bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Slipped strand mispairing in the gene encoding cell wall associated sialidase
2 # NanH3 in *Gardnerella* spp.

3

4 Shakya P. Kurukulasuriya, Mo H. Patterson, Janet E. Hill[#]

5

6 Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Canada

7

8 Running head: Slipped-strand mispairing in *Gardnerella* sialidase

9

10 #Address correspondence to Janet E. Hill, Janet.Hill@usask.ca

11

12

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13 **Abstract**

14 Cell wall proteins with sialidase activity are involved in carbohydrate assimilation, adhesion to
15 mucosal surfaces, and biofilm formation. *Gardnerella* spp. inhabit the human vaginal
16 microbiome and encode up to three sialidase enzymes, two of which are cell wall associated.
17 Here we demonstrate that the gene encoding extracellular sialidase NanH3 is found almost
18 exclusively in *G. piotii* and closely related *Gardnerella* genome sp. 3, and its presence correlates
19 with sialidase positive phenotype in a collection of 112 *Gardnerella* isolates. The *nanH3* gene
20 sequence includes a homopolymeric repeat of cytosines that varies in length within cell
21 populations, indicating that this gene is subject to slipped-strand mispairing, a mechanisms of
22 phase variation in bacteria. Variation in the length of the homopolymer sequence results in
23 encoding of either the full length sialidase protein or truncated peptides due to introduction of
24 reading-frame shifts and premature stop codons. Phase variation in this extracellular, cell wall
25 associated sialidase may be involved in immune evasion or modulation of adhesion to host
26 epithelial cells, and formation of biofilms characteristic of the vaginal dysbiosis known as
27 bacterial vaginosis.

28

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

29 **Introduction**

30 Bacterial vaginosis (BV) is a condition that is characterized by altered composition of the
31 vaginal microbiota and occurs when the healthy microbiota is replaced by an overgrowth of
32 mixed aerobic and anaerobic species, including *Gardnerella* spp. (1). An abundance of
33 *Gardnerella* spp. is often found in cases of symptomatic BV, although they are also found in
34 healthy women with no clinical signs or symptoms of BV (2). *Gardnerella* spp. can be resolved
35 into four subgroups based on cpn60 barcodes sequencing (3) or whole genome sequencing (4).
36 Recently, the description of *Gardnerella vaginalis* was amended and three new species were
37 defined within the genus *Gardnerella*: *G. leopoldii*, *G. swidsinskii*, and *G. piotii* (5). *G. piotii* and
38 *G. vaginalis* correspond to cpn60 subgroup B and C, respectively. *G. leopoldii* and G. *swidsinskii*
39 were previously grouped together as subgroup A based on cpn60 sequences of the isolates
40 available at that time. Isolates belonging to cpn60 subgroup D corresponded to three distinct
41 genome species: genome species 8, 9 and 10 (6).

42 In addition to the characteristic change in microbiota, and elevated pH, sialidase activity
43 in vaginal fluid is a diagnostic marker of BV (7, 8). Sialidase enzymes cleave the glycosidic
44 linkages of sialic acids from terminal glycans including vaginal mucins, immunoglobulin A
45 molecules, and epithelial cell surface glycoproteins (9, 10). Activity on the latter may be related
46 to the adhesion of bacteria to epithelial cells; the initial step in biofilm formation (11).
47 Previously, we have demonstrated that sialidase activity is almost exclusively confined to *G.*
48 *piotii* and *Gardnerella* genome species 3 (cpn60 subgroup B) (12).

49 A putative sialidase gene, *nanH1* (sialidase A) identified in *Gardnerella* spp. and was
50 initially thought to be the gene responsible for sialidase activity (13). Although *nanH1* appears to
51 be present in all sialidase positive strains, it is also found in sialidase activity negative strains

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52 (12). This observation combined with the lack of signal peptide on the NanH1 proteins suggests
53 that this protein is an intracellular enzyme, likely involved in manipulation of nutritional
54 substrates as is the case in many bacteria (14). The discrepancy between *nanH1* presence and
55 enzyme activity led to the recognition of two additional sialidase genes (*nanH2* and *nanH3*) in
56 some *Gardnerella* isolates (15). The authors of this report concluded that NanH2 and NanH3 are
57 extracellular and that they are the proteins responsible for sialidase activity detected in vaginal
58 fluid. Cell wall anchored proteins with sialidase activity in other species have been found to be
59 involved in adhesion to mucosal surfaces and biofilm initiation as well as carbohydrate
60 assimilation (16–19).

61 Interestingly, *nanH3* contains a homopolymeric tract of cytosine residues. Genomic
62 regions that contain short, homogenous or heterogenous repeats are susceptible to slipped-strand
63 mispairing in which the length of the repeat region can change with each replication (20). The
64 result of this modulation is phase variation: a reversible process in which the expression of the
65 encoded protein can be rapidly switched on and off (21). Phase variation in cell surface proteins
66 can result in immune evasion and alteration of biofilm phenotypes (22, 23).

67 Here we determined the distribution of genes encoding cell-wall associated sialidases
68 NanH2 and NanH3 in the context of newly reclassified *Gardnerella* spp. and demonstrated that
69 *nanH3* is subject to slipped-strand mispairing.

70

71 **Methods**

72 **Protein domain identification and sequence alignment**

73 DNA and protein sequence alignments were performed with Clustal Omega
74 (EMBL_EBI) and NCBI BLAST (basic local alignment search tool). InterproScan

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

75 (https://www.ebi.ac.uk/interpro/) and SignalP was used to predict the location of functional
76 domains signal peptides (24).

77 **Bacterial strains and culture conditions**

78 *Gardnerella* strains (n = 112) from a previously described culture collection were used in
79 the study (12). Sialidase activity for all isolates (12), and whole genome sequences for 33 of
80 these isolates (25) were previously determined. Complete strain information and sequence
81 accessions are provided in Table S1.

82 *Gardnerella* isolates were grown on Columbia Sheep Blood Agar (BBL, Becton,
83 Dickinson and Company, Sparks, MD, USA) plates incubating at 37°C for 48 hours with
84 anaerobic BD GasPak EZ (Becton, Dickinson and Company, Sparks, MD, USA). For broth
85 cultures a few colonies from the plate were collected with a 10 µl inoculation loop and used to
86 inoculate NYC III (ATCC 1685 medium; per litre: 2.4g HEPES, 15 g Proteose peptone, 3.8 g
87 Yeast extract, 5 g NaCl, 5 g Glucose). Broth cultures were incubated at 37°C for 48 hours in
88 anaerobic conditions.

89 **PCR screen for *nanH3***

90 Genomic DNA was purified from broth cultures using a modified salting-out procedure
91 (26). All DNA extracts were initially tested by PCR for the universal cpn60 barcode to confirm
92 the quality of the DNA.

93 To screen isolates for the presence of *nanH3*, PCR primers were designed based on
94 multiple sequence alignments of 15 *nanH3* sequences obtained from the Integrated Microbial
95 Genomes database (https://img.jgi.doe.gov/). Degenerate primers were designed to account for
96 sequence variability within the gene sequence and to amplify a product of 375 bp in length
97 (JH0684: 5'-GTT GTA GAR CTT TCT GAT GG-3', JH0685: 5'-YRY TAT TAT CGC CCT

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

98 CAT ATA-3'). PCR reactions contained 1 × PCR Buffer (0.2 M Tris-HCl at pH 8.4, 0.5 M KCl),
99 2.5 μM $MgCl_2$, 0.40 μM dNTP, 0.20 μM forward primer, 0.20 μM reverse primer, 2 U Taq DNA
100 Polymerase, ultrapure water and 2 μl of template DNA in a final volume of 50 μl. PCR reactions
101 were conducted using the following thermocycling parameters in a Mastercycler Pro 6321
102 (Eppendorf AG, Hamburg, Germany): 94 °C for 3 minutes, 40 cycles of (94 °C for 30 seconds,
103 55 °C for 30 seconds, 72 °C for 30 seconds), 72 °C for 1 minute, hold at 20 °C. PCR products
104 were visualized under UV light on a 1.0 % agarose gel containing ethidum bromide.

105 **Homopolymer PCR, cloning and sequencing**

106 To determine the length of the homopolymer region of *nanH3,* four strains (W11,
107 VN014, VN015, NR032) were cultured on Columbia Agar plates with 5% Sheep Blood. Primers
108 JH0780 (5'-ATG ATT GGA ACA GCG CAT AAA G-3') and JH0781 (5'-GAT TTC TCC ACC
109 TAC AGT TAC C-3') were designed to PCR amplify a region including basepairs 2-310 of the
110 open reading frame of *nanH3*.

111 The DNA sequence (308 bp) was amplified by PCR in a Mastercycler Pro 6321
112 (Eppendorf AG, Hamburg, Germany). The components of the PCR reaction mix (50 μl per
113 reaction) were added to achieve final concentrations of 1×High Fidelity PCR buffer (60 mM
114 Tris-SO_4 (pH 8.9), 18 mM $(NH_4)_2SO_4$), 0.2 mM dNTP mix, 2 mM $MgSO_4$, 1U Platinum High
115 Fidelity (Hi-Fi) proof reading Taq polymerase (Invitrogen, Carlsbad, CA, USA). Two colonies
116 of each *G. piotii* or genome sp. 3 strain were randomly picked and added to separate PCR
117 reactions using sterile toothpicks. Thermocycling conditions included 35 cycles of denaturation
118 at 94°C for 15 seconds, annealing at 55°C for 30 seconds, extension at 68°C for 1 minute, and
119 final extension at 68°C for 5 minutes. PCR products were visualized on a 1% agarose gel. PCR

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

120 products were purified using QiaQuick PCR purification kit (Qiagen, Hilden, Germany) and
121 purified PCR products were sequenced using the amplification primers (JH0780, JH0781).
122 To clarify the exact length of the homopolymer region, amplicons generated from the two
123 colonies of each of the four strains were A-tailed in 10 µl reactions containing 1× Platinum PCR
124 buffer (20 mM Tris HCl (pH 8.4), 50 mM KCl), 2 mM $MgCl_2$, 0.5 mM dATP, 5 U Platinum Taq
125 polymerase (Invitrogen, Carlsbad, CA, USA) and less than 500 ng of the purified PCR product.
126 The reaction mixture was incubated at 72 °C for 20 minutes in Mastercycler Pro 6321
127 (Eppendorf AG, Hamburg, Germany). End-modified PCR products were ligated into pGEM-T
128 Easy vector (Promega, Madison, WI, USA). The vector-insert construct was used to transform
129 chemically competent DH5α cells or OneShot Top 10 *E. coli* (Invitrogen, Carlsbad, CA, USA)
130 and plated on LB/AMP/X-gal agar media. Ten white colonies were randomly selected from the
131 transformants of each *Gardnerella* strain and transferred into LB + ampicillin broth. Cultures
132 were grown overnight at 37 °C. Plasmid DNA was isolated using QiaPrep Spin Miniprep kit
133 (Qiagen, Hilden, Germany). Plasmids were sequenced using vector primers T7 and SP6.
134
135 **Results**
136 **Distribution of extracellular sialidases in *Gardnerella* spp. isolates**.
137 Robinson et al. (15) had previously reported that sialidase activity was associated with
138 the presence of either *nanH2* or *nanH3* in a collection of 34 *Gardnerella* isolates but did not
139 report the species or subgroup affiliation. In order to confirm this observation and to reconcile
140 the distribution of genes and activities with the new *Gardnerella* taxonomic framework (5), we
141 queried 33 isolates in our culture collection for which whole genome sequence data and sialidase
142 activity data were available (Table S1). The presence of the genes was assessed by aligning

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

143 *nanH2* of JCP8151B (ATJH01000056) and *nanH3* of W11 to the genome sequences using
144 BLASTn. Sialidase activity was reported in 9/33 isolates (Table 1) and all nine were *G. piotii* or
145 genome sp. 3 (cpn60 subgroup B). The intact ORF of *nanH3* was found in 8/9 sialidase activity
146 positive isolates while *nanH2* was present in only 4/9 isolates. Three isolates possessed both
147 *nanH2* and *nanH3* genes in their sequences. VN002 was sialidase activity positive but only
148 *nanH2* was identified in the genome sequence. This isolate subsequently screen PCR positive for
149 *nanH3* (see next section), suggesting that the gene was not included in the shotgun assembly of
150 the genome, which consisted of multiple gap-containing scaffolds. None of the *G. vaginalis, G.*
151 *swidsinskii, G. leopoldi* or subgroup D isolates were sialidase positive and none contained *nanH2*
152 or *nanH3*.

153 **Correlation of *nanH3* with sialidase activity**

154 In order to examine further the relationship of *nanH3* to sialidase activity, 112
155 *Gardnerella* spp. isolates for which sialidase activity data was available (12) was screened for
156 the presence of *nanH3* using PCR primers JH0684/JH0685 (Figure 1). Sialidase activity was
157 detected in 32/33 *G. piotii* isolates and 3/33 *G. vaginalis* isolates. All sialidase activity positive
158 isolates were positive for *nanH3* by PCR with the exception of one *G. piotii* isolate (WP027) that
159 was PCR negative. Unfortunately, genome sequence data was not available for WP027.

160 **Characterization of a poly-C homopolymer in *nanH3***

161 The protein sequences encoded by all *nanH3* sequences were predicted to encode a signal
162 peptide (amino acids 1-32), a sialidase domain (amino acids 181-852 in W11) and a C-terminal
163 transmembrane domain (amino acids 783-805 in W11). We observed a poly-cytosine
164 homopolymer (8-14 cytosines) in all *nanH3* genes identified in the whole genome sequences,
165 approximately 100 bases from the start of the open reading frame. The homopolymer occurred

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

166 immediately following the region of the sequence predicted to encode the signal peptide (amino
167 acids 1-32) (Figure 2). When Sanger sequencing was performed on PCR products corresponding
168 to nucleotides 2-310 of *nanH3* amplified from genomic DNA extracted from broth cultures of
169 strain W11, the results suggested variable lengths of the homopolymer. Specifically, clean data
170 was obtained upstream of the poly-C tract but sequence 3' to that region was indicative of a
171 mixed template (Figure S1). Since it has been demonstrated that simple repeats such as
172 homopolymers can be subject to slipped strand mispairing, we set out to determine if the
173 homopolymer region of *nanH3* varied in length within and between strains of *G. piotii*.

174 PCR product libraries were made from two colonies each of isolates W11, VN014,
175 VN015 and NR032 (8 clone libraries total). Plasmids were purified from 10 colonies from each
176 of the 8 clone libraries and sequenced. High quality sequence data was obtained from 15, 14, 18
177 and 16 clones from the W11, VN014, VN015 and NR032 libraries, respectively. Example results
178 are shown in Figure S2. The length of the homopolymer varied from 8 to 14 among all strains
179 (Table 2), and within each strain the sequence flanking the homopolymer was identical. *In silico*
180 translation of the encoded polypeptides showed that most homopolymer lengths resulted in a
181 truncated peptide (37-43 amino acids), while full-length protein (812 aa) would result when there
182 were nine or twelve cytosines in the homopolymer region. Interestingly, the most common
183 length of the poly-C region varied among strains. C_{10} was most common in VN014 and VN015,
184 while C_9 and C_{12} were most frequently observed in NR032 and W11, respectively. VN015 had
185 the highest number of homopolymer variants (Figure 3). A second open reading frame within
186 the *nanH3* open reading frame was also identified, corresponding to amino acids Met_{397}-Tyr_{812} of
187 the NanH3 protein. These 415 aa polypeptides encompass part of the predicted sialidase domain,
188 and the C-terminal transmembrane domain but lacks a signal peptide.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

189

190 **Discussion**

191 The genus *Gardnerella* is known as a hallmark of BV and their abundance is used as a
192 criterion for the laboratory diagnosis of the condition (27). In addition to participating in biofilms
193 that coat epithelial cells in BV, *Gardnerella* spp. produce enzymes and a cholesterol dependent
194 cytolysin that contribute to degrading the protective barriers of the vaginal mucosa (28).
195 Sialidase activity can provide nutrients by releasing sialic acid moieties from vaginal sialylated
196 mucins (10), altering the physical properties of vaginal mucus. Removal of sialic acid residues
197 from epithelial cell surface glycans can facilitate bacterial adhesion and initiation of biofilm
198 formation (11, 16, 18, 19).

199 The two putative sialidase genes (*nanH2* and *nanH3*) in *Gardnerella* spp. have been
200 demonstrated to encode proteins with high enzymatic potency (15). Taken together, the result of
201 this study and our current study show that *nanH3* is more common and *nanH2* is virtually never
202 found without *nanH3*. It is now clear that extracellular sialidase activity is a property of *G. piotii*
203 and the closely related *Gardnerella* genome sp. 3. We found two occurrences of sialidase
204 positive, *nanH3* positive *G. vaginalis* isolates (2/33 isolates screened), and Vaneechoutte et al.
205 (5) reported sialidase activity in 1/4 *G. vaginalis* isolates used in the amendment of the genus.
206 This infrequent prevalence of *nanH3* in *G. vaginalis* could be the result of lateral gene transfer
207 from *G. piotii* since these species co-exist in the same microbiome and women are usually
208 colonized by more than one species (6, 29, 30).

209 All of the *nanH3* sequences we examined were predicted to encode a signal peptide, a
210 sialidase domain and a C-terminal membrane domain, suggesting a cell-wall tethered protein,
211 similar to SiaBb2 of *Bifidobacterium bifidum*, which enhances adhesion to intestinal mucosal

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

212 surfaces and contributes to carbohydrate assimilation (19). Interestingly, the NanH3 of
213 *Gardnerella* strain JCP8151B investigated by Robinson et al. was reported to lack a signal
214 peptide. When we examined the sequence upstream of *nanH3* in the JCP8151B sequence
215 (Genbank accession ATJH01000033), we found that there is an alternative start codon, and the
216 homopolymer we observed in all of our sequences. The frame-shift that resulted in the
217 annotation of the gene without the signal peptide-encoding N-terminus is caused by the
218 homopolymer.

219 Genomic regions that contain homogenous or heterogenous repeats are prone to changes
220 in length of the repeat at each replication due to slipped strand mispairing (20). This can lead to
221 consequent changes in the transcription or translation product of a gene depending on whether
222 the slipped strand mispairing occurs within the open reading frame or in extragenic regions such
223 as promoter sequences. Phase variation results when slipped strand mispairing creates "on" and
224 "off" states of expression of a protein. Phase variation of cell surface proteins has been
225 documented in many bacterial species including *Neisseria* spp. (31–33), *Salmonella* spp. (34–
226 36), *Treponema pallidum* (37, 38) and *Helicobacter pylori* (23). It was first described in the *opa*
227 genes that encode opacity surface proteins of *Neisseria* spp. (39). The *opa* genes contain CTCTT
228 pentamer repeats within the signal peptide encoding region and the expression of the Opa protein
229 is regulated by slipped-strand mispairing. With six, nine or twelve CTCTT repeats the initiation
230 codon is in frame with the remaining *opa* gene translating the Opa protein. Four or eight coding
231 repeats makes the initiation codon out of frame with rest of the *opa* codons. The phase "on" state
232 allows them to adhere to specific surface receptors on host cells (40).

233 Our results clearly show variation of the homopolymer length among *Gardnerella piotii*
234 within colonies growing on agar, with anywhere from 8 to 14 C's observed. The effect of this

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

235 change would be a mixture of cells with translation of NanH3 on or off. Why would *Gardnerella*
236 *piotii* have a sialidase subject to phase variation? Two possible reasons are immune evasion and
237 cell adhesion (biofilm initiation and dispersal).

238 Although BV is often referred to as a non-inflammatory condition because of the lack of
239 typical clinical signs of inflammation (leukocyte infiltration, pain, redness, swelling), there is
240 evidence of a host immune response to bacteria associated with BV, including *Gardnerella* spp..
241 IgA specific for *Gardnerella* produced vaginolysin has been detected in women with clinical
242 signs of BV (41), and IgA levels correlate with IL-8 expression in vaginal secretions (42).
243 *Gardnerella* has also been shown to stimulate pro-inflammatory cytokines *in vitro* (43).
244 Balancing this inflammatory response are the actions of bacterial enzymes like prolidases and
245 sialidases that can degrade the effectors of inflammation. Whether there is a specific host
246 response to NanH3 remains to be determined.

247 The initial step in biofilm formation is adhesion, which in the case of BV associated
248 biofilms is to epithelial cells. *Streptococcus pneumoniae* surface proteins with sialidase activity
249 have been shown to reveal carbohydrate ligands for bacterial adhesion to host cells (16).
250 Similarly, Soong et al. (17) showed that a cell wall associated sialidase of *Pseudomonas*
251 *aeruginosa* plays a critical role in facilitating respiratory mucosa infection through biofilm
252 formation. Specific antibodies against *B. bifidum* SiaBb2, a cell wall tethered sialidase, inhibit
253 adhesion to cells (19). The involvement of *Gardnerella* sialidases in adhesion has not been
254 demonstrated directly, although it has been reported that inhibition of sialidase activity in
255 *Gardnerella* strain JCP8066 by an antiviral drug (Zanamavir) resulted in reduction of adherence
256 to vaginal epithelial cells *in vitro* (11). *Gardnerella* spp. can form multi-species biofilms *in vitro*
257 (44) and although "*Gardnerella vaginalis*" has been shown to participate in multispecies

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

258 biofilms with other BV associated bacteria *in vivo* (45, 46)*,* it is not yet known how each of the
259 *Gardnerella* spp. participate in this process. Interestingly, *G. piotii* and *Gardnerella* genome sp.
260 3 (cpn60 subgroup B) have been previously associated with "intermediate" grades of vaginal
261 dysbiosis (6, 29, 47, 48) and may contribute to transition from eubyosis to dysbiosis and enhance
262 colonization by other BV associated anaerobes. Phase variation in a cell surface associated
263 sialidase enzyme might be critical for turning on or off the initial adhesion of *Gardnerella* to
264 epithelial cells and thus the cascade of events following that result in the biofilm covered "clue
265 cells" typical of BV (49).

266 It seems likely that some of the NanH3 protein produced by *G. piotii* does not remain
267 tethered to the cell surface, as is the case with *S. pneumoniae* (50), since the sequence upstream
268 of the C-terminal transmembrane domain may be susceptible to proteolytic cleavage, and
269 heterologously expressed NanH3 lacking the transmembrane domain is active (15). Taken
270 together, the activity of NanH3 while anchored to the cell surface or released into the
271 extracellular environment, and the likelihood that it is subject to phase variation through slipped-
272 strand mispairing, make this protein an intriguing puzzle for future study. Development of
273 genetic tools to create mutants where NanH3 production is locked "on" or "off" would be a
274 significant step toward understanding the role of this protein in the initiation and maintenance of
275 vaginal dysbiosis.

276

277 **Acknowledgements**

278 This research was funded by a Natural Sciences and Engineering Council of Canada Discovery
279 Grant to JEH. The authors are grateful to Champika Fernando for excellent technical assistance,
280 and to all our colleagues in the Hill Lab for helpful discussions and feedback.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

281

282 **References**

283 1. Nugent RP, Krohn MA, Hillier SL. 1991. Reliability of diagnosing bacterial vaginosis is
284 improved by a standardized method of gram stain interpretation. J Clin Microbiol 29:297–
285 301.

286 2. van de Wijgert JH, Borgdorff H, Verhelst R, Crucitti T, Francis S, Verstraelen H, Jespers
287 V. 2014. The vaginal microbiota: what have we learned after a decade of molecular
288 characterization? PLoS One 9:e105998.

289 3. Paramel Jayaprakash T, Schellenberg JJ, Hill JE. 2012. Resolution and characterization of
290 distinct cpn60-based subgroups of *Gardnerella vaginalis* in the vaginal microbiota. PLoS
291 ONE 7:e43009.

292 4. Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, Powell E, Janto B, Eutsey
293 R, Hiller NL, Boissy R, Dahlgren ME, Hall BG, Costerton JW, Post JC, Hu FZ, Ehrlich
294 GD. 2012. Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis*
295 provide evidence of multiple genetically isolated clades consistent with subspeciation into
296 genovars. J Bacteriol 194:3922–37.

297 5. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P.
298 2019. Emended description of *Gardnerella vaginalis* and description of *Gardnerella*
299 *leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with
300 delineation of 13 genomic species within the genus *Gardnerella*. Int J Syst Evol Microbiol
301 69:679–687.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

302 6. Hill JE, Albert AYK. 2019. Resolution and co-occurrence patterns of *Gardnerella*
303 *leopoldii, Gardnerella swidsinskii, Gardnerella piotii* and *Gardnerella vaginalis* within the
304 vaginal microbiome. Infect Immun 87:e00532-19.

305 7. McGregor JA, French JI, Jones W, Milligan K, McKinney PJ, Patterson E, Parker R. 1994.
306 Bacterial vaginosis is associated with prematurity and vaginal fluid mucinase and sialidase:
307 results of a controlled trial of topical clindamycin cream. Am J Obstet Gynecol 170:1048–
308 1059.

309 8. Myziuk L, Romanowski B, Johnson SC. 2003. BVBlue test for diagnosis of bacterial
310 vaginosis. J Clin Microbiol 41:1925–8.

311 9. Briselden AM, Moncla BJ, Stevens CE, Hillier SL. 1992. Sialidases (neuraminidases) in
312 bacterial vaginosis and bacterial vaginosis-associated microflora. J Clin Microbiol 30:663–
313 6.

314 10. Lewis WG, Robinson LS, Gilbert NM, Perry JC, Lewis AL. 2013. Degradation, foraging,
315 and depletion of mucus sialoglycans by the vagina-adapted Actinobacterium *Gardnerella*
316 *vaginalis*. J Biol Chem 288:12067–79.

317 11. Govinden G, Parker JL, Naylor KL, Frey AM, Anumba DOC, Stafford GP. 2018. Inhibition
318 of sialidase activity and cellular invasion by the bacterial vaginosis pathogen *Gardnerella*
319 *vaginalis*. Arch Microbiol 200:1129–1133.

320 12. Schellenberg JJ, Paramel Jayaprakash T, Withana Gamage N, Patterson MH, Vaneechoutte
321 M, Hill JE. 2016. *Gardnerella vaginalis* subgroups defined by cpn60 sequencing and
322 sialidase activity in isolates from Canada, Belgium and Kenya. PLoS One 11:e0146510.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

323 13. Santiago GL, Deschaght P, El Aila N, Kiama TN, Verstraelen H, Jefferson KK,
324 Temmerman M, Vaneechoutte M. 2011. *Gardnerella vaginalis* comprises three distinct
325 genotypes of which only two produce sialidase. Am J Obstet Gynecol 204:450.e1–7.

326 14. Corfield T. 1992. Bacterial sialidases—roles in pathogenicity and nutrition. Glycobiology
327 2:509–521.

328 15. Robinson LS, Schwebke J, Lewis WG, Lewis AL. 2019. Identification and characterization
329 of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium
330 *Gardnerella vaginalis*. J Biol Chem 294:5230–5245.

331 16. Navarre WW, Schneewind O. 1999. Surface proteins of Gram-positive bacteria and
332 mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev MMBR
333 63:174–229.

334 17. Soong G, Muir A, Gomez MI, Waks J, Reddy B, Planet P, Singh PK, Kanetko Y, Wolfgang
335 MC, Hsiao Y-S, Tong L, Prince A. 2006. Bacterial neuraminidase facilitates mucosal
336 infection by participating in biofilm production. J Clin Invest 116:2297–2305.

337 18. Parker D, Soong G, Planet P, Brower J, Ratner AJ, Prince A. 2009. The NanA
338 neuraminidase of *Streptococcus pneumoniae* is involved in biofilm formation. Infect
339 Immun 77:3722–30.

340 19. Nishiyama K, Yamamoto Y, Sugiyama M, Takaki T, Urashima T, Fukiya S, Yokota A,
341 Okada N, Mukai T. 2017. *Bifidobacterium bifidum* extracellular sialidase enhances
342 adhesion to the mucosal surface and supports carbohydrate assimilation. MBio 8.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

343 20. Woude MW van der, Bäumler AJ. 2004. Phase and antigenic variation in bacteria. Clin
344 Microbiol Rev 17:581–611.

345 21. Phillips ZN, Tram G, Seib KL, Atack JM. 2019. Phase-variable bacterial loci: how bacteria
346 gamble to maximise fitness in changing environments. Biochem Soc Trans 47:1131–1141.

347 22. Seib KL, Jen FE-C, Scott AL, Tan A, Jennings MP. 2017. Phase variation of DNA
348 methyltransferases and the regulation of virulence and immune evasion in the pathogenic
349 Neisseria. Pathog Dis 75.

350 23. Bergman M, Del Prete G, van Kooyk Y, Appelmelk B. 2006. Helicobacter pylori phase
351 variation, immune modulation and gastric autoimmunity. 2. Nat Rev Microbiol 4:151–159.

352 24. Nielsen H, Tsirigos KD, Brunak S, von Heijne G. 2019. A brief history of protein sorting
353 prediction. Protein J 38:200–216.

354 25. Khan S, Vancuren SJ, Hill JE. 2020. A generalist lifestyle allows rare *Gardnerella* spp. to
355 persist at low levels in the vaginal microbiome. bioRxiv 2020.08.25.267146.

356 26. Martin-Platero AM, Valdivia E, Maqueda M, Martinez-Bueno M. 2007. Fast, convenient,
357 and economical method for isolating genomic DNA from lactic acid bacteria using a
358 modification of the protein “salting-out” procedure. Anal Biochem 366:102–4.

359 27. Nugent RP, Krohn MA, Hillier SL. 1991. Reliability of diagnosing bacterial vaginosis is
360 improved by a standardized method of gram stain interpretation. J Clin Microbiol 29:297–
361 301.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

362 28. Schellenberg JJ, Patterson MH, Hill JE. 2017. *Gardnerella vaginalis* diversity and ecology
363 in relation to vaginal symptoms. Res Microbiol doi:10.1016/j.resmic.2017.02.011.

364 29. Balashov SV, Mordechai E, Adelson ME, Gygax SE. 2014. Identification, quantification
365 and subtyping of *Gardnerella vaginalis* in noncultured clinical vaginal samples by
366 quantitative PCR. J Med Microbiol 63:162–75.

367 30. Albert AY, Chaban B, Wagner EC, Schellenberg JJ, Links MG, van Schalkwyk J, Reid G,
368 Hemmingsen SM, Hill JE, Money D. 2015. A study of the vaginal microbiome in healthy
369 Canadian women utilizing cpn60-based molecular profiling reveals distinct *Gardnerella*
370 subgroup community state types. PLoS One 10:e0135620.

371 31. Bayliss CD, Hoe JC, Makepeace K, Martin P, Hood DW, Moxon ER. 2008. *Neisseria*
372 *meningitidis* escape from the bactericidal activity of a monoclonal antibody is mediated by
373 phase variation of lgtG and enhanced by a mutator phenotype. Infect Immun 76:5038–5048.

374 32. Apicella MA, Shero M, Jarvis GA, Griffiss JM, Mandrell RE, Schneider H. 1987.
375 Phenotypic variation in epitope expression of the *Neisseria gonorrhoeae*
376 lipooligosaccharide. Infect Immun 55:1755–1761.

377 33. van Putten JP. 1993. Phase variation of lipopolysaccharide directs interconversion of
378 invasive and immuno-resistant phenotypes of *Neisseria gonorrhoeae*. EMBO J 12:4043–
379 4051.

380 34. Lederberg J, Iino T. 1956. Phase variation in *Salmonella*. Genetics 41:743–757.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

381 35. Silverman M, Zieg J, Hilmen M, Simon M. 1979. Phase variation in *Salmonella*: genetic
382 analysis of a recombinational switch. Proc Natl Acad Sci 76:391–395.

383 36. Bonifield HR, Hughes KT. 2003. Flagellar phase variation in *Salmonella enterica* is
384 mediated by a posttranscriptional control mechanism. J Bacteriol 185:3567–3574.

385 37. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, Gwinn M, Hickey
386 EK, Clayton R, Ketchum KA, Sodergren E, Hardham JM, McLeod MP, Salzberg S,
387 Peterson J, Khalak H, Richardson D, Howell JK, Chidambaram M, Utterback T, McDonald
388 L, Artiach P, Bowman C, Cotton MD, Fujii C, Garland S, Hatch B, Horst K, Roberts K,
389 Sandusky M, Weidman J, Smith HO, Venter JC. 1998. Complete genome sequence of
390 *Treponema pallidum*, the syphilis spirochete. Science 281:375–388.

391 38. Pinto M, Borges V, Antelo M, Pinheiro M, Nunes A, Azevedo J, Borrego MJ, Mendonça J,
392 Carpinteiro D, Vieira L, Gomes JP. 2016. Genome-scale analysis of the non-cultivable
393 *Treponema pallidum* reveals extensive within-patient genetic variation. 1. Nat Microbiol
394 2:1–11.

395 39. Stern A, Brown M, Nickel P, Meyer TF. 1986. Opacity genes in *Neisseria gonorrhoeae*:
396 Control of phase and antigenic variation. Cell 47:61–71.

397 40. Stern A, Nickel P, Meyer TF, So M. 1984. Opacity determinants of *Neisseria gonorrhoeae*:
398 Gene expression and chromosomal linkage to the gonococcal pilus gene. Cell 37:447–456.

399 41. Cauci S, Scrimin F, Driussi S, Ceccone S, Monte R, Fant L, Quadrifoglio F. 1996. Specific
400 immune response against *Gardnerella vaginalis* hemolysin in patients with bacterial
401 vaginosis. Am J Obstet Gynecol 175:1601–5.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

402 42. Cauci S, Driussi S, Guaschino S, Isola M, Quadrifoglio F. 2002. Correlation of local
403 Interleukin-1beta levels with specific IgA response against *Gardnerella vaginalis* cytolysin
404 in women with bacterial vaginosis. Am J Reprod Immunol 47:257–264.

405 43. Eade CR, Diaz C, Wood MP, Anastos K, Patterson BK, Gupta P, Cole AL, Cole AM. 2012.
406 Identification and characterization of bacterial vaginosis-associated pathogens using a
407 comprehensive cervical-vaginal epithelial coculture assay. PLoS One 7:e50106.

408 44. Khan S, Voordouw MJ, Hill JE. 2019. Competition among *Gardnerella* subgroups from the
409 human vaginal microbiome. Front Cell Infect Microbiol 9:374.

410 45. Machado A, Cerca N. 2015. Influence of biofilm formation by *Gardnerella vaginalis* and
411 other anaerobes on bacterial vaginosis. J Infect Dis 212:1856–61.

412 46. Machado D, Castro J, Palmeira-de-Oliveira A, Martinez-de-Oliveira J, Cerca N. 2015.
413 Bacterial vaginosis biofilms: Challenges to current therapies and emerging solutions. Front
414 Microbiol 6:1528.

415 47. Hilbert DW, Schuyler JA, Adelson ME, Mordechai E, Sobel JD, Gygax SE. 2017.
416 *Gardnerella vaginalis* population dynamics in bacterial vaginosis. Eur J Clin Microbiol
417 Infect Dis https://doi.org/10.1007/s10096-017-2933-8.

418 48. Shipitsyna E, Krysanova A, Khayrullina G, Shalepo K, Savicheva A, Guschin A, Unemo
419 M. 2019. Quantitation of all four *Gardnerella vaginalis* clades detects abnormal vaginal
420 microbiota characteristic of bacterial vaginosis more accurately than putative *G. vaginalis*
421 sialidase A gene count. Mol Diagn Ther 23:139–147.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

422 49. Simoes JA, Discacciati MG, Brolazo EM, Portugal PM, Dini DV, Dantas MC. 2006.
423 Clinical diagnosis of bacterial vaginosis. Int J Gynaecol Obstet 94:28–32.

424 50. Lock RA, Paton JC, Hansman D. 1988. Purification and immunological characterization of
425 neuraminidase produced by *Streptococcus pneumoniae*. Microb Pathog 4:33–43.

426

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

427

428 **Table 1.** Distribution of *nanH2* and *nanH3* genes in *Gardnerella* spp. whole genome sequences

Isolate	cpn60 Subgroup	Species	Sialidase activity[1]	*nanH2*	*nanH3*
GH007	B	*G. piotii*	+	+	+
GH019		*G. piotii*	+	-	+
GH020		*G. piotii*	+	-	+
GH22		*G. piotii*	+	+	+
N170		Genome species 3	+	-	+
NR026		Genome species 3	+	-	+
VN002		*G. piotii*	+	+	-[2]
N144		Genome species 3	+	+	+
W11		Genome species 3	+	-	+
GH005	A	*G. leopoldii*	-	-	-
NR015		Unknown	-	-	-
NR016		*G. swidsinskii*	-	-	-
NR017		*G. leopoldii*	-	-	-
NR019		*G. leopoldii*	-	-	-
NR020		*G. swidsinskii*	-	-	-
NR021		*G. swidsinskii*	-	-	-
VN003		*G. leopoldii*	-	-	-
WP021		*G. swidsinskii*	-	-	-
WP022		*G. swidsinskii*	-	-	-
NR010		*G. leopoldii*	-	-	-
N072		*G. swidsinskii*	-	-	-
GH021	C	*G. vaginalis*	-	-	-
NR001		*G. vaginalis*	-	-	-
NR037		*G. vaginalis*	-	-	-
NR038		*G. vaginalis*	-	-	-
NR039		*G. vaginalis*	-	-	-
WP023		*G. vaginalis*	-	-	-
N165		*G. vaginalis*	-	-	-
GH015		*G. vaginalis*	-	-	-
NR003	D	Genome sp. 8	-	-	-
NR047		Unknown	-	-	-
WP012		Genome sp. 9	-	-	-
N160		Genome sp. 10	-	-	-

429 [1]Previously determined (12)
430 [2]PCR positive for *nanH3*
431

432

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

433

434 **Table 2.** Homopolymer length variants in *Gardnerella piotii* and genome sp. 3 isolates.

Isolate	Homopolymer sequence	Predicted peptide length (amino acids)[1]
W11	CAACTA- C_{11}-ATGAACAAA	43
	CAACTA- C_{12}-ATGAACAAA	812
	CAACTA- C_{13}-ATGAACAAA	38
	CAACTA- C_{14}-ATGAACAAA	44
VN014	CAACTA- C_{9}-TCGAACAAA	812
	CAACTA- C_{10}-TCGAACAAA	37
	CAACTA- C_{11}-TCGAACAAA	43
VN015	CAACTA- C_{9}-TCGAACAAA	812
	CAACTA- C_{10}-TCGAACAAA	37
	CAACTA- C_{11}-TCGAACAAA	43
	CAACTA- C_{12}-TCGAACAAA	812
	CAACTA- C_{13}- - -GAACAAA	43
NR032	CAACTA- C_{8}-ATGAACAAA	42
	CAACTA- C_{9}-ATGAACAAA	812

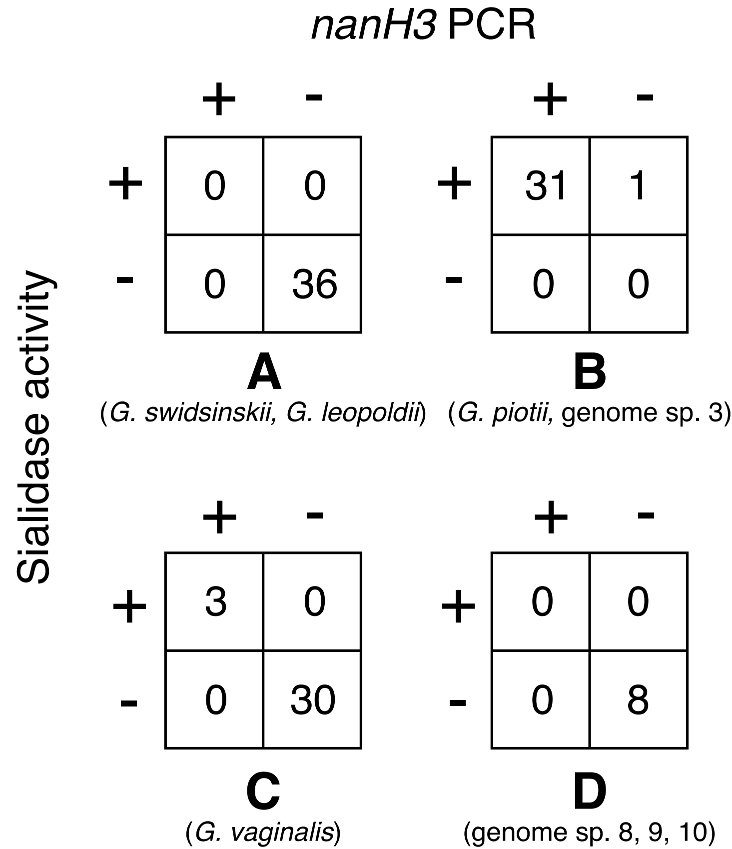
435

436

437

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

438



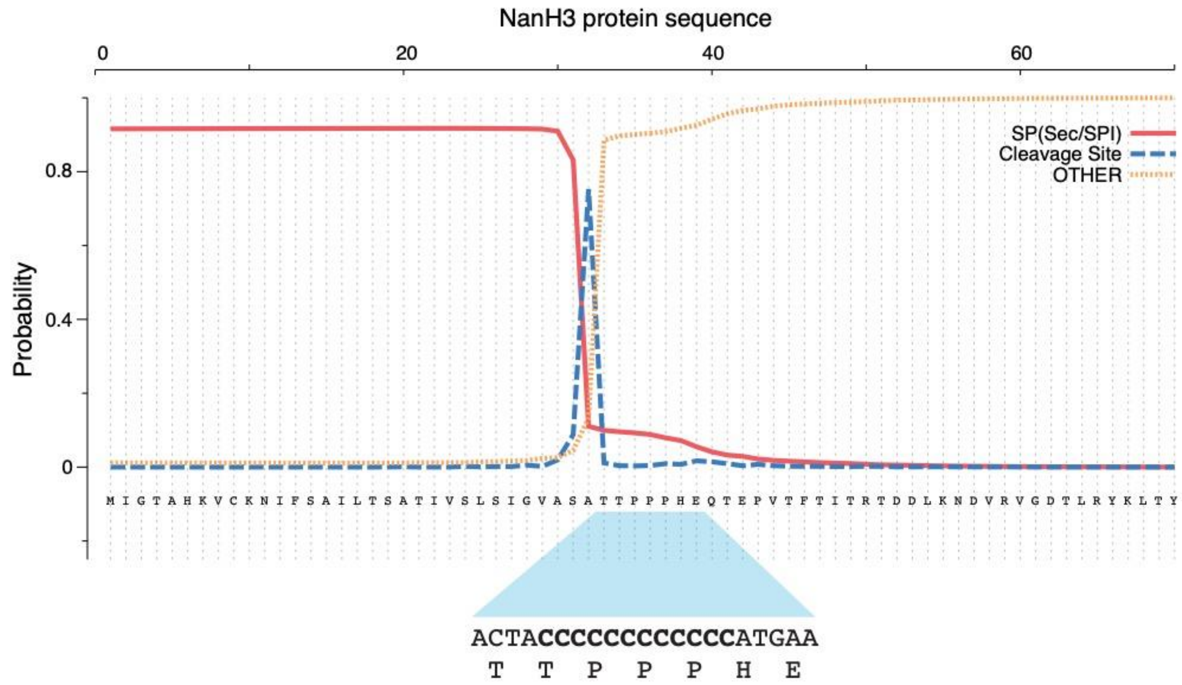
439

440 **Figure 1.** Correlation of *nanH3* and sialidase activity in 112 *Gardnerella* isolates. Numbers of
441 isolates positive or negative for sialidases activity and *nanH3* PCR. cpn60 subgroup affiliation
442 and species are indicated below each crosstab. Sialidase results were determined by Schellenberg
443 et al. (12).

444

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

445



446

447 **Figure 2.** Location of predicted signal peptide, cleavage site and homopolymer in the *nanH3*

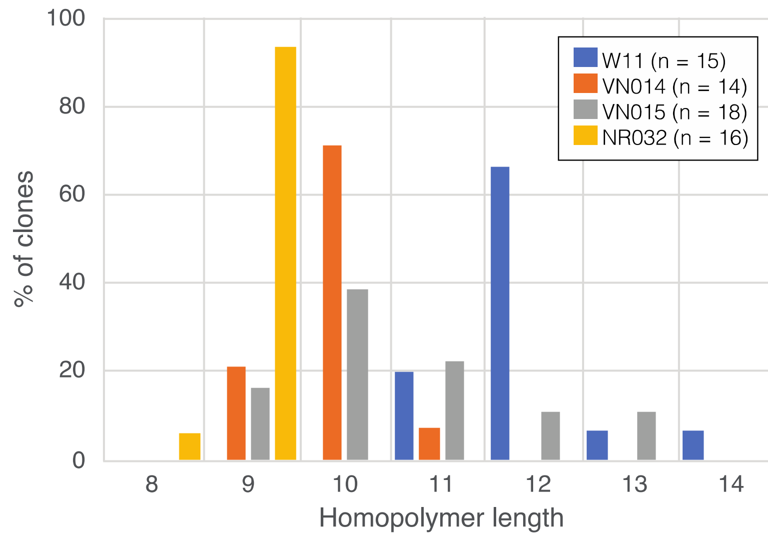
448 gene of strain W11. Analysis was conducted with SignalP-5.0.

449

450

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

451



452

453 **Figure 3.** Frequencies of the homopolymer length variants in isolates W11, VN014, VN015 and

454 NR032.

455

456

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.300947; this version posted September 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

457 **Supplemental Materials**

458

459 **Table S1.** *Gardnerella* strains used in the study, cpn60 subgroup affiliations, whole genome
460 sequence accession numbers, and sialidase activity phenotypes.

461

462 **Figure S1.** Sequencing of PCR products amplified directly from colonies of isolate W11.
463 PCR was performed with primers flanking the homopolymer region as described in the methods.
464 In each electropherogram, after eleven peaks of cytosine, the sequencing signal shows multiple
465 overlapping peaks suggesting the presence of multiple template sequences with different
466 homopolymer lengths.

467

468 **Figure S2.** Electropherograms of the homopolymer regions in two W11 colonies. Colony 6 had
469 individuals with 11, 12 and 13 C's in the homopolymeric region. Five out of eight colonies had
470 12C's in the homopolymer region. Colony 8 had individuals with 11, 12 and 14 C's in the
471 homopolymeric region. Five out of eight colonies had 12 C's in the homopolymer region.
472 Having 12 C's makes the coding region in-frame. When there is 11,13 or 14 C's, a premature
473 STOP codon is generated possibly making the protein truncated.

474

475