman positive for
901 bacterial vaginosis in Seattle, WA, USA. Accessions for the 16S rRNA gene sequence and whole
902 genome sequence for DNF00502^T are PZ013890 and CP058399.

903 Strain used in previous classifying publications: CMW7778B.

904 **Description of *Gardnerella hutchinsoni* sp. nov.**

905 *Gardnerella hutchinsoni* (hutch.in.so'ni. N.L. gen. n. *hutchinsoni*, named in honor of Fred
906 Hutchinson Cancer Center, where the isolate was first cultivated and the first complete genomes
907 of this subspecies were sequenced).

908 We included six genomes of *Gardnerella hutchinsoni* isolates in our curated collection, all
909 cultivated from the vagina of humans in the United States and Zambia (Supplementary Table 2).
910 The median genome length of *Gardnerella hutchinsoni* is 1.67 Mb [range 1.64 – 1.72] and the

911 median G+C content is 40.64% [range 40.53% – 40.76%]. Cells are gram variable. Optimal growth
912 occurs within one to two days on Brucella agar in an anerobic atmosphere at 37°C but can occur
913 in aerobic environments. In a genome comparison of *Gardnerella hutchinsoni* strain KA00225^T
914 with validly published *Gardnerella* type strains and the type strains proposed in this study,
915 *Gardnerella leopoldii* subsp. *aggregata* C0026F5^T has the highest ANI value (86.69%) (Figure 2).
916 Within the genome, 69 CAZymes were identified, including β-galactoside phosphorylase (GH112)
917 and β-glucosidase (GH3)[112]. KEGG analysis identified a total of 70 transporters, 13 secretion
918 genes, and 341 enzymes[112]. Metabolism predictions using GapMind[103-105] include biosynthesis of
919 asparagine, glutamine, glycine, methionine, and proline, and catabolism of ethanol (Supplementary
920 Table 6). The type strain of *Gardnerella hutchinsoni* is KA00225^T, isolated from the vaginal fluid
921 of a woman positive for bacterial vaginosis in Seattle, WA, USA. Accessions for the 16S rRNA
922 gene sequence and whole genome sequence for KA00225^T are PZ013891 and CP058376.

923 Strains used in previous classifying publications: 3-005-3V and HUZ601990-0.

924 **Description of Candidatus *Gardnerella kenyensis* sp. nov.**

925 *Gardnerella kenyensis* (ken.yen'sis. N.L. fem. adj. *kenyensis*, from Kenya, referring to the country
926 of isolation of the type strain).

927 All known strains of Candidatus *Gardnerella kenyensis* to date (NR002, NR021, and NR047) have
928 been isolated from the vaginal fluid of a human in Nairobi, Kenya (Supplementary Table 2). Two

929 of these are genomic duplicates (NR002 and NR047) and one (NR021) did not pass our genome
930 quality threshold metrics. As such, only NR047^{T} was included in our curated collection
931 (Supplementary Table 2). The genome length of this isolate is 1.66 Mb, and its G+C content is
932 43.62%. In a genome comparison of Candidatus *Gardnerella kenyensis* strain NR047T with validly
933 published *Gardnerella* type strains and the type strains proposed in this study, the highest ANI
934 values correspond with *Gardnerella insolita* subsp. *libera* C0085C4^{T} (90.59%) (Figure 2). Within
935 the genome, 66 CAZymes were identified, including sialidase (GH33), and KEGG analysis
936 identified a total of 84 transporters, 13 secretion genes, and 340 enzyme[112]. Metabolism predictions
937 using GapMind[103-105] include biosynthesis of asparagine, cysteine, glutamine, glycine, methionine,
938 proline, and threonine, and catabolism of ethanol (Supplementary Table 6). The type strain of
939 Candidatus *Gardnerella kenyensis* is NR047^{T} and the accession for the whole genome sequence is
940 NQOH00000000.2.

941 **Description of *Gardnerella insolita* sp. nov.**

942 *Gardnerella insolita* (in.so'li.ta. L. fem. adj. *insolita*, unusual, strange, uncommon, referring to the
943 low number of species cultivated).

944 To date, only three strains of *Gardnerella insolita* have been cultivated, all isolated from the vagina
945 of humans in the United States and Zambia (Supplementary Table 2). The median genome length
946 of *Gardnerella insolita* is 1.61 Mb [range 1.59 – 1.66] and the median G+C content is 43.60%
947 [range 41.27% – 43.80%]. The *Gardnerella insolita* type strain is C0396A3^{T}, isolated from the

948 vagina of a pregnant woman at 19 weeks of gestation and who delivered preterm in Zambia.
949 Accession for the 16S rRNA gene sequence C0396A3^{T} is PZ236487.

950 **Description of *Gardnerella insolita* subsp. *insolita* comb. nov.**

951 Basonym: *Gardnerella insolita* Bouzek *et al*. 2026

952 *Gardnerella insolita* (in.so'li.ta. L. fem. adj. *insolita*, unusual, strange, uncommon, referring to the
953 low number of species cultivated).

954 There are two strains of *Gardnerella insolita* subsp. *insolita* in our curated genome collection. The
955 genome lengths of these isolates are 1.66 and 1.61 Mb, and the G+C contents are 41.27% and
956 43.80%. This isolate is a gram-positive bacillus and produces pinpoint (≤ 1 mm) colonies on
957 Brucella agar after two days of anaerobic incubation at 37˚C. Colonies are round, clear, and glossy.
958 In a genome comparison of *Gardnerella insolita* subsp. *insolita* strain C0396A3^{T} with validly
959 published *Gardnerella* type strains and the type strains proposed in this study, the highest ANI
960 values correspond with *Gardnerella insolita* subsp. *libera* C0085C4^{T} (95.24%), followed by
961 *Gardnerella kenyensis* NR047^{T} (89.89%) (Figure 2). Within the genome, 66 CAZymes were
962 identified, including β-glucosidase (GH3)[112]. KEGG analysis identified a total of 85 transporters,
963 13 secretion genes, and 347 enzymes[112]. Metabolism predictions using GapMind[103-105] include
964 biosynthesis of asparagine, glutamine, glycine, proline, and threonine, and catabolism of ethanol
965 (Supplementary Table 6).

966 Strain used in previous classifying publications: 1-004-1.

967 **Description of *Gardnerella insolita* subsp. *libera* subsp. nov.**

968 *Gardnerella insolita* subsp. *libera* (li'be.ra. L. fem. adj. *libera*, independent, free, referring to the
969 rarity of the subspecies).

970 To date, only one strain of *Gardnerella insolita* subsp. *libera* has been cultivated, C0085C4^T, from
971 the vagina of a woman positive for *Chlamydia trachomatis* in Baltimore, Maryland, USA. Its
972 genome length is 1.59 Mb, and its G+C content is 43.60%. This isolate is a gram variable
973 coccobacillus and produces pinpoint colonies (≤1 mm) on Brucella agar after two days of anaerobic
974 growth at 37˚C. Colonies are grey, glistening, round, and β-hemolytic. In a genome comparison of
975 *Gardnerella insolita* subsp. *libera* strain C0085C4^T with validly published Gardnerella type strains
976 and the type strains proposed in this study, the highest ANI values correspond with *Gardnerella*
977 *insolita* subsp. *insolita* C0396A3^T (95.30%), followed by Candidatus *Gardnerella kenyensis*
978 NR047^T (90.82%) (Figure 2). Within the genome, 70 CAZymes were identified, including Class
979 II α-mannosidase and α-glucosidase[112]. KEGG analysis identified a total of 84 transporters, 12
980 secretion genes, and 342 enzymes[112]. Metabolism predictions using GapMind[103-105] include
981 biosynthesis of cysteine, glutamine, glycine, proline, and methionine, and catabolism of ethanol
982 and serine (Supplementary Table 6). The type strain of *Gardnerella insolita* subsp. *libera* is
983 C0085C4^T. Accession for the 16S rRNA gene sequence for C0085C4^T is PZ016852.

984 **Acknowledgements**

985 Research reported in this publication was supported in part by the National Human Genome
986 Research Institute as part of the Human Microbiome Project Initiative focused on technology
987 development under award number R01HG005816 (D.N.F.), the National Institute for Allergy and
988 Infectious Diseases of the National Institutes of Health under award number U19AI084044 (J.R.),
989 the National Institute of Dental and Craniofacial Research (NIDCR) of the National Institutes of
990 Health Office of the Director (OD) Common Fund under award number R01DE027850 (C.D.J).
991 and the Gates Foundation under award INV010470 to (J.R.). In addition, startup funds were
992 provided by Fred Hutch Cancer Center and MD Anderson Cancer Center (C.D.J.).

993 **Contributions**

994 H.K.B., S.S., C.D.J., and D.N.F. conceived the study. E.F.M., D.S.J., S.M.S., T.F.L., E.M.L., and
995 M.F. conducted experiments including strain isolation, DNA extraction, whole-genome
996 sequencing, cloning, and testing of the shuttle vector. H.K.B. and M.Z.R. conducted all
997 bioinformatics and curated data. S.S., C.D.J, J.R., and D.N.F. supervised and obtained funding for
998 the project. H.K.B., M.Z.R., S.S., S.M.S., C.D.J, and D.N.F. wrote the original draft. J.R. and M.K.
999 provided critical edits to the manuscript and assisted in nomenclature. M.K. provided crucial
1000 nomenclature feedback and provided edits to the protologues and manuscript. All authors were
1001 involved in reviewing and editing this submission.

1002 **Data availability**

1003 NCBI Accession numbers of genomes used are available in Supplementary Table 1.

1004 **Declaration of interests**

1005 C.D.J is an inventor on intellectual property related to SyngenicDNA/SyMPL technology licensed
1006 by Fred Hutchinson Cancer Center to Azitra Inc., which received compensation related to this
1007 license, and C.D.J serves as a consultant to Azitra Inc. J.R. is the cofounder of LUCA Biologics.
1008 Azitra Inc. and LUCA Biologics had no role in this study. The remaining authors declare no
1009 competing interests.

1010 **References**

1011 1 Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. Molecular Identification of Bacteria
1012 Associated with Bacterial Vaginosis. *New England Journal of Medicine* **353**, 1899-1911
1013 (2005). https://doi.org/10.1056/nejmoa043802
1014 2 Amsel, R. *et al.* Nonspecific vaginitis. Diagnostic criteria and microbial and
1015 epidemiologic associations. *Am J Med* **74**, 14-22 (1983). https://doi.org/10.1016/0002-
1016 9343(83)91112-9
1017 3 Harwich, M. D. *et al.* Drawing the line between commensal and pathogenic Gardnerella
1018 vaginalis through genome analysis and virulence studies. *BMC Genomics* **11**, 375 (2010).
1019 https://doi.org/10.1186/1471-2164-11-375
1020 4 Wiesenfeld, H. C., Hillier, S. L., Krohn, M. A., Landers, D. V. & Sweet, R. L. Bacterial
1021 Vaginosis Is a Strong Predictor ofNeisseria
1022 gonorrhoeaeandChlamydiatrachomatisInfection. *Clinical Infectious*
1023 *Diseases* **36**, 663-668 (2003). https://doi.org/10.1086/367658
1024 5 Atashili, J., Poole, C., Ndumbe, P. M., Adimora, A. A. & Smith, J. S. Bacterial vaginosis
1025 and HIV acquisition: a meta-analysis of published studies. *AIDS* **22** (2008).
1026 6 Hillier, S. L. *et al.* Association between bacterial vaginosis and preterm delivery of a low-
1027 birth-weight infant. *New England journal of medicine* **333**, 1737-1742 (1995).
1028 7 Klebanoff, M. A. & Brotman, R. M. Treatment of bacterial vaginosis to prevent preterm
1029 birth. *Lancet* **392**, 2141-2142 (2018). https://doi.org/10.1016/s0140-6736(18)32115-9
1030 8 Berman, H. L., Goltsman, D. S. A., Anderson, M., Relman, D. A. & Callahan, B. J.
1031 Gardnerella diversity and ecology in pregnancy and preterm birth. *mSystems* **9**, e0133923
1032 (2024). https://doi.org/10.1128/msystems.01339-23
1033 9 Gardner, H. L. & Dukes, C. D. Haemophilus vaginalis vaginitis: a newly defined specific
1034 infection previously classified non-specific vaginitis. *Am J Obstet Gynecol* **69**, 962-976
1035 (1955).
1036 10 Schwebke, J. R., Muzny, C. A. & Josey, W. E. Role of Gardnerella vaginalis in the
1037 Pathogenesis of Bacterial Vaginosis: A Conceptual Model. *The Journal of Infectious*
1038 *Diseases* **210**, 338-343 (2014). https://doi.org/10.1093/infdis/jiu089

11 Muzny, C. A. *et al.* An Updated Conceptual Model on the Pathogenesis of Bacterial Vaginosis. *The Journal of Infectious Diseases* **220**, 1399-1405 (2019). https://doi.org/10.1093/infdis/jiz342

12 Gelber, S. E., Aguilar, J. L., Lewis, K. L. T. & Ratner, A. J. Functional and Phylogenetic Characterization of Vaginolysin, the Human-Specific Cytolysin fromGardnerella vaginalis. *Journal of Bacteriology* **190**, 3896-3903 (2008). https://doi.org/10.1128/jb.01965-07

13 Patterson, J. L., Stull-Lane, A., Girerd, P. H. & Jefferson, K. K. Analysis of adherence, biofilm formation and cytotoxicity suggests a greater virulence potential of Gardnerella vaginalis relative to other bacterial-vaginosis-associated anaerobes. *Microbiology* **156**, 392-399 (2010). https://doi.org/10.1099/mic.0.034280-0

14 Yeoman, C. J. *et al.* Comparative Genomics of Gardnerella vaginalis Strains Reveals Substantial Differences in Metabolic and Virulence Potential. *PLoS ONE* **5**, e12411 (2010). https://doi.org/10.1371/journal.pone.0012411

15 Alves, P., Castro, J., Sousa, C., Cereija, T. B. & Cerca, N. Gardnerella vaginalis Outcompetes 29 Other Bacterial Species Isolated From Patients With Bacterial Vaginosis, Using in an In Vitro Biofilm Formation Model. *The Journal of Infectious Diseases* **210**, 593-596 (2014). https://doi.org/10.1093/infdis/jiu131

16 Castro, J. *et al.* Using an in-vitro biofilm model to assess the virulence potential of Bacterial Vaginosis or non-Bacterial Vaginosis Gardnerella vaginalis isolates. *Scientific Reports* **5**, 11640 (2015). https://doi.org/10.1038/srep11640

17 Castro, J. *et al.* Comparative transcriptomic analysis of Gardnerella vaginalis biofilms vs. planktonic cultures using RNA-seq. *NPJ Biofilms Microbiomes* **3**, 3 (2017). https://doi.org/10.1038/s41522-017-0012-7

18 Janulaitiene, M. *et al.* Phenotypic characterization of Gardnerella vaginalis subgroups suggests differences in their virulence potential. *PLoS One* **13**, e0200625 (2018). https://doi.org/10.1371/journal.pone.0200625

19 Shvartsman, E., Hill, J. E., Sandstrom, P. & Macdonald, K. S. Gardnerella Revisited: Species Heterogeneity, Virulence Factors, Mucosal Immune Responses, and Contributions to Bacterial Vaginosis. *Infection and Immunity* **91** (2023). https://doi.org/10.1128/iai.00390-22

20 Machado, A. & Cerca, N. Influence of Biofilm Formation byGardnerella vaginalisand Other Anaerobes on Bacterial Vaginosis. *Journal of Infectious Diseases* **212**, 1856-1861 (2015). https://doi.org/10.1093/infdis/jiv338

21 Janulaitiene, M. *et al.* Prevalence and distribution of Gardnerella vaginalis subgroups in women with and without bacterial vaginosis. *BMC Infectious Diseases* **17** (2017). https://doi.org/10.1186/s12879-017-2501-y

22 Munch, M. M. *et al.* Gardnerella Species and Their Association With Bacterial Vaginosis. *J Infect Dis* **230**, e171-e181 (2024). https://doi.org/10.1093/infdis/jiae026

23 Khan, S., Vancuren, S. J. & Hill, J. E. A Generalist Lifestyle Allows Rare Gardnerella spp. to Persist at Low Levels in the Vaginal Microbiome. *Microbial Ecology* **82**, 1048-1060 (2021). https://doi.org/10.1007/s00248-020-01643-1

24 Castro, J., Machado, D. & Cerca, N. Unveiling the role of Gardnerella vaginalis in polymicrobial Bacterial Vaginosis biofilms: the impact of other vaginal pathogens living as neighbors. *ISME J* **13**, 1306-1317 (2019). https://doi.org/10.1038/s41396-018-0337-0

25 Balashov, S. V., Mordechai, E., Adelson, M. E. & Gygax, S. E. Identification, quantification and subtyping of Gardnerella vaginalis in noncultured clinical vaginal samples by quantitative PCR. *Journal of Medical Microbiology* **63**, 162-175 (2014). https://doi.org/https://doi.org/10.1099/jmm.0.066407-0

26 Piot, P. *et al.* Biotypes of Gardnerella vaginalis. *Journal of Clinical Microbiology* **20**, 677-679 (1984). https://doi.org/10.1128/jcm.20.4.677-679.1984

27 Benito, R., Vazquez, J. A., Berron, S., Fenoll, A. & Saez-Nieto, J. A. A Modified Scheme for Biotyping Gardnerella Vaginalis. *Journal of Medical Microbiology* **21**, 357-359 (1986). https://doi.org/https://doi.org/10.1099/00222615-21-4-357

28 Nath, K., Devlin, D. & Beddoe, A. M. Heterogeneity in restriction patterns of Gardnerella vaginalis isolates from individuals with bacterial vaginosis. *Research in Microbiology* **143**, 199-209 (1992). https://doi.org/https://doi.org/10.1016/0923-2508(92)90009-D

29 Ingianni, A., Petruzzelli, S., Morandotti, G. & Pompei, R. Genotypic differentiation ofGardnerella vaginalisby amplified ribosomal DNA restriction analysis (ARDRA). *FEMS Immunology & Medical Microbiology* **18**, 61-66 (1997). https://doi.org/10.1111/j.1574-695x.1997.tb01028.x

30 Ahmed, A. *et al.* Comparative Genomic Analyses of 17 Clinical Isolates of Gardnerella vaginalis Provide Evidence of Multiple Genetically Isolated Clades Consistent with Subspeciation into Genovars. *Journal of Bacteriology* **194**, 3922-3937 (2012). https://doi.org/10.1128/jb.00056-12

31 Cornejo, O. E., Hickey, R. J., Suzuki, H. & Forney, L. J. Focusing the diversity of Gardnerella vaginalis through the lens of ecotypes. *Evolutionary Applications* **11**, 312-324 (2018). https://doi.org/10.1111/eva.12555

32 Vaneechoutte, M. *et al.* Emended description of Gardnerella vaginalis and description of Gardnerella leopoldii sp. nov., Gardnerella piotii sp. nov. and Gardnerella swidsinskii sp. nov., with delineation of 13 genomic species within the genus Gardnerella. *Int J Syst Evol Microbiol* **69**, 679-687 (2019). https://doi.org/10.1099/ijsem.0.003200

33 Potter, R. F., Burnham, C. D. & Dantas, G. In Silico Analysis of Gardnerella Genomospecies Detected in the Setting of Bacterial Vaginosis. *Clin Chem* **65**, 1375-1387 (2019). https://doi.org/10.1373/clinchem.2019.305474

34 Innamorati, K. A. *et al.* Metronidazole response profiles of Gardnerella species are congruent with phylogenetic and comparative genomic analyses. *Genome Medicine* **17** (2025). https://doi.org/10.1186/s13073-025-01446-4

35 Tarracchini, C. *et al.* Assessing the Genomic Variability of Gardnerella vaginalis through Comparative Genomic Analyses: Evolutionary and Ecological Implications. *Applied and Environmental Microbiology* **87** (2020). https://doi.org/10.1128/aem.02188-20

36 Links, M. G., Dumonceaux, T. J., Hemmingsen, S. M. & Hill, J. E. The Chaperonin-60 Universal Target Is a Barcode for Bacteria That Enables De Novo Assembly of Metagenomic Sequence Data. *PLoS ONE* **7**, e49755 (2012). https://doi.org/10.1371/journal.pone.0049755

37 Jayaprakash, T. P., Schellenberg, J. J. & Hill, J. E. Resolution and Characterization of Distinct cpn60-Based Subgroups of Gardnerella vaginalis in the Vaginal Microbiota. *PLoS ONE* **7**, e43009 (2012). https://doi.org/10.1371/journal.pone.0043009

38 Zhang, G., Gao, B., Adeolu, M., Khadka, B. & Gupta, R. S. Phylogenomic Analyses and Comparative Studies on Genomes of the Bifidobacteriales: Identification of Molecular

Signatures Specific for the Order Bifidobacteriales and Its Different Subclades. *Frontiers in Microbiology* **7** (2016). https://doi.org/10.3389/fmicb.2016.00978

39 Lugli, G. A. *et al.* Comparative genomic and phylogenomic analyses of the Bifidobacteriaceae family. *BMC Genomics* **18** (2017). https://doi.org/10.1186/s12864-017-3955-4

40 Huys, G. *et al.* Alloscardovia omnicolens gen. nov., sp. nov., from human clinical samples. *International Journal of Systematic and Evolutionary Microbiology* **57**, 1442-1446 (2007). https://doi.org/https://doi.org/10.1099/ijs.0.64812-0

41 García-Aljaro, C. *et al.* Neoscardovia arbecensis gen. nov., sp. nov., isolated from porcine slurries. *Systematic and Applied Microbiology* **35**, 374-379 (2012). https://doi.org/https://doi.org/10.1016/j.syapm.2012.06.007

42 Simpson, P. J., Ross, R. P., Fitzgerald, G. F. & Stanton, C. Bifidobacterium psychraerophilum sp. nov. and Aeriscardovia aeriphila gen. nov., sp. nov., isolated from a porcine caecum. *International Journal of Systematic and Evolutionary Microbiology* **54**, 401-406 (2004). https://doi.org/https://doi.org/10.1099/ijs.0.02667-0

43 Jian, W. & Dong, X. Transfer of Bifidobacterium inopinatum and Bifidobacterium denticolens to Scardovia inopinata gen. nov., comb. nov., and Parascardovia denticolens gen. nov., comb. nov., respectively. *International Journal of Systematic and Evolutionary Microbiology* **52**, 809-812 (2002). https://doi.org/https://doi.org/10.1099/00207713-52-3-809

44 Gerhardt, K. *et al.* FastAAI: efficient estimation of genome average amino acid identity and phylum-level relationships using tetramers of universal proteins. *Nucleic Acids Research* **53** (2025). https://doi.org/10.1093/nar/gkaf348

45 Schellenberg, J. J. *et al.* Gardnerella vaginalis Subgroups Defined by cpn60 Sequencing and Sialidase Activity in Isolates from Canada, Belgium and Kenya. *PLOS ONE* **11**, e0146510 (2016). https://doi.org/10.1371/journal.pone.0146510

46 Bohr, L. L., Mortimer, T. D. & Pepperell, C. S. Lateral Gene Transfer Shapes Diversity of Gardnerella spp. *Front Cell Infect Microbiol* **10**, 293 (2020). https://doi.org/10.3389/fcimb.2020.00293

47 Hill, J. E. & Albert, A. Y. K. Resolution and Cooccurrence Patterns of Gardnerella leopoldii, G. swidsinskii, G. piotii, and G. vaginalis within the Vaginal Microbiome. *Infection and Immunity* **87** (2019). https://doi.org/10.1128/iai.00532-19

48 Sousa, M. *et al.* Gardnerella pickettii sp. nov. (formerly Gardnerella genomic species 3) and Gardnerella greenwoodii sp. nov. (formerly Gardnerella genomic species 8) isolated from female urinary microbiome. *Int J Syst Evol Microbiol* **73** (2023). https://doi.org/10.1099/ijsem.0.006140

49 Allini Ntiguemassa, P. A. *et al.* Further dissection of Gardnerella vaginalis: description of Gardnerella lacydonensis sp. nov. (formerly genomic species 2), Gardnerella bretellae sp. nov. (formerly genomic species 9), Gardnerella massiliensis sp. nov. (formerly genomic species 14) and Gardnerella phocaeensis sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **76** (2026). https://doi.org/https://doi.org/10.1099/ijsem.0.007028

50 Teixeira, P., Sousa, M., Ribeiro, T. G., Grosso, F. & Peixe, L. *Unraveling theGardnerellaPangenome: Insights into Genetic Diversity, Virulence, and Antibiotic Resistance* (Cold Spring Harbor Laboratory, 2025).

51 Malki, K. *et al.* Genomes of Gardnerella Strains Reveal an Abundance of Prophages within the Bladder Microbiome. *PLOS ONE* **11**, e0166757 (2016). https://doi.org/10.1371/journal.pone.0166757

52 Gautreau, G. *et al.* PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. *PLOS Computational Biology* **16**, e1007732 (2020). https://doi.org/10.1371/journal.pcbi.1007732

53 Flores Ramos, S. *et al.* Genomic Stability and Genetic Defense Systems in Dolosigranulum pigrum , a Candidate Beneficial Bacterium from the Human Microbiome. *mSystems* **6** (2021). https://doi.org/10.1128/msystems.00425-21

54 Dillard, L. R., Glass, E. M., Lewis, A. L., Thomas-White, K. & Papin, J. A. Metabolic Network Models of the Gardnerella Pangenome Identify Key Interactions with the Vaginal Environment. *mSystems* **8** (2023). https://doi.org/10.1128/msystems.00689-22

55 Garcia, E. M., Kraskauskiene, V., Koblinski, J. E. & Jefferson, K. K. Interaction of Gardnerella vaginalis and Vaginolysin with the Apical versus Basolateral Face of a Three-Dimensional Model of Vaginal Epithelium. *Infection and Immunity* **87** (2019). https://doi.org/10.1128/iai.00646-18

56 Cauci, S. *et al.* Pore-forming and haemolytic properties of the Gardnerella vaginalis cytoiysin. *Molecular Microbiology* **9**, 1143-1155 (1993). https://doi.org/10.1111/j.1365-2958.1993.tb01244.x

57 Robinson, L. S., Schwebke, J., Lewis, W. G. & Lewis, A. L. Identification and characterization of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium Gardnerella vaginalis. *Journal of Biological Chemistry* **294**, 5230-5245 (2019). https://doi.org/10.1074/jbc.ra118.006221

58 Agarwal, K. & Lewis, A. L. Vaginal sialoglycan foraging byGardnerella vaginalis: mucus barriers as a meal for unwelcome guests? *Glycobiology* **31**, 667-680 (2021). https://doi.org/10.1093/glycob/cwab024

59 Pelayo, P. *et al.* Prevotella are major contributors of sialidases in the human vaginal microbiome. *Proceedings of the National Academy of Sciences* **121** (2024). https://doi.org/10.1073/pnas.2400341121

60 Perna, V. N., Barrett, K., Meyer, A. S. & Zeuner, B. Substrate specificity and transglycosylation capacity of α-L-fucosidases across GH29 assessed by bioinformatics-assisted selection of functional diversity. *Glycobiology* **33**, 396-410 (2023). https://doi.org/10.1093/glycob/cwad029

61 Watanabe, S. Characterization of l-2-keto-3-deoxyfuconate aldolases in a nonphosphorylating l-fucose metabolism pathway in anaerobic bacteria. *Journal of Biological Chemistry* **295**, 1338-1349 (2020). https://doi.org/10.1016/s0021-9258(17)49890-5

62 Watanabe, S. Characterization of a novel L-fuconate dehydratase involved in the non-phosphorylated pathway of L-fucose metabolism from bacteria. *Bioscience, Biotechnology, and Biochemistry* **88**, 177-180 (2024). https://doi.org/10.1093/bbb/zbad161

63 Kurukulasuriya, S. P., Patterson, M. H. & Hill, J. E. Slipped-Strand Mispairing in the Gene Encoding Sialidase NanH3 in Gardnerella spp. *Infection and Immunity* **89** (2021). https://doi.org/10.1128/iai.00583-20

64 Bhandari, P., Tingley, J., Abbott, D. W. & Hill, J. E. Glycogen-Degrading Activities of Catalytic Domains of α-Amylase and α-Amylase-Pullulanase Enzymes Conserved in Gardnerella spp. from the Vaginal Microbiome. *Journal of Bacteriology* **205** (2023). https://doi.org/10.1128/jb.00393-22

65 Bordigoni, A., Bouchard, S. & Desnues, C. Isolation and Characterization of Gardnerella Phage vB_Gva_AB1, a Bacteriophage Infecting a Clinical Strain of Gardnerella vaginalis. *Microbiology Resource Announcements* **10** (2021). https://doi.org/10.1128/mra.00053-21

66 Meier-Kolthoff, J. P. & Göker, M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* **33**, 3396-3404 (2017). https://doi.org/10.1093/bioinformatics/btx440

67 Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nature Reviews Microbiology* **21**, 686-700 (2023). https://doi.org/10.1038/s41579-023-00934-x

68 Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Research* **42**, 10618-10631 (2014). https://doi.org/10.1093/nar/gku734

69 Tock, M. R. & Dryden, D. T. F. The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8**, 466-472 (2005). https://doi.org/https://doi.org/10.1016/j.mib.2005.06.003

70 Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nature Communications* **13** (2022). https://doi.org/10.1038/s41467-022-30269-9

71 Tal, N. *et al.* Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nature Microbiology* **7**, 1200-1209 (2022). https://doi.org/10.1038/s41564-022-01158-0

72 Garvey, P., Fitzgerald, G. F. & Hill, C. Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Applied and Environmental Microbiology* **61**, 4321-4328 (1995). https://doi.org/10.1128/aem.61.12.4321-4328.1995

73 Deng, Y.-M., Harvey, M. L., Liu, C.-Q. & Dunn, N. W. A novel plasmid-encoded phage abortive infection system from Lactococcus lactis biovar. diacetylactis. *FEMS Microbiology Letters* **146**, 149-154 (1997). https://doi.org/10.1111/j.1574-6968.1997.tb10185.x

74 Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691-695 (2019). https://doi.org/10.1038/s41586-019-1605-5

75 Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host & Microbe* **30**, 1556-1569.e1555 (2022). https://doi.org/10.1016/j.chom.2022.09.017

76 Mestre, M. R. *et al.* UG/Abi: a highly diverse family of prokaryotic reverse transcriptases associated with defense functions. *Nucleic Acids Research* **50**, 6084-6101 (2022). https://doi.org/10.1093/nar/gkac467

77 Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077-1084 (2020). https://doi.org/10.1126/science.aba0372

78 Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018). https://doi.org/10.1126/science.aar4120

79 Chen, C. *et al.* Development of a Pre-Modification Strategy to Overcome Restriction–Modification Barriers and Enhance Genetic Engineering in Lactococcus lactis for Nisin Biosynthesis. *International Journal of Molecular Sciences* **26**, 2200 (2025). https://doi.org/10.3390/ijms26052200

80 Hu, S. *et al.* Altering under-represented DNA sequences elevates bacterial transformation efficiency. *mBio* **14** (2023). https://doi.org/10.1128/mbio.02105-23

81 Finn, M. B., Ramsey, K. M., Tolliver, H. J., Dove, S. L. & Wessels, M. R. Improved transformation efficiency of group A Streptococcus by inactivation of a type I restriction modification system. *PLOS ONE* **16**, e0248201 (2021). https://doi.org/10.1371/journal.pone.0248201

82 Klimowicz, A. K., Garcia, E. M., Jefferson, K. K. & Dillard, J. P. Genetic transformation of Gardnerella species and characterization of vaginolysin and sialidase mutants. *Infection and Immunity* **93** (2025). https://doi.org/10.1128/iai.00299-25

83 Kularatne, B. M. D. N. & Hill, J. E. Transformation of Gardnerella vaginalis with a Bifidobacterium-Escherichia coli shuttle vector plasmid. *Microbiology Spectrum* **13** (2025). https://doi.org/10.1128/spectrum.00481-25

84 Zhu, M. *et al.* Vaginal Lactobacillus fatty acid response mechanisms reveal a novel strategy for bacterial vaginosis treatment. *bioRxiv*, 2023.2012.2030.573720 (2023). https://doi.org/10.1101/2023.12.30.573720

85 Kharsany, A. B., Hoosen, A. A. & Van Den Ende, J. Antimicrobial susceptibilities of Gardnerella vaginalis. *Antimicrobial Agents and Chemotherapy* **37**, 2733-2735 (1993). https://doi.org/10.1128/aac.37.12.2733

86 Johnston, C. D. *et al.* Systematic evasion of the restriction-modification barrier in bacteria. *Proceedings of the National Academy of Sciences* **116**, 11454-11459 (2019). https://doi.org/10.1073/pnas.1820256116

87 Srinivasan, S. *et al.* More Easily Cultivated Than Identified: Classical Isolation With Molecular Identification of Vaginal Bacteria. *Journal of Infectious Diseases* **214**, S21-S28 (2016). https://doi.org/10.1093/infdis/jiw192

88 Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Research* **44**, D67-D72 (2016). https://doi.org/10.1093/nar/gkv1276

89 Meier-Kolthoff, J. P. & Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications* **10** (2019). https://doi.org/10.1038/s41467-019-10210-3

90 Alessandri, G., Van Sinderen, D. & Ventura, M. The genus Bifidobacterium: from genomics to functionality of an important component of the mammalian gut microbiota. *Computational and Structural Biotechnology Journal* **19**, 1472-1487 (2021). https://doi.org/10.1016/j.csbj.2021.03.006

91 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043-1055 (2015). https://doi.org/10.1101/gr.186072.114

92 Aziz, R. K. *et al.* The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**, 75 (2008). https://doi.org/10.1186/1471-2164-9-75

93 Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9** (2018). https://doi.org/10.1038/s41467-018-07641-9

94 Gu, Z. Complex heatmap visualization. *iMeta* **1** (2022). https://doi.org/10.1002/imt2.43

95 Auch, A. F., Von Jan, M., Klenk, H.-P. & Göker, M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences* **2**, 117-134 (2010). https://doi.org/10.4056/sigs.531120

96 Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60 (2013). https://doi.org/10.1186/1471-2105-14-60

97 Blum, M. *et al.* InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Research* **53**, D444-D456 (2025). https://doi.org/10.1093/nar/gkae1082

98 barrnap v. 0.9 (Australia, 2018).

99 Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004). https://doi.org/10.1186/1471-2105-5-113

100 Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593 (2011). https://doi.org/10.1093/bioinformatics/btq706

101 Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**, W78-W82 (2024). https://doi.org/10.1093/nar/gkae268

102 Gardner, S. N., Slezak, T. & Hall, B. G. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* **31**, 2877-2878 (2015). https://doi.org/10.1093/bioinformatics/btv271

103 Price, M. N., Deutschbauer, A. M. & Arkin, A. P. GapMind: Automated Annotation of Amino Acid Biosynthesis. *mSystems* **5** (2020). https://doi.org/10.1128/msystems.00291-20

104 Price, M. N., Deutschbauer, A. M. & Arkin, A. P. Filling gaps in bacterial catabolic pathways with computation and high-throughput genetics. *PLOS Genetics* **18**, e1010156 (2022). https://doi.org/10.1371/journal.pgen.1010156

105 Price, M. N. *et al. Improving the annotation of amino acid biosynthesis pathways: GapMind 2024* (openRxiv, 2024).

106 Payne, L. J. *et al.* PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. *Nucleic Acids Research* **50**, W541-W550 (2022). https://doi.org/10.1093/nar/gkac400

107 Payne, L. J. *et al.* Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Research* **49**, 10868-10878 (2021). https://doi.org/10.1093/nar/gkab883

108 Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8** (2020). https://doi.org/10.1186/s40168-020-00867-0

109 Perfilyev, A. *et al.* Large-scale analysis of bacterial genomes reveals thousands of lytic phages. *Nature Microbiology* **11**, 42-52 (2025). https://doi.org/10.1038/s41564-025-02203-4

110 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530-1534 (2020). https://doi.org/10.1093/molbev/msaa015

111 Greenwood, J. R. & Pickett, M. J. Transfer of Haemophilus vaginalis Gardner and Dukes to a New Genus, Gardnerella: G. vaginalis (Gardner and Dukes) comb. nov. *International Journal of Systematic Bacteriology* **30**, 170-178 (1980). https://doi.org/10.1099/00207713-30-1-170

112 Hitch, T. C. A. *et al.* Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria. *ISME Communications* **1** (2021). https://doi.org/10.1038/s43705-021-00017-z