Intrapartum antibiotic prophylaxis selects for mutators in group B streptococci among persistently colonized patients

Macy E. Pell[1], Heather M. Blankenship[2], Jennifer A. Gaddy[3,4,5],

H. Dele Davies[6], Shannon D. Manning[1]*

[1]Michigan State University, Department of Microbiology, Genetics, and Immunology (MGI), E. Lansing, MI;

[2]Michigan Department of Health and Human Services, Bureau of Laboratories, Division of Infectious Diseases, Lansing, MI;

[3]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN;

[4]Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN;

[5]Tennessee Valley Healthcare Systems, Department of Veterans Affairs, Nashville, TN; and

[6]University of Nebraska Medical Center, Omaha, NE.

*Materials and correspondence to mannin71@msu.edu

ABSTRACT

Through vaginal colonization, GBS causes severe pregnancy outcomes including neonatal sepsis and meningitis. Although intrapartum antibiotic prophylaxis (IAP) has reduced early-onset disease rates, persistent GBS colonization has been observed in patients following prophylaxis. To determine whether IAP selects for genomic signatures that enhance GBS survival and persistence in the vaginal tract, whole-genome sequencing was performed on 97 isolates from 58 patients before (prenatal) and after (postpartum) IAP/childbirth. Core-gene mutation analysis identified 7,025 mutations between the paired isolates. Three postpartum isolates accounted for 98% of mutations and were classified as “mutators” because of point mutations within DNA repair systems. *In vitro* assays revealed stronger biofilms in two mutators. These findings suggest that antibiotics select for mutations that promote survival *in vivo*, which increases the likelihood of transmission to neonates. They also demonstrate how mutators can provide a reservoir of beneficial mutations that enhance fitness and genetic diversity in the GBS population.

16 **INTRODUCTION**

17

18 As an opportunistic bacterial pathogen, Group B *Streptococcus* (GBS) colonizes ~35% of
19 pregnant patients[1] and colonization is the main risk factor for invasive neonatal infection[2,3].
20 During pregnancy, rectovaginal colonization can also lead to sepsis of the pregnant person and
21 other adverse outcomes including preterm birth and stillbirth[4,5]. GBS disease in neonates occurs
22 as early-onset disease (EOD), which presents as sepsis and/or pneumonia within the first week
23 of life, or late-onset disease (LOD) that presents as bacteremia and/or meningitis between 1
24 week and 3 months after birth[6].

25

26 In the U.S., screening for vaginal-rectal colonization is recommended at 35-37 weeks'
27 gestation[7]. To prevent neonatal infections, GBS-positive patients are given intrapartum antibiotic
28 prophylaxis (IAP) during labor, for which penicillin is the most effective[3]. While IAP has
29 successfully reduced the incidence of EOD, it has not impacted LOD rates and is not helpful for
30 preventing preterm births or stillbirths[8]. It can also negatively impact the microbiome[9,10].
31 Furthermore, GBS colonization is intermittent during pregnancy and colonization frequencies
32 vary across locations[11,12]. One study of EOD cases, for example, showed that 60% were born to
33 individuals who were GBS-negative at the prenatal screening[13], thereby highlighting the need
34 for improved treatment and prevention methods.

35

36 Classifying GBS based on the polysaccharide capsule (cps), which dictates the serotype, and
37 the multilocus sequence type (ST), has identified associations between molecular traits and
38 disease severity[14,15]. Capsule type III ST-17 strains predominate among neonates with invasive
39 disease[14–16] and possess unique features that help withstand oxidative stress, survive inside
40 macrophages, associate with host cells, and elicit proinflammatory responses relative to other
41 STs[17–21]. Less definitive epidemiological associations have been observed for other lineages

42 (e.g., STs 19 and 23) despite their high prevalence[15]. The effectiveness of IAP was also found
43 to vary across STs in our prior study[22], with STs 17 and 19 more commonly colonizing patients
44 before IAP and persisting up to 6 weeks postpartum than other STs. ST-12 strains, however,
45 were more frequently lost after IAP, which was also independently associated with persistent
46 colonization. Despite these associations, the mechanisms that promote GBS persistence in the
47 vaginal tract and evasion of antibiotic-mediated killing remain elusive. This is a serious concern
48 as persistent colonization enhances the risk of invasive disease, especially for subsequent
49 pregnancies and babies after IAP cessation[10,22], and may partly explain why rates of GBS LOD
50 have not decreased in the U.S.[8,10].
51
52 Altogether, these findings led us to hypothesize that IAP imposes a selective pressure on GBS,
53 such that genomic features that promote colonization are selected for to enhance fitness. We
54 used whole-genome sequencing (WGS) to compare GBS isolates recovered from pregnant
55 patients sampled before (prenatal) and after (postpartum) IAP[23]. This study has enhanced
56 understanding of how IAP impacts the colonizing GBS population, concurrently providing insight
57 for treatment and prevention improvements.
58
59
60
61 **RESULTS**
62
63 **GBS colonization dynamics and antibiotic exposures differ across patients**
64 Isolates (n=97) were collected from 58 pregnant patients in a prior study[23]. Forty-one (70.7%)
65 patients had GBS at the prenatal and postpartum samplings indicating persistent colonization
66 despite IAP. Two postpartum strains from two persistently colonized patients (IDs 11 and 14)
67 were not available for characterization. The remaining 17 patients were colonized only at the

68 prenatal sampling indicating loss of GBS colonization. Most (n=53; 91.4%) patients received
69 IAP with ampicillin (n=26), penicillin (n=17), cefazolin (n=2), or clindamycin (n=6). One patient
70 received both ampicillin and clindamycin, and another had penicillin and metronidazole (Flagyl).
71 Five (8.6%) patients did not receive IAP, including three with persistent colonization and two
72 with prenatal colonization only; four of these patients had neonates with signs of infection.
73 Moreover, eight (13.8%) patients received postpartum antibiotics, four of whom had persistent
74 colonization.
75
76 As determined previously[22], the 97 GBS strains represent 20 STs comprising clonal complex
77 (CC)-1 (n=25), CC-12 (n=15), CC-17 (n=16), CC-19 (n=17), CC-22 (n=2), CC-23 (n=20), and
78 CC-26 (n=1). One strain represented a singleton (ST-67). Eight distinct cps types were
79 represented, with cps III (n=27) predominating followed by Ia (n=21) and V (n=17); four isolates
80 were non-typeable (NT).
81
82 **Pangenome analysis reveals a large accessory genome and clustering by CCs**
83 Ninety-two high-quality genomes were included in this analysis; three prenatal and two
84 postpartum genomes from five patients were excluded due to poor assembly quality. Among the
85 92 genomes, 68 represented paired strains from 34 patients with persistent colonization
86 recovered from both the prenatal and postpartum samplings in the same patients. Additionally,
87 seven unpaired strains from seven persistently colonized patients collected at the prenatal (n=4)
88 or postpartum (n=3) sampling were included plus the 17 strains from patients who lost GBS
89 following IAP and were only colonized at the prenatal sampling.
90
91 The GBS pangenome consists of 5,054 unique genes comprising a small core-genome with
92 1,368 core and 213 soft-core genes. It has a large accessory genome, which contained 945
93 shell and 2,528 cloud genes accounting for 69% of the total genes. A maximum likelihood (ML)

94 tree based on the 1,368 core genes resolved five sequence clusters grouping together with
95 >95% bootstrap support (**Figure 1**). Each cluster was associated with a CC. The 17 CC-19
96 strains and 16 CC-17 strains grouped into two separate clusters with 100% bootstrap support,
97 while the CC-1 (n=23) and CC-12 (n=23) strains were part of two separate clusters with 99%
98 and 95% bootstrap support, respectively. Although the 19 CC-23 strains clustered together, they
99 comprised a mixture of outgroups and smaller clades with 2-5 genomes each. A core-gene
100 neighbor-net tree confirmed these groupings (**Figure S1**) and detected evidence for
101 recombination (pairwise homoplasy index (PHI) p-value = 0.00) across the 17,825 parsimony-
102 informative (PI) sites.
103
104 Among the 34 pairs of strains from patients with persistent colonization, 29 (85.3%) clustered
105 together with >91% bootstrap support and five did not. Both strains from two of these five
106 patients (IDs 3 and 8) had different STs and were on different branches of the tree, suggesting
107 acquisition of a new strain type following childbirth and IAP. For the remaining three patients
108 (IDs 22, 43, and 56), the prenatal and postpartum strains shared a ST but did not cluster
109 together in the ML tree. This finding suggests that the phylogenetic distance between these
110 paired strains is due to nucleotide-level variation within the core gene sequences.
111

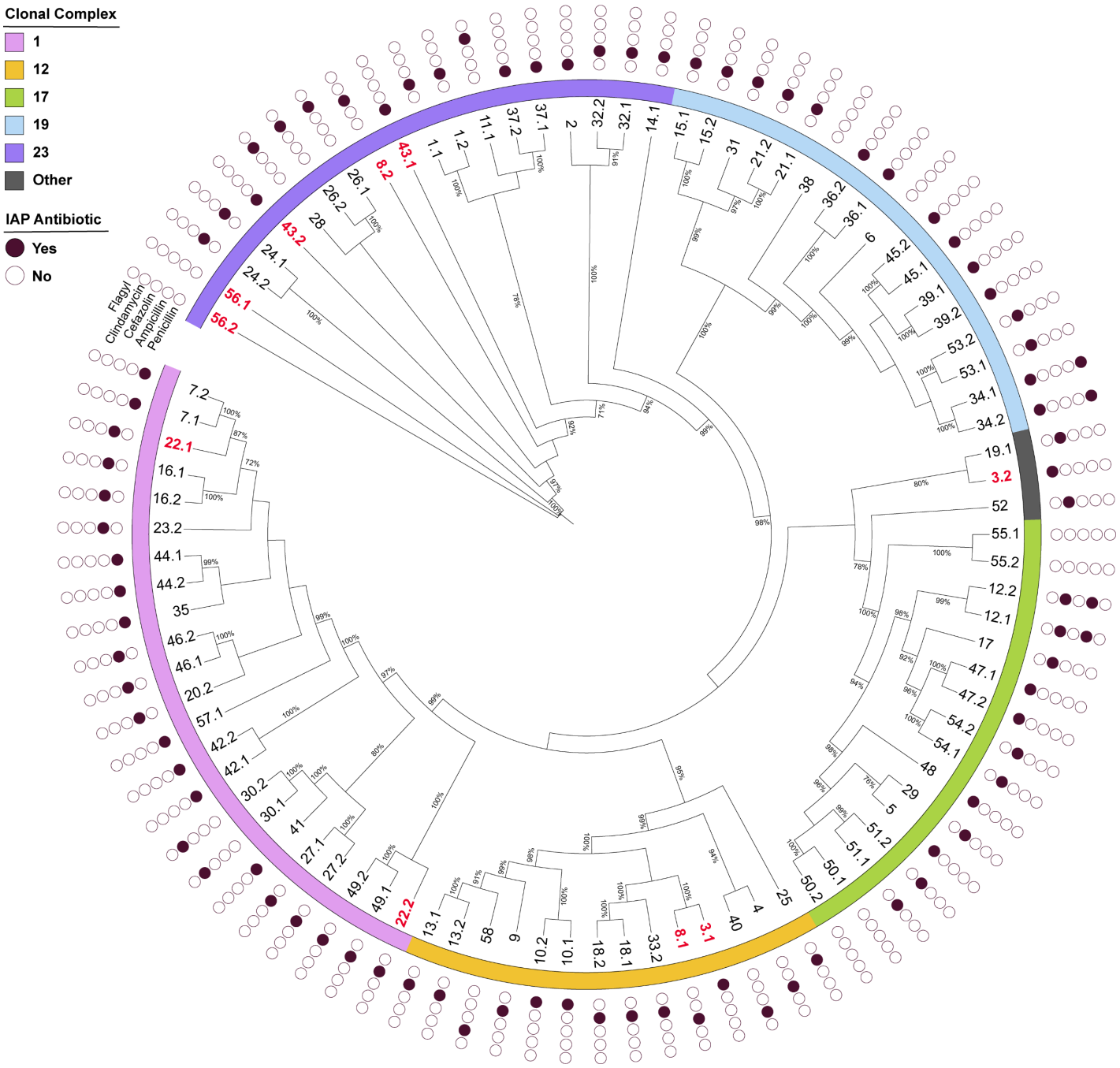


Figure 1: Maximum likelihood tree of 92 GBS genomes based on 1,368 core genes. Patient strain IDs are noted at the end of each branch such that persistent strains contain a ".1" (prenatal) or ".2" (postpartum) to indicate the sampling timepoint with respect to IAP and childbirth. The clonal complex (CC) is represented by the colored strips at the outside of the tree and bootstrap support values >70% are noted on the tree branches. The maroon circles on the outside of the phylogeny show the antibiotic(s) used for IAP; the paired strains from the same patient have the same antibiotic usage profiles. IDs in red bold face font highlight the paired prenatal and postpartum strains from five patients (IDs 3, 8, 22, 43, and 56) that do not cluster together in the tree.

123 **Acquisition of antibiotic resistance genes is not important for persistent colonization**

124 To determine whether GBS residing in the vaginal tract acquired antibiotic resistance genes
125 (ARGs) to promote colonization, an assembly-based analysis was performed to extract these
126 genes from the 92 genomes. Nine distinct ARGs were identified conferring resistance to
127 tetracyclines (*tetL, tetM, tetO, tetW, lmrP*), macrolides-lincosamides-streptogramins (MLS)
128 (*ermA, lmrP, mreA*), fluoroquinolones (*norB*), and cationic peptides (*mprF*). Both *lmrP* and *mprF*
129 were found in all strains, while *norB* and *mreA* were found in all but one (**Figure 2**).

130

131 Among the variably present ARGs, *tetM* predominated and was found in 63 (68.5%) genomes
132 including 47 of the 68 (69.1%) paired strains from the 34 patients. A similar proportion (n=11;
133 64.7%) was detected in the 17 prenatal strains that were lost following IAP. Although the
134 prenatal and postpartum strains from patient 22 had the same ST, *tetM* was only detected in the
135 prenatal strain (22.1), indicating gene loss. This result is not surprising since *tetM* is often
136 associated with plasmids and the two strains did not cluster together in the core-gene ML tree
137 (**Figure 1**). The same was true for patient 3, though this finding is less surprising in that the two
138 paired strains had different STs, indicating acquisition of a new strain by the postpartum visit.

139

140 **Specific virulence gene profiles are not linked to persistent colonization**

141 An assembly-based analysis was also performed to identify virulence genes in the 92 genomes;
142 50 distinct genes were detected and classified based on function. Genes important for
143 adherence and invasion of host cells (n=18, 36%), immune modulation (n=17, 34%),
144 metabolism (n=1, 2%), toxin production (n=13, 26%), and dissemination (n=1, 2%) were found.

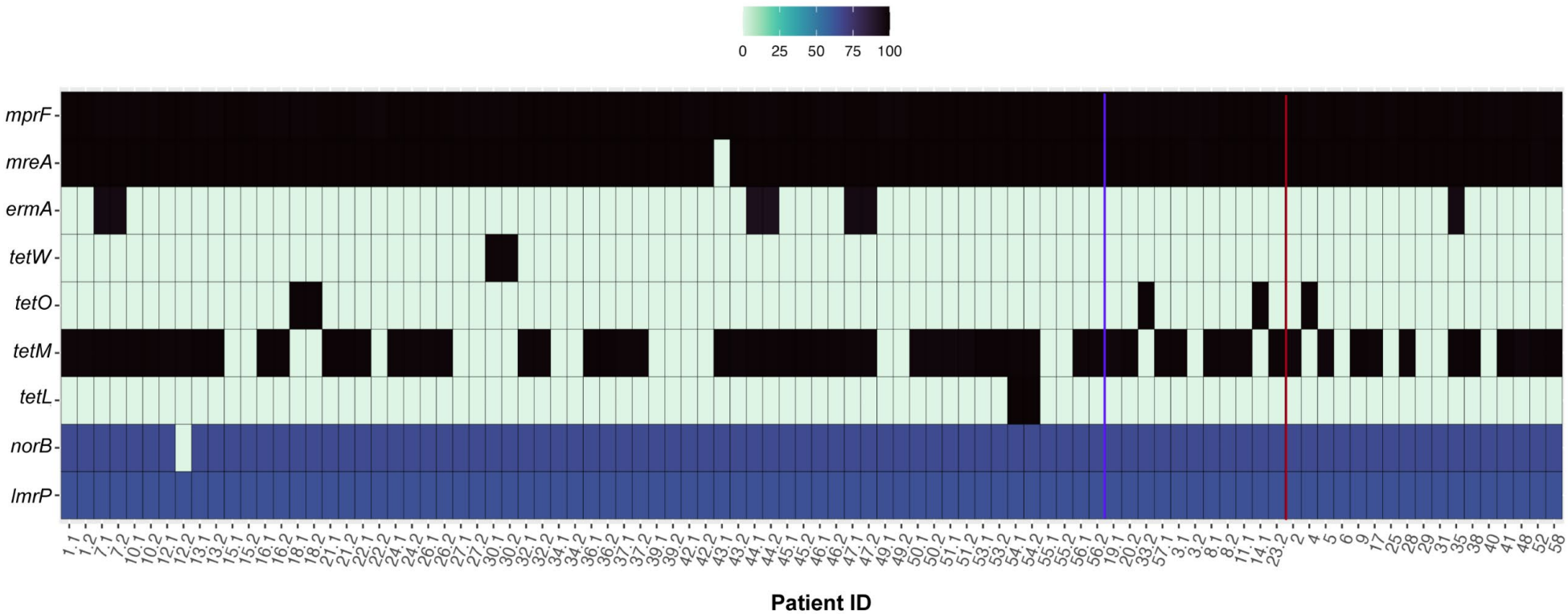


Figure 2: Antibiotic resistance genes (ARGs) detected in the 92 genomes. ARGs were extracted from the 92 genome assemblies using five different ARG databases. The x-axis shows the patient strains with a ".1" (prenatal) and ".2" (postpartum) after the "Patient ID" indicating the sampling timepoint. The gene profiles for the paired isolates from the same 32 patients with isolates of the same sequence type (ST) are shown next to each other on the left side of the vertical purple line. The two sets of paired strains (IDs 3 and 8) with different STs along with seven unpaired strains are shown between the purple and red vertical lines. The ARG profiles for the 17 "lost" strains, which were recovered only at the prenatal sampling and did not persist, are to the right of the red vertical line. Gene names are on the y-axis. The color gradient shows the average percent identity of a gene across the five databases, with black representing 100% identity and mint green representing <10% identity. ARGs identified in this analysis confer resistance to tetracyclines (*tetL, tetM, tetO, tetW, lmrP*), macrolides-lincosamides-streptogramins (*ermA, lmrP, mreA*), fluoroquinolones (*norB*), and cationic peptides (*mprF*).

155 The greatest variation in virulence genes was observed for those encoding capsule (*cps*)
156 production and pili (**Figure 3**). Among the 32 persistently colonized patients with strains of the
157 same ST, several discrepancies in gene presence were observed. Most of these differences,
158 however, involved genes that fell below the identity and coverage thresholds, indicating they are
159 present but may be distinct. Patients 3 and 8 with prenatal and postpartum strains belonging to
160 different STs, also had distinct virulence gene profiles, indicating acquisition of a new strain.
161
162 To further confirm that the prenatal and postpartum strain populations had distinct virulence
163 gene profiles, a principal component analysis was performed. No distinct clustering patterns
164 were observed for the prenatal versus postpartum strains (**Figure S2A**) or after stratifying by
165 antibiotic, cps type, and colonization phenotype (persistent or lost) (**Figure S2B-E**). Additionally,
166 no difference was observed between colonization phenotype and the number of adherence
167 (n=1,031, Chi-square p = 0.10) or immune modulatory (n=1,231, Chi-square p=0.52) genes, as
168 determined by coverage and nucleotide identity thresholds. These data suggest that possession
169 of specific combinations of virulence genes is not linked to a strain's ability to persistently
170 colonize the vaginal tract after IAP.
171
172 **Point mutations in persistent strains highlight microevolution after IAP**
173 To detect nucleotide-level variation, we conducted a reads-based analysis of the 1,368 core
174 genes in the 64 paired prenatal and postpartum strains that shared STs and were recovered
175 from the 32 patients with persistent colonization. In this analysis, each prenatal genome was
176 used as the reference genome for its paired postpartum genome and the number and type of
177 mutations in the postpartum genomes was determined. Notably, 7,025 mutations were identified
178 in 24 (75.0%) of the 32 postpartum genomes. No mutations were detected in the remaining
179 eight postpartum genomes, which had identical core gene sequences as the paired prenatal
180 genomes from the same patients.

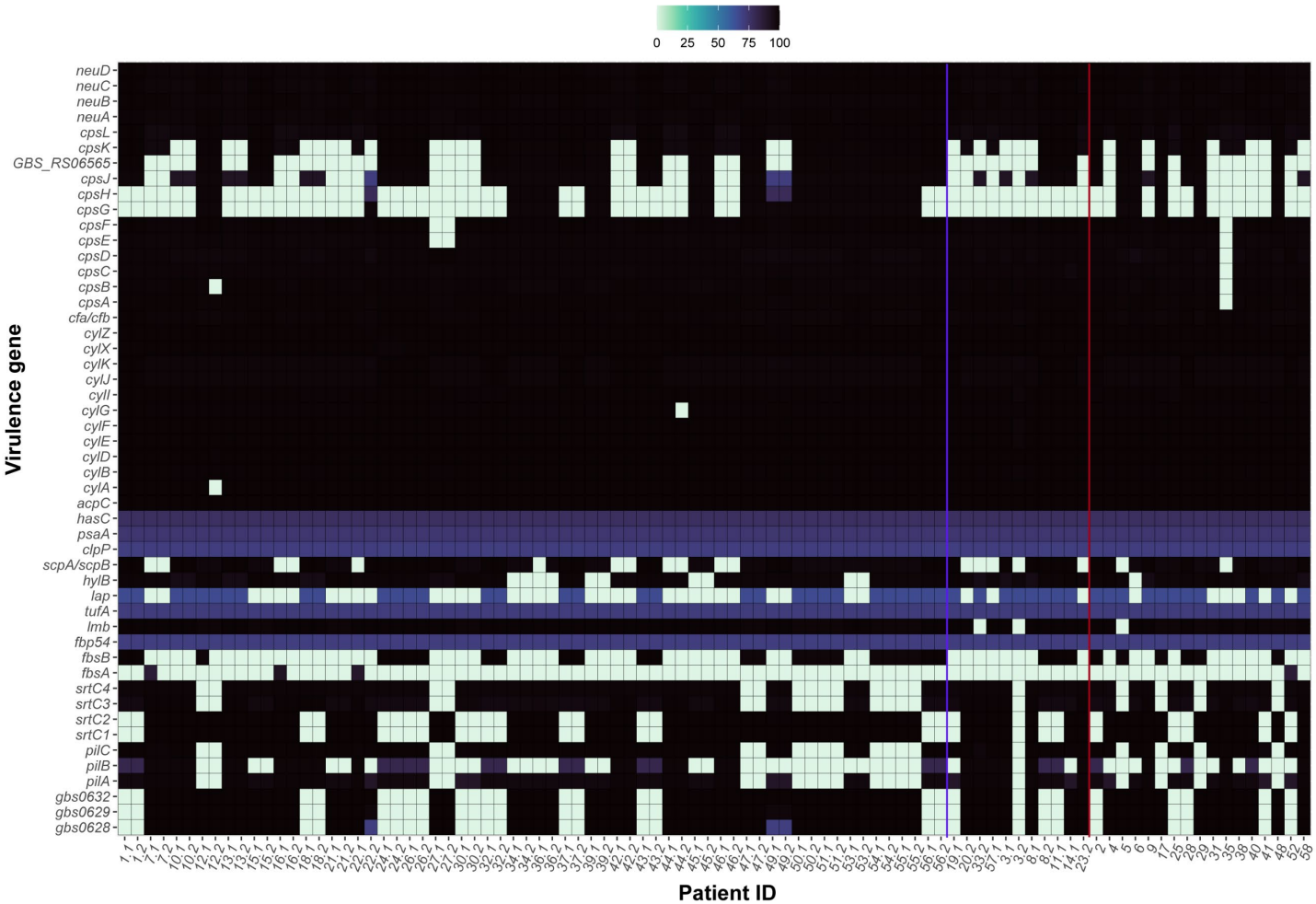
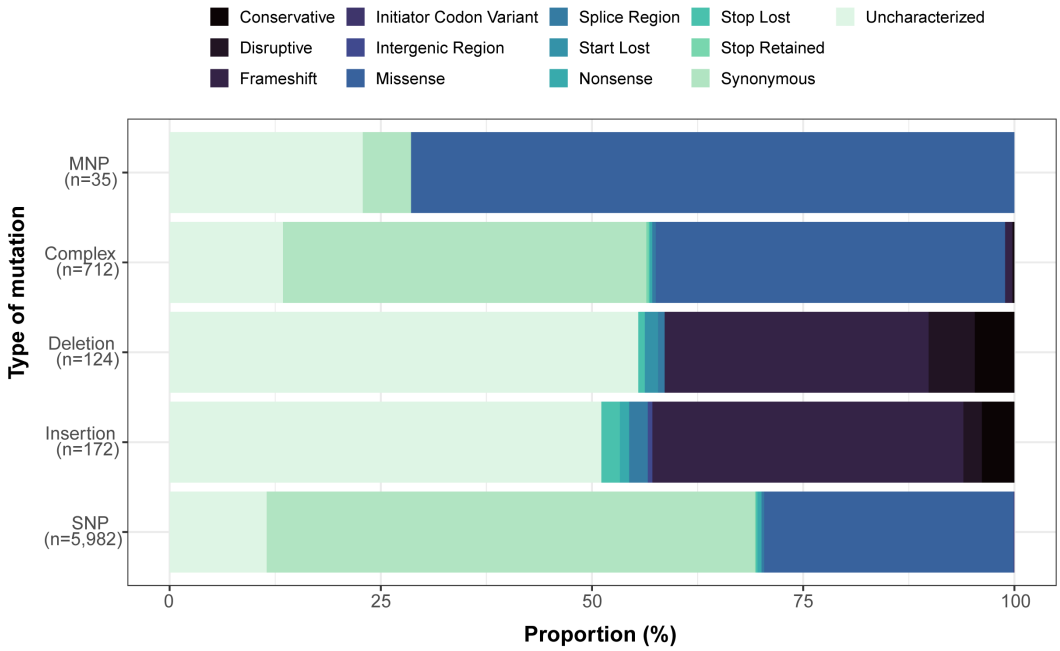


Figure 3: Virulence genes detected in 92 GBS genomes. The x-axis shows the patient ID with the ".1" (prenatal) and ".2" (post-partum) indicating the sampling timepoint. Gene names are displayed on the y-axis. The gene profiles for the paired isolates from 32 patients with isolates of the same sequence type (ST) are next to each other, left of the purple line. The seven unpaired strains plus the two paired strains with different STs from patients 3 and 8 are shown between the purple and red vertical lines. Profiles for the 17 "lost" strains are to the right of the red line. The color gradient displays the average percent identity of a gene, with black representing 100% identity and mint green representing <10% identity.

Among the 24 genomes with ≥1 mutation, most mutations (n=6,065; 86.3%) were in coding regions, though an additional 958 (13.6%) were in uncharacterized regions and two were in rRNAs. The type of mutations varied across genomes (**Figure S3A**). Most (n=5,982; 85.2%) were single nucleotide polymorphisms (SNPs) including synonymous (n=3,469; 58.0%), missense (n=1,771; 29.6%), or nonsense (n=26; 0.4%) mutations (**Figure 4**). Complex

196 mutations (n=712), defined as a combination of multiple nucleotide polymorphisms (MNPs) and
197 an insertion or deletion (indels), were also detected and classified as synonymous (n=311;
198 43.7%) and missense (n=299; 41.9%) mutations. Among the indels (172 insertions and 124
199 deletions), frameshift mutations (n=107; 36.1%) were most common. Similar frequencies of
200 conservative outcomes (n=6), defined as an indel of one or more entire codons, and disruptive
201 outcomes (n=7), defined as a codon change plus a codon indel, were among the deletions.
202 Another deletion resulted in the loss of a stop codon, while two others resulted in the loss of a
203 start codon. For the insertions, more were conservative (n=7) than disruptive (n=4) mutations
204 with two resulting in nonsense mutations and four in the loss of a stop codon. Among the 35
205 MNPs detected, most were missense mutations (n=25; 71.4%).

206



226 **Figure 4: Core gene mutations detected in the postpartum genomes of GBS strains from**
227 **patients following IAP.** Five types of mutations (y-axis) were detected in 24 of the 32
228 postpartum genomes when each genome was compared to its respective prenatal genome from
229 the same patient. No mutations were identified in eight of the 32 postpartum genomes. Types of
230 mutations identified include multiple nucleotide polymorphisms (MNP), complex mutations,
231 deletions, insertions, and single nucleotide polymorphisms (SNP). The raw mutation count is
232 noted in parentheses for each type of mutation, while the mutation outcomes are distinguished
233 by color and are shown as a proportion of each mutation type (x-axis). The categories are not
234 mutually exclusive as some mutations were classified as multiple types (i.e., an insertion
235 classified as disruptive and as a frameshift will be included in both mutation categories).

236 **A skewed distribution of mutations reveals mutator strains in GBS**

237 Comparing mutation frequencies across the postpartum genomes revealed a skewed
238 distribution with three outliers (**Figure S3B**). These outlier genomes contained 204, 262, and
239 6,454 mutations relative to an average of five mutations and range of 1-16 mutations for the 21
240 postpartum genomes with at least one mutation (**Figure S3C**). Based on the total assembly
241 length, we calculated the mutation rate, or the number of mutations per number of base pairs
242 (bp) for each postpartum genome relative to its respective prenatal genome. The rates for the
243 three outlier genomes are 9.37×10^{-5} (patient genome ID=36.2), 1.32×10^{-4} (ID=43.2) and 0.003
244 (ID=22.2). These rates were 53-, 75-, and 1,747-fold higher, respectively, than the average
245 mutation rate (1.77×10^{-6} mutations/bp) for the remaining 29 genomes, including the eight
246 genomes lacking any detectable mutations.

247

248 Because mutators arise at a low frequency in the population due to mutations within genes
249 encoding DNA replication and proofreading machinery, mismatch repair (MMR) pathways, and
250 the 8-oxo-dG (GO) system[24,25], we examined these genes in the three outlier postpartum
251 genomes. Genome ID 22.2, which contains the most mutations, also has the greatest number of
252 nonsynonymous mutations in genes encoding DNA repair systems including the MMR (n=14)
253 and GO (n=8) pathways, and in genes involved in DNA replication (n=18), DNA repair (n=7),
254 and DNA recombination (n=9). Similarly, genome ID 36.2 has a deletion in the gene encoding
255 DNA polymerase IV, a missense SNP in *mutY* of the GO system, and a 33 bp insertion within
256 an MMR locus, while genome ID 43.2 has a missense SNP in the DNA topoisomerase gene
257 and a 1 bp deletion downstream of the MMR loci. The increased mutation rates along with
258 detectable mutations in these key regions meet the criteria[24,26] for classifying these three strains
259 as “mutators” that arose after IAP.

260

261 **Distinct classes of MMR loci exist but none are associated with hypermutability**

262 Since mutations in MMR loci were identified in all three mutators, we examined the six MMR
263 pathway genes (7,467 bp region) in the 64 paired genomes from the 32 patients. These genes
264 encode DNA mismatch repair proteins, MutS (*mutS*) and MutL (GBS_230), a cold shock protein,
265 a MFS transporter/LmrP homolog (GBS_436), the Holliday junction branch migration protein
266 (RuvA), and DNA-3-methyladenine glycosylase I (GBS_235). A multiple sequence alignment
267 using MUSCLE[27] detected only 104 variable nucleotides.

268

269 Construction of a neighbor joining (NJ) phylogeny based on the MMR gene sequences (7,467
270 bp) revealed four clusters, which we arbitrarily named Class 1-4 (**Figure 5**); three of these
271 classes had subclasses grouping with >60% bootstrap support. Class 4 predominated (n=28,
272 43.7%) and contains three subclasses (4A-4C), followed by Class 2 (n=21, 32.8%) with
273 subclasses 2A-2E. Class 1 (n=12, 18.7%) has two subclasses, while Class 3 has none.
274 Although a neighbor-net tree based on 103 PI sites within the 7,467 bp MMR region provided no
275 evidence of recombination (PHI p-value=0.0899), the four distinct MMR classes are still evident
276 (**Figure S4**).

277

278 The MMR loci classes and subclasses are associated with specific CCs. All CC-17 strains, for
279 instance, possess Class 1 MMR loci, with subclasses 1A and 1B clustering separately in the
280 core gene phylogeny (**Figure S5**). By contrast, the Class 2 and Class 4 subclasses are split
281 between CCs, with subclass 4A being found exclusively within CC-23 strains and 4B and 4C
282 clustering separately within the CC-19 strains. The subclasses of the Class 2 MMR loci did not
283 group together within the phylogeny, as 2B-2E were found exclusively in the CC-1 or CC-12
284 strains and subclass 2A strains were unrelated and split between the two CCs.

285

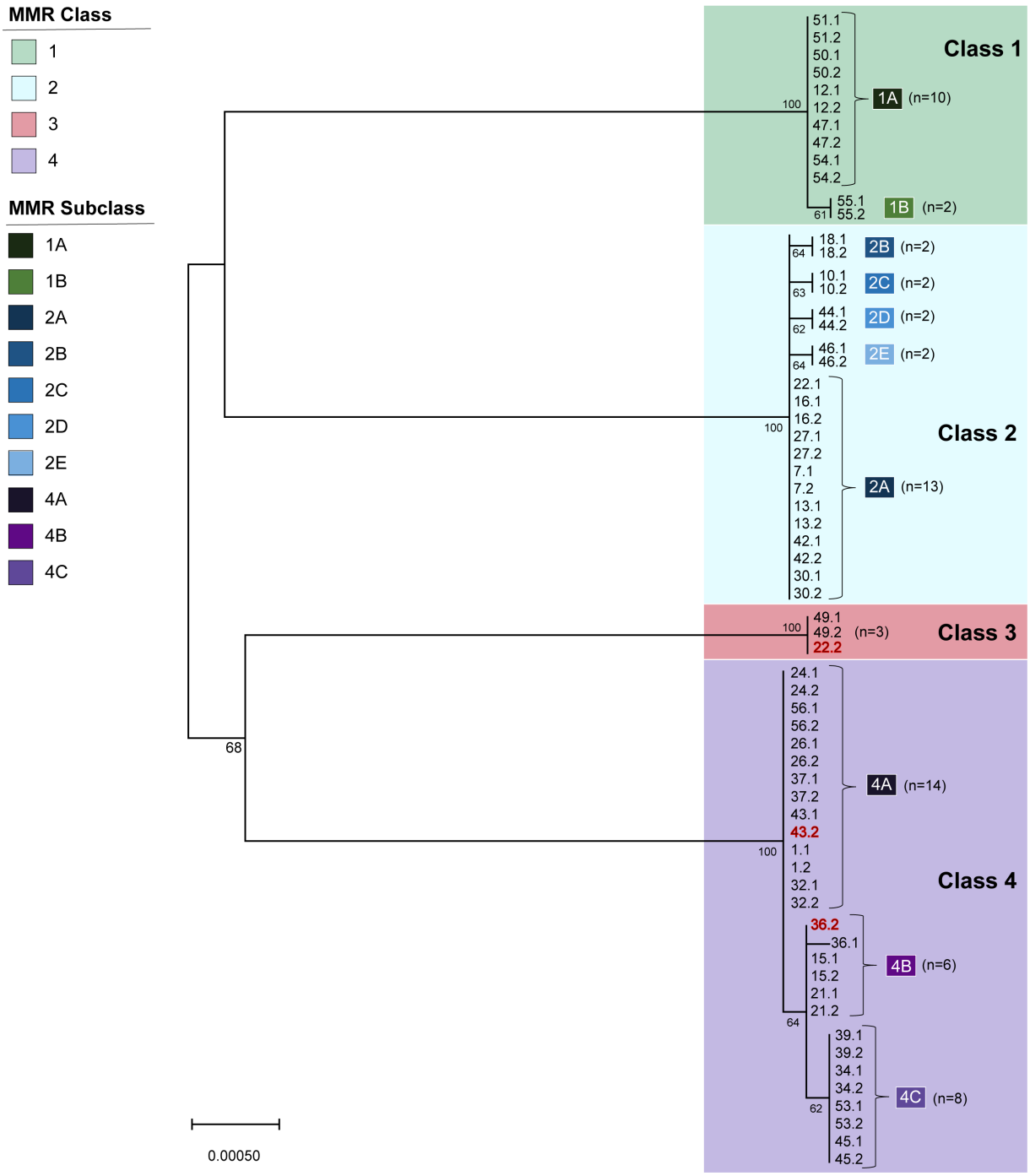


Figure 5: Neighbor joining phylogeny based on the MMR loci (7,467 bp) extracted from 32 pairs of strains from pregnant patients collected before and after IAP/childbirth. The evolutionary distances are shown as the number of base substitutions per site. Four sequence clusters (Classes 1-4) were detected using the bootstrap test with 1,000 replicates. Subclasses were identified with three classes, which grouped with >60% bootstrap support; the number in parentheses indicates the number of the genomes with the same subclass (100% identity) among the 64 genomes. Bootstrap values are shown on the tree branches and the three mutator strains are in red, bolded text.

294 A comparison of the strain pairs (prenatal versus postpartum) from the same patients revealed
295 that the MMR classes and subclasses were identical in all but two pairs. The two postpartum
296 mutator strains (IDs 22.2 and 36.2) differed from the prenatal genomes (IDs 22.1 and 36.1) by
297 56 nucleotides (21 amino acids) and 1 nucleotide (1 amino acid), respectively. Although two of
298 the three mutators belong to MMR Class 4, both had distinct MMR subclasses representing 4A
299 (genome ID 43.2) and 4B (genome ID 36.2). The third mutator (genome ID 22.2) belonged to
300 Class 3 (**Figure 5**), suggesting that a single MMR allele is not associated with hypermutability.
301
302 To determine how common each MMR class is among GBS genomes and identify novel
303 classes, a Basic Local Alignment Search Tool (BLAST)[28] analysis was performed on GBS
304 genomes available in the National Center for Biotechnology Information (NCBI; access date
305 3/21/2023). Among ~100 GBS genomes with sequence matches, MMR subclass 2A
306 predominated as 43 genomes were identified with 100% nucleotide identity. Five MMR loci
307 subclasses (1B, 2B, 2C, 2D, and 4C), however, were unique to our strains, as no published
308 GBS genomes had matching sequences. The Class 3 loci, which was found in mutator strain
309 22.2, returned three GBS genomes with 100% identity including strains B111 (CP021772.1),
310 SG-M8 (CP021868.1), and GBS-M002 (CP013908.1), all of which represent invasive, human-
311 derived strains from Asia. Similar to strain 22.2, each was classified as
312 serotype VI and belong to ST-1 using PubMLST[29].
313
314 **IAP elicits a selective pressure that differentially impacts lineages**
315 To better understand which point mutations could alter protein functionality, we further classified
316 the subset of 5,291 characterized SNPs out of the 5,982 SNPs identified in the 1,368 core
317 genes among the 32 postpartum strains from patients with persistent colonization. While
318 specific mutations were not associated with individual lineages, CCs 1 (n=4,953), 19 (n=121),
319 and 23 (n=199) had the greatest number of SNPs. These numbers, however, are skewed by

320 inclusion of the three mutator strains. Excluding the mutators revealed that the CC-23 strains
321 had the most SNPs (n=21) followed by CCs 12 (n=15) and 1 (n=12), averaging 3, 5, and 1
322 SNP(s) per strain, respectively, and ranging from 0-10, 0-9, and 0-7 SNPs per strain. By
323 comparison, the postpartum non-mutator strains of CCs 17 (n=3) and 19 (n=7) acquired fewer
324 SNPs overall, averaging between 0-1 SNPs per strain with a range of 0-3. Among the few SNPs
325 identified in CC-17 strains, only one was nonsynonymous, indicating a pattern of negative or
326 purifying selection for strains in this lineage. Comparatively, the non-mutator strains of other
327 lineages had more nonsynonymous (n=38) than synonymous (n=17) SNPs, suggesting a
328 pattern of positive/diversifying selection and a trend of adaptive evolution for these lineages.
329
330 Based on prior medical record reviews[23], the three mutator strains had been exposed to either
331 penicillin (patient IDs 36 and 43) or ampicillin (ID 22) for IAP. Both antibiotics contributed to an
332 average of 292 and 31 SNPs in the postpartum strains from patients given ampicillin (n=17) or
333 penicillin (n=10), respectively. After excluding the mutators, however, the average decreased to
334 two SNPs/strain for both drugs. While comparisons could not be made for the other antibiotics
335 due to low numbers, the three strains from patients who did not receive IAP, had an average of
336 only one SNP/strain. Collectively, these data demonstrate that the emergence of mutators can
337 impact mutation frequencies, and further, they suggest that certain strain types are more
338 susceptible to IAP-associated evolutionary pressures.
339
340 **Mutators have a high frequency of missense SNPs and genes under positive selection**
341 Since mutators can serve as a reservoir of beneficial mutations for the surrounding bacterial
342 population, we sought to further examine those SNPs identified within coding sequences in the
343 three mutator strains. Among the subset of 5,291 characterized SNPs, 5,233 (98.9%) were in
344 the three mutators and mapped to 1,167 genes, yielding an average of two SNPs/gene (**Figure**
345 **6A**). Comparatively, 58 (1.1%) SNPs were detected within 56 genes in the non-mutator strains,

346 resulting in an average of 1 SNP/gene (**Figure 6B**). Within the mutator genomes, most genes
347 (n=683) acquired ≤4 SNPs relative to the respective prenatal genomes with 390, 142, 86, and
348 65 genes having 1, 2, 3, and 4 SNPs, respectively. Moreover, >4 SNPs were detected within
349 366 genes; 26 of these genes had >20 SNPs and were important for DNA replication/repair,
350 antimicrobial resistance, phage proteins, and virulence. Among all 366 genes, most (77.6%,
351 n=284) had evidence of negative selection based on the ratio of nonsynonymous (dN) to
352 synonymous (dS) mutations. Another 61 genes (16.7%) were under positive selection, while the
353 remaining 21 (5.74%) genes had evidence for neutral selection.
354
355 Restricting our analysis to genes with >4 nonsynonymous SNPs, we identified 662 SNPs in 91
356 genes, with most (98.3%, n=651) representing missense mutations (**Figure 6C**). Eight genes
357 had strong signals of positive selection (dN/dS >4) and are important for oxidative stress
358 (*yaaA_2,* GBS_639), competition (GBS_534), translation (*asnS*), virulence (GBS_675), and
359 aminoglycoside resistance (GBS_318) Two hypothetical genes (GBS_724, GBS_776) were also
360 of interest (Table S9) along with *aspS* (tRNA ligase), *recX* (recombination regulator), and *esaA*
361 (type VII secretion protein). Because multiple virulence and stress response genes were
362 identified, it is likely that such processes play a role in persistent colonization following IAP.
363
364 Of the 651 missense SNPs identified in the mutator strains, several were located in the same
365 set of genes, ranging from 4-26 missense SNPs/gene (**Figure 6C**). Virulence genes like *esaA,*
366 *hylB* (hyaluronate lyase)*,* and *pilA* (PI-2a pilus adhesin), which are important for colonization,
367 acquired >10 missense SNPs/gene. Furthermore, 10 nonsense mutations were acquired in
368 eight genes including *asp3* (accessory Sec-system protein), *fbsA* (fibrinogen-binding adhesin),
369 and GBS_550 (pullulanase), which also had a mutation contributing to the loss of a start codon.
370

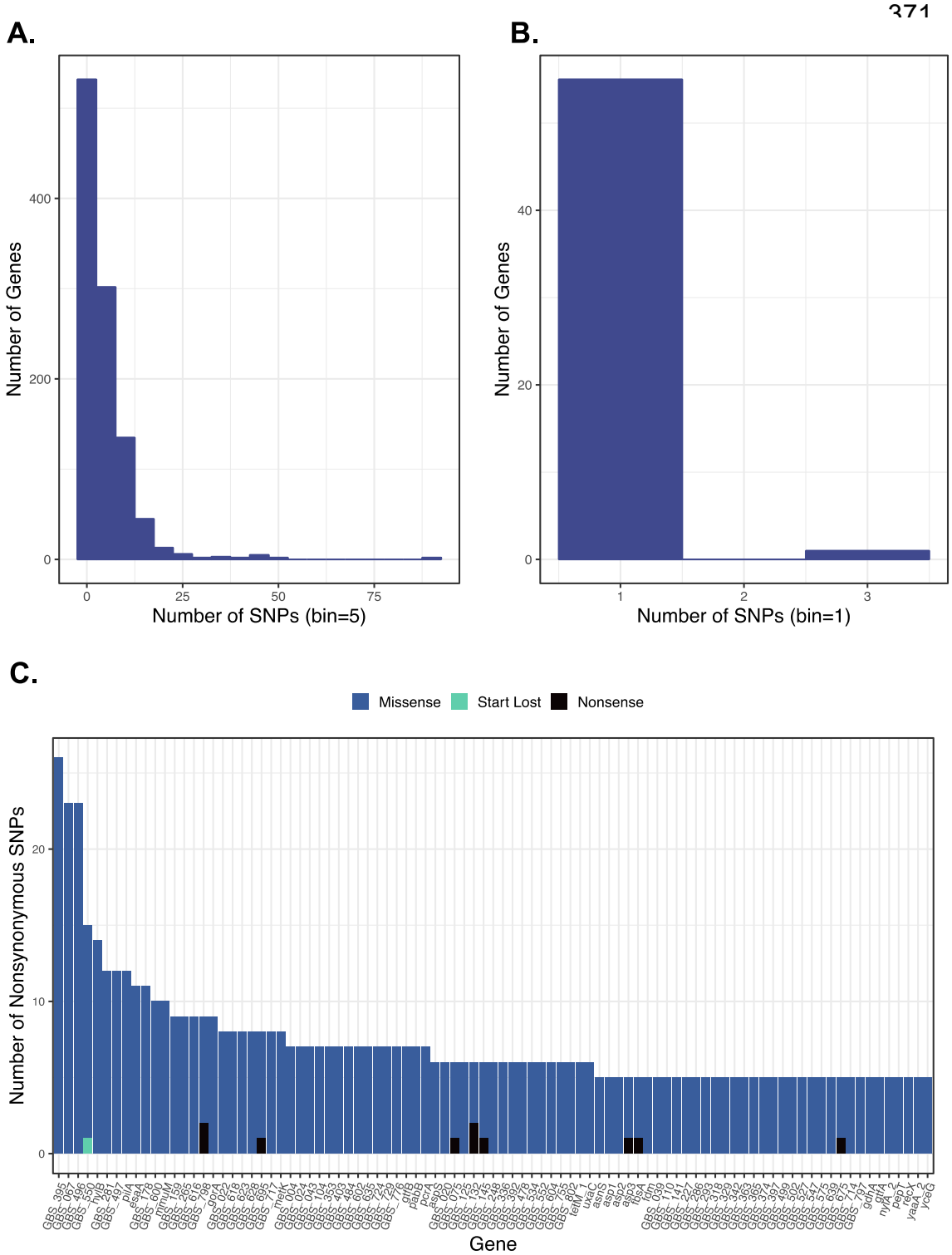
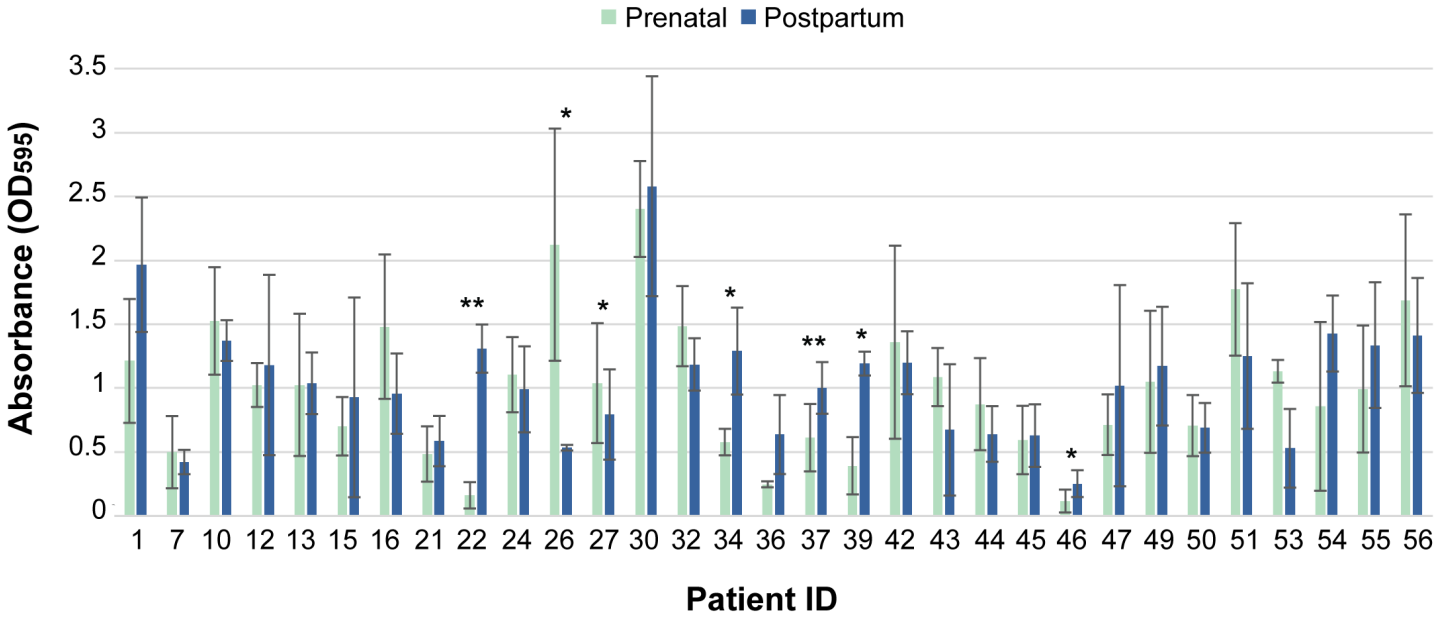


Figure 6: Number and distribution of SNPs detected within genes from the 32 postpartum genomes relative to the respective prenatal genomes. The distribution of SNP frequencies (x-axis) by the number of genes (y-axis) in the A) mutator and B) non-mutator genomes. C) The number of nonsynonymous SNP types (missense=blue, start lost=green, nonsense=black) is shown by gene (x-axis) for the subset of genes with >4 nonsynonymous SNPs in the three mutator genomes.

382 **Postpartum strains have enhanced biofilm production relative to prenatal strains**

383 Because biofilm formation is a common colonization strategy and nonsynonymous mutations
384 were found in biofilm-related genes (e.g., *pilA*, *gtfB*, GBS_067), we examined this phenotype in
385 the paired strains. Indeed, an increase in biofilm production was observed in over half (54.8%,
386 n=17) of the 31 postpartum strains relative to each paired prenatal strain (**Figure 7**); five of
387 these increases were significant. Among the mutators, two had increased biofilm production,
388 one of which (pair 22) had the largest increase observed across all pairs evaluated (fold-change
389 = 8.14) (**Figure S6**). Two postpartum isolates from pairs 26 and 27 exhibited a significant
390 decrease in biofilm formation; these isolates had acquired 11 and 2 mutations, respectively,
391 some of which occurred within genes encoding binding and surface proteins such as penicillin-
392 binding protein Pbp2B, capsule protein CpsC, and a LPXTG anchor domain surface protein.



393
394 **Figure 7: A comparison of biofilm formation between prenatal and postpartum strains**
395 **from 31 patients with persistent colonization.** Biofilm formation was assessed across the 31
396 paired strains by measuring the absorbance (OD_{595}) values (y-axis) and averaging the biological
397 replicates (n=3). Prenatal (mint green) and postpartum (dark blue) isolates are shown by patient
398 ID (x-axis). One-tailed, paired t-tests were calculated to identify significant differences in biofilm
399 formation between the paired strains (*=p <0.05, **=p<0.01). Each measurement was
400 normalized to a negative media control. Error bars display standard deviation.
401

402 **DISCUSSION**

403

404 Through this analysis, we have identified genomic traits and alterations that influence persistent
405 colonization by GBS, which we had observed in 59.4% of colonized pregnant people following
406 IAP[22], and increases the risk of consequent invasive disease. The pangenome analysis of 1,368
407 core genes uncovered five distinct sequence clusters, with most paired strains of the same ST
408 clustering together. This finding along with a comprehensive evaluation of point mutations
409 indicates that in most cases, the strains that persisted in each patient are highly similar to those
410 recovered prior to IAP. Nonetheless, we identified evidence of microevolution in GBS as point
411 mutations were found in 72% (n=21) of the 29 postpartum strains that clustered together with
412 each respective prenatal strain. It is therefore likely that the accumulation of key mutations may
413 lead to enhanced fitness in some strains, thereby promoting persistent colonization.

414

415 Indeed, we observed increased biofilm formation in over half of the postpartum strains, meaning
416 that the strains exposed to antibiotics formed better biofilms than the respective prenatal strains.
417 Consistent with this finding is the identification of point mutations in pili-associated genes (*pilA-*
418 *C, srtC1*-4) and Sec-system related genes (GBS_067, *gtfB*, *asp3*), which were linked to biofilms
419 in *Streptococcus pneumoniae*[30] as well as *cps* loci, which we found to be important for GBS
420 biofilms and ascending infection[31]. Although only five of the biofilm increases were significant,
421 many prenatal strains exhibited robust biofilm phenotypes prior to antibiotic exposure (OD_{595}
422 >1); hence, it is possible that these strains had been persisting in each host prior to our first
423 prenatal sampling. The variation in biofilm production observed is consistent with our study
424 showing differences across lineages and pilus types[32]. These data suggest that biofilms may
425 enhance persistence in some strain backgrounds, however, confirmation is required in more
426 clinically relevant conditions.

427

428 Antibiotic resistance is another mechanism for survival in antibiotics, yet we failed to detect
429 newly acquired ARGs in the postpartum strains relative to the prenatal strains from each patient.
430 Therefore, resistance to the antibiotics used for IAP was ruled out as a mechanism of persistent
431 colonization. Among those postpartum genomes from individuals (88.6%) who received β-
432 lactam antibiotics for IAP, none had β-lactam resistance genes, which is not surprising since
433 GBS is universally susceptible to the β-lactams[10]. Although resistance to the MLS antibiotics is
434 more common in GBS[10], the 11.3% of patients given clindamycin had postpartum strains that
435 were neither phenotypically resistant[23] or harbored macrolide resistance genes. Acquisition of
436 specific virulence genes was also not associated with persistence since distinct genes were not
437 found in the postpartum versus prenatal strains. Nonetheless, this assembly-based analysis has
438 some limitations in that genes labeled as absent may have fallen below our detection thresholds
439 due to poor coverage or high nucleotide variation.
440
441 Most noteworthy was our observation that three pairs (IDs 22, 43, 56) of strains from patients
442 with persistent colonization did not cluster together in the core gene phylogeny despite having
443 the same ST. The reads-based nucleotide analysis of the 1,368 core genes identified 6454,
444 262, and 4 mutations in the three postpartum genomes relative to each prenatal genome and
445 the two strains from pairs 22 and 43 were classified as mutators. This classification is based on
446 the increased mutation rates and identification of mutations in genes encoding DNA replication
447 and proofreading machinery[24,26]. The same was true for the postpartum genome from patient
448 36, which acquired 204 mutations despite having the same ST and clustering with the prenatal
449 genome in the core gene phylogeny. While mutators have been described in other streptococcal
450 species including *Streptococcus pyogenes*[33], *S. pneumoniae*[34], *S. mutans*[25], and *S. iniae*[35], this
451 study is the first to describe mutators in GBS.
452

453 Mutators (or hypermutators) represent bacterial cells within a population that have a higher-
454 than-average mutation rate[24] and typically arise in bacterial populations undergoing adaptation
455 to stressful conditions[36,37]. They have also been linked to persistent bacterial infections, such as
456 *Pseudomonas aeruginosa* infections in the lungs of cystic fibrosis patients[38]. In most cases,
457 mutators arise at low frequencies due to DNA polymerase errors, dysfunctional proofreading
458 mechanisms, and/or failure of mismatch/error correcting machinery, including the MMR
459 pathway[24,26]. Dysfunction of this system increases the likelihood of horizontal gene transfer and
460 spread of antibiotic resistance[26]. Notably, all three mutators had mutations within the MMR-
461 encoding region, which comprised four distinct classes. Each mutator strain had a different
462 subclass, demonstrating that hypermutability is not associated with a specific MMR genotype,
463 but is more likely due to the mutations that disrupt the functionality of this system.
464
465 Each mutator also belongs to a different CC. More SNPs were detected among the non-mutator
466 persistent pairs belonging to CC-23 compared to those within CCs 17 and 19. The average
467 mutation rate across non-mutator strains of CCs 1, 12, and 23 (n=11, 3.66×10^{-6}) was also
468 significantly higher than the average mutation rate of the CC17 and CC19 non-mutator strains
469 (n=6, 1.37×10^{-6}; T-test p=0.02). Together, these observations support the classification of CCs
470 1, 12, and 23 as generalists, which are more vulnerable to genomic alterations with an
471 enhanced ability to colonize a wide range of hosts[14,39]. In contrast, CCs 17 and 19, which cause
472 higher frequencies of neonatal disease[14–16], are likely to be more specialized. Nonetheless, this
473 core-gene mutation analysis has some limitations including the possibility of infrequent
474 sequencing errors. To ensure that the variation observed between paired genomes was not due
475 to sequencing artifacts, however, we manually confirmed each mutation while determining that
476 each was in an area of >10x read coverage.
477

478 Studies in *S. pneumoniae* have identified increased mutation rates after antibiotic exposure[37,40],
479 suggesting that antibiotic stress may drive genetic diversity and contribute to more robust and fit
480 populations. Although three postpartum strains still acquired mutations without IAP exposure,
481 the average mutation rate (1.50×10^{-6}) was considerably lower than for the strains with IAP
482 exposure (1.16×10^{-4}). It is notable that one of these patients (ID=56) received cefazolin
483 following childbirth instead and the postpartum strain did not cluster with the prenatal strain in
484 the core-gene phylogeny, highlighting genomic variation. Consistent with these findings, >16%
485 of genes with mutations had evidence of positive selection, suggesting that IAP exerts a
486 selective pressure on GBS, fostering selection and expansion of more resilient strains. As
487 nonsynonymous SNPs are more likely to influence protein functionality and we observed
488 multiple genes with similar mutations, it is also likely that these genes and/or gene combinations
489 have evolved in parallel and enhance the colonization potential of strains in the vaginal tract.
490
491 Among the three mutators, one strain (genome ID 22.2) underwent a capsule switch from cps V
492 to VI after IAP, which was observed previously[41] and was suggested to occur via extensive
493 recombination events that introduce genomic variation beyond the *cps* loci[41,42]. Intriguingly, this
494 mutator strain also possessed MMR loci belonging to Class 3, which was only detected in two
495 cps VI genomes in our dataset and three others among ~100 published GBS genomes
496 evaluated. Although cps VI is less common in North America, it has been shown to predominate
497 in some regions[15,43], which could be attributable to differential selective pressures across
498 locations. Identification of capsule switching *in vivo* following IAP is concerning in that cps V
499 strains along with types Ia, Ib, II, III, and IV are responsible for 98% of invasive infections in
500 babies worldwide and are targeted in the GBS6 conjugate vaccine[44,45].
501
502 Indeed, a cps switch from type V to type VI in a mutator strain background could threaten the
503 effectiveness of this vaccine in some populations due to evasion of vaccine-induced immunity

504 along with the effectiveness of IAP. The same is true for mutations outside of the *cps* loci, as
505 mutators can increase the potential for recombination events through which rare alleles and
506 genotypes can arise. Because current GBS vaccines are based on only a subset of circulating
507 cps types[44], there are concerns that vaccination could introduce a selective pressure from which
508 rare and more resilient strains can prevail as has been observed for *S. pneumoniae*[46].
509 Nonetheless, using vaccines and antibiotics to protect newborns from invasive GBS infections
510 far outweighs the risks imposed by evolutionary events that could enhance immune evasion, but
511 also highlights the need for continuously monitoring strains circulating across populations.
512 Future studies should therefore focus on better characterizing these strains to fully understand
513 the role they play in colonization and disease.
514

515 **METHODS**

516 **Bacterial isolate selection and characterization**

517 A subset of 97 GBS isolates recovered from vaginal-rectal swabs from 58 pregnant patients[23]
518 were included in this analysis. Isolates were recovered from the same patient during a prenatal
519 (35-37 weeks gestation) and postpartum (6-weeks after childbirth) visit[23], and were previously
520 characterized by MLST, serotyping, and cps typing[22]. Patients with GBS at both samplings were
521 considered to have persistent colonization resulting in 80 strains from 41 patients. The
522 remaining 17 patients had GBS only at the prenatal sampling and were considered to have lost
523 the pathogen by the postpartum sampling.
524

525 **Whole-genome sequencing (WGS)**

526 A set of 41 of the 97 isolates was sequenced previously and the raw reads were downloaded
527 from the NCBI Sequence Read Archive (SRA) (**Table S2**). The remaining 56 isolates were
528 sequenced for this study following DNA extraction from overnight cultures of GBS grown in
529 Todd-Hewitt broth (THB) at 37°C + 5% CO_2 using the E.Z.N.A. (Omega Bio-tek Inc., Norcross,

530 GA, USA) or Wizard HMW (Promega, Madison, WI, USA) kits. Construction of DNA libraries
531 was performed by the Michigan Department of Health and Human Services using the Nextera
532 XT library prep kit (Illumina, San Diego, CA, USA) followed by sequencing on the MiSeq
533 (Illumina) with 2x250 bp paired end reads.
534
535 Raw reads for the 97 genomes were trimmed with Trimmomatic v0.39[47] using "gentle trimming"
536 parameters to trim adapters and remove sequences with average quality scores <15 or less
537 than 36 nucleotides in length. The paired-end reads, or single-end reads for SRR517012 that
538 was sequenced using an unpaired method, were assessed for quality using FastQC v0.11.9
539 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/); high-quality genomes were
540 defined as those that failed <10% of quality modules. High-quality trimmed reads were
541 assembled *de novo* using SPAdes v3.13.1 (kmers 21, 33, 55, 77, 99, 127) with mismatch error
542 correction[48] and quality was assessed using QUAST v5.0.2[49] and MultiQC v1.15[50] with default
543 parameters. Assemblies were considered high-quality if N50 and L50 scores were >15,000 and
544 <50, respectively. Assembled genomes were annotated using Prokka v1.14.6 with the –proteins
545 option that allows for the use of a custom database[51]. This database was curated from 25
546 closed *S. agalactiae* genomes downloaded from NCBI to ensure that gene naming was specific
547 for GBS (**Table S10**). STs were confirmed using PubMLST[29], though previously determined
548 STs[22] were used for three strains with poor assembly quality for one of the seven MLST loci.
549
550 **Pangenome and phylogenetic analysis**
551 Roary v3.11.2 was used to create a multiFASTA alignment of core genes from the 92 high-
552 quality assemblies using the -i (blastp of 95%) and -e (PRANK aligner 170427) parameters[52].
553 ML phylogenies were generated with RAxML v8.2.12 (parameters -m GTRGAMMA, 500
554 bootstrap replicates)[53], and phylogeny visualizations and annotations were performed using the

555 Interactive tree of life (iToL, https://itol.embl.de/). Analysis of the six MMR loci was performed by
556 manually extracting each gene from the assemblies using Geneious v2022.2.2
557 (https://www.geneious.com/) followed by a MUSCLE[27] nucleotide alignment and construction of
558 a neighbor-joining phylogeny in MEGA X[54]. Neighbor-net trees were generated in SplitsTree
559 v4.19.0[55] and the PHI test was used to detect recombination with a window size of 100 and a k
560 of 1; a p-value of <0.05 was considered significant.
561
562 **Gene and nucleotide-level analyses**
563 ARGs and virulence genes were extracted from the 92 assemblies with ABRicate
564 (https://github.com/tseemann/abricate) using five ARG databases (CARD[56], ARG-annot[57],
565 MEGARes[58], Resfinder[59], NCBI AMRFinderPlus[60]) and the virulence-finder-database (VFDB)[61],
566 respectively. A gene was considered present if it had ≥10% identity and ≥80% coverage and for
567 the ARGs, if it was detected in two or more databases. Antibiotic resistance phenotypes were
568 determined previously by disk diffusion[23] and were compared to the ARG data.
569
570 Core gene mutations were extracted from the assemblies of each pair of strains recovered from
571 patients with persistent colonization (n=64, 32 pairs). For each pair, the assembled prenatal
572 genome was used as the reference, while the trimmed reads from the postpartum isolate were
573 mapped against the reference. Each core genome was interrogated for mutations including
574 SNPs, insertions, deletions, MNPs, and complex mutations using Snippy v4.6.0
575 (https://github.com/tseemann/snippy). For discrepancies in gene presence/absence within a pair
576 of strains, Geneious v2022.2.2 was used to confirm ≥15X read coverage and the up and
577 downstream regions were mapped with respect to a given gene to confirm location and
578 completeness of the gene within the contig. Core gene mutations were validated using a ≥10X
579 read coverage cutoff.

580

581 **Biofilm Assays**

582 Biofilm assays were performed in triplicate on all but one pair of strains from persistently
583 colonized patients (n=62 isolates, 31 pairs). Values were normalized to blank media controls as
584 described[32] with the modification of stagnant growth for 24hrs. The magnitude (fold-change) of
585 the difference in biofilm formation was determined for the postpartum strain relative to the paired
586 prenatal strain.

587

588 **Data analysis and statistics**

589 Raw output from WGS analyses were managed in Microsoft Excel. R v4.1.2 (https://www.r-
590 project.org/) was used in RStudio v2022.07.1+554 (https://www.rstudio.com/) for importing data
591 (readr v2.1.4), data wrangling (devtools v2.4.5, dplyr v1.1.3, forcats v1.0.0, plyr v1.8.8, tidyr
592 v1.3.0, tidyverse v2.0.0), and visualization (ggplot2 v3.4.3, viridis v0.6.4) (https://cran.r-
593 project.org/). R package factoextra v1.0.7 was used to perform and visualize PCA with 95%
594 confidence ellipses for virulence genes using the prcomp() and fviz_pca() functions
595 (https://cran.r-project.org/). Chi-square tests were performed in Microsoft Excel or Epi Info™ v7
596 (Centers for Disease Control and Prevention, Atlanta, GA, USA); statistically significant
597 associations were classified at $p \leq 0.05$. For unnamed genes that acquired mutations, arbitrary
598 "gene ids" using the GBS_00X notation were assigned for easier reference. For all SNP-only
599 analyses, SNPs were filtered to exclude those that occurred in rRNA and/or uncharacterized,
600 non-coding regions.

601

602 As characterized by Snippy, nonsynonymous SNPs included those classified as "initiator codon
603 variant", "missense", "start lost", "stop lost", and "stop gained" (nonsense), while synonymous
604 SNPs included those referred to as "stop retained" and "synonymous variant". Measures of
605 evolutionary selection were calculated based on the nonsynonymous to synonymous SNP ratio

606 ($\pi N/\pi S$) across the genomes, as well as across each gene of interest to discern areas under
607 positive ($\pi N/\pi S > 1$), neutral ($\pi N/\pi S = 1$), or negative ($\pi N/\pi S < 1$) selection. Calculations for
608 evolutionary selection across genes were limited to genes that acquired a total of 5 or more
609 SNPs. Student t-tests were used to analyze differences in biofilm formation between isolate
610 pairs and mutation rates across CCs; a p-value <0.05 was deemed significant.
611
612 **Data availability**
613 The datasets are available in the supplementary tables and the raw sequencing reads for all 97
614 genomes have been deposited in GenBank.
615
616 **Acknowledgements**
617 We thank the Research Technology Support Facility and Michigan Department of Agriculture
618 and Rural Development for DNA sequencing, as well as the National Center for *Streptococcus*
619 (Edmonton, Alberta), particularly Marguerite Lovgren, for strain collection and serotyping.
620 Financial support was provided by the Michigan State University (MSU) Research Foundation
621 (to SDM) and in part, by the Michigan Sequencing and Academic Partnerships for Public Health
622 Innovation and Response (MI-SAPPHIRE) initiative at the MDHHS via CDC through the
623 Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious
624 Diseases Enhancing Detection Expansion program (6NU50CK000510-02-07; to HMB and
625 SDM). Additional support was provided by the National Institutes of Health (R01HD090061 to
626 JAG and SDM, R01AI134036 to JAG) and Department of Veterans Affairs Merit Award
627 I01BX005352 (Office of Research) to JAG. Student support was provided to MEP by the MSU
628 Department of MGI via the Philipp and Vera Gerhardt Travel and Ralph Evans Awards and the
629 MSU College of Natural Science via two Outstanding Scholar Fellowships.
630

631 **Author contributions**

632 Conceptualization: MEP, SDM; methodology: HMB, MEP; validation and formal analysis: MEP,
633 SDM; resources: SDM, HMB, HDD; supervision and insight: SDM, HMB, JAG, HDD; data
634 curation: MEP, SDM; writing—original draft preparation, MEP; writing—review and editing and
635 data visualization: MEP, SDM; funding acquisition: SDM, HMB, JAG. All authors have read and
636 agreed to the final version of the manuscript.

637

638 **Competing interests**

639 The authors declare no competing interests or other interests that could influence the
640 interpretation of the article.

641 **REFERENCES**

642 1. Russell, N. J. *et al.* Maternal colonization with group B *Streptococcus* and serotype
643 distribution worldwide: Systematic review and meta-analyses. *Clin Infect Dis* **65**
644 (suppl_2), S100-S111 (2017).
645 2. Boyer, K. M. & Gotoff, S. P. Prevention of early-onset neonatal group B streptococcal
646 disease with selective intrapartum chemoprophylaxis. *N Engl J Med* **314**, 1665–9 (1986).
647 3. Verani, J. R., McGee, L., Schrag, S. J. Prevention of perinatal group B streptococcal
648 disease--revised guidelines from CDC, 2010. *MMWR Recomm Rep* **59**, 1–36 (2010).
649 4. Phares, C. R. *et al.* Epidemiology of invasive group B streptococcal disease in the United
650 States, 1999-2005. *JAMA* **299**, 2056-65 (2008).
651 5. Seale, A. C. *et al.* Stillbirth with group B *Streptococcus* disease worldwide: Systematic
652 review and meta-analyses. *Clin Infect Dis* **65,** S125–S132 (2017).
653 6. Schuchat, A. Epidemiology of group B streptococcal disease in the United States: Shifting
654 paradigms. *Clin Microbiol Rev* 11, 497-513 (1998).
655 7. Mei, J. Y. & Silverman, N. S. Group B *Streptococcus* in pregnancy. *Obstet Gynecol Clin*
656 *North Am* **50**, 375–387 (2023).
657 8. Schrag, S. J. *et al.* Group B streptococcal disease in the era of intrapartum antibiotic
658 prophylaxis. *N Engl J Med* **342**, 15–20 (2000).
659 9. Azad, M. B. *et al.* Impact of maternal intrapartum antibiotics, method of birth and
660 breastfeeding on gut microbiota during the first year of life: A prospective cohort study.
661 *BJOG* **123**, 983–993 (2016).
662 10. Schrag, S. J. & Verani, J. R. Intrapartum antibiotic prophylaxis for the prevention of
663 perinatal group B streptococcal disease: Experience in the United States and implications
664 for a potential group B streptococcal vaccine. *Vaccine* **31**, D20–D26 (2013).
665 11. Anthony, B. F., Okada, D. M. & Hobel, C. J. Epidemiology of the group B *Streptococcus*:
666 Maternal and nosocomial sources for infant acquisitions. *J Pediatr* **95**, (1979).
667 12. Kwatra, G. *et al.* Prevalence of maternal colonisation with group B *Streptococcus*: a
668 systematic review and meta-analysis. *Lancet Infect Dis* **16**, 1076–1084 (2016).
669 13. Van Dyke, M. K. *et al.* Evaluation of Universal Antenatal Screening for Group B
670 *Streptococcus*. *Obstet Gynecol Surv* **64**, 703–704 (2009).
671 14. Jones, N. *et al.* Multilocus sequence typing system for group B *Streptococcus*. *J Clin*
672 *Microbiol* **41**, 2530–2536 (2003).

673 15. Bianchi-Jassir, F. *et al.* Systematic review of group B streptococcal capsular types,
674 sequence types and surface proteins as potential vaccine candidates. *Vaccine* **38**,
675 (2020).
676 16. Manning, S. D. *et al.* Multilocus sequence types associated with neonatal Group B
677 streptococcal sepsis and meningitis in Canada. *J Clin Microbiol* **47**, 1143–1148 (2009).
678 17. Springman, A. C. *et al.* Selection, recombination, and virulence gene diversity among
679 group B streptococcal genotypes. *J Bacteriol* **191**(17):5419-27 (2009).
680 18. Korir, M. L. *et al.* Association and virulence gene expression vary among serotype III
681 group B *Streptococcus* isolates following exposure to decidual and lung epithelial cells.
682 *Infect Immun* **82**, (2014).
683 19. Korir, M. L. *et al.* Differing mechanisms of surviving phagosomal stress among group B
684 *Streptococcus* strains of varying genotypes. *Virulence* **8**, 924–937 (2017).
685 20. Korir, M. L. *et al. Streptococcus agalactiae cadD* alleviates metal stress and promotes
686 intracellular survival in macrophages and ascending infection during pregnancy. *Nat*
687 *Commun* **13**, (2022).
688 21. Flaherty, R. A., Aronoff, D. M., Gaddy, J. A., Petroff, M. G. & Manning, S. D. Distinct
689 Group B *Streptococcus* sequence and capsule types differentially impact macrophage
690 stress and inflammatory signaling responses. *Infect Immun* **89**, e00647-20 (2021).
691 22. Manning, S. D. *et al.* Genotypic diversity and serotype distribution of group B
692 *Streptococcus* isolated from women before and after delivery. *Clin Infect Dis* **46**, 1829–
693 1837 (2008).
694 23. Spaetgens, R. *et al.* Perinatal antibiotic usage and changes in colonization and
695 resistance rates of group B *Streptococcus* and other pathogens. *Obstet Gynecol* **100**,
696 525–533 (2002).
697 24. Miller, J. H. Spontaneous mutators in bacteria: Insights into pathways of mutagenesis and
698 repair. *Annu Rev Microbiol* **50**, 625–643 (1996).
699 25. Gonzalez, K., Faustoferri, R. C. & Quivey, R. G. Role of DNA base excision repair in the
700 mutability and virulence of Streptococcus mutans. *Mol Microbiol* **85**, 361–377 (2012).
701 26. Chopra, I., O'Neill, A. J. & Miller, K. The role of mutators in the emergence of antibiotic-
702 resistant bacteria. *Drug Resist Updat* **6**, (2003).
703 27. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high
704 throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
705 28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
706 search tool. *J Mol Biol* **215**, (1990).

707 29. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics:
708 BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*
709 **3**, 1–20 (2018).
710 30. Bandara, M. *et al.* The accessory Sec system (SecY2A2) in *Streptococcus pneumoniae*
711 is involved in export of pneumolysin toxin, adhesion and biofilm formation. *Microbes*
712 *Infect* **19**, 402–412 (2017).
713 31. Noble, K. *et al.* Group B *Streptococcus cpsE* is required for serotype V capsule
714 production and aids in biofilm formation and ascending infection of the reproductive tract
715 during pregnancy. *ACS Infect Dis* **7**, (2021).
716 32. Parker, R. E. *et al.* Association between genotypic diversity and biofilm production in
717 group B *Streptococcus*. *BMC Microbiol* **16**, 1–12 (2016).
718 33. Scott, J., Thompson-Mayberry, P., Lahmamsi, S., King, C. J. & McShan, W. M. Phage-
719 associated mutator phenotype in group A *Streptococcus*. *J Bacteriol* **190**, 6290–6301
720 (2008).
721 34. Samrakandi, M. M. & Pasta, F. Hyperrecombination in *Streptococcus pneumoniae*
722 depends on an atypical *mutY* homologue. *J Bacteriol* **182**, (2000).
723 35. Rudenko, O., Engelstädter, J. & Barnes, A. C. Evolutionary epidemiology of
724 *Streptococcus iniae*: Linking mutation rate dynamics with adaptation to novel
725 immunological landscapes. *Infect Gen Evol* **85**, (2020).
726 36. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. Evolution of high mutation rates in
727 experimental populations of *E. coli*. *Nature* **387**, (1997).
728 37. Mehta, H. H. *et al.* The essential role of hypermutation in rapid adaptation to antibiotic
729 stress. *Antimicrob Agents Chemother* **63**, (2019).
730 38. Oliver, A., Cantón, R., Campo, P., Baquero, F. & Blázquez, J. High frequency of
731 hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science (1979)*
732 **288**, (2000).
733 39. Richards, V. P. *et al.* Population gene introgression and high genome plasticity for the
734 zoonotic pathogen *Streptococcus agalactiae*. *Mol Biol Evol* **36**, (2019).
735 40. Henderson-Begg, S. K., Livermore, D. M. & Hall, L. M. C. Effect of subinhibitory
736 concentrations of antibiotics on mutation frequency in *Streptococcus pneumoniae*. *J*
737 *Antimicrob Chemother* **57**, 849–854 (2006).
738 41. Neemuchwala, A., Teatero, S., Athey, T. B. T., Mcgeer, A. & Fittipaldi, N. Capsular
739 switching and other large-scale recombination events in invasive sequence type 1 group
740 B *Streptococcus*. *Emerg Infect Dis* **22**, 1941–1944 (2016).

741 42. Brochet, M. *et al.* Shaping a bacterial genome by large chromosomal replacements, the
742 evolutionary history of *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* **105**, 15961–
743 15966 (2008).
744 43. Lachenauer, C. S. *et al.* Serotypes VI and VIII predominate among group B streptococci
745 isolated from pregnant Japanese women. *J Infect Dis* **179**, (1999).
746 44. Buurman, E. T. *et al.* A novel hexavalent capsular polysaccharide conjugate vaccine
747 (GBS6) for the prevention of neonatal group B streptococcal infections by maternal
748 immunization. *J Infect Dis* **220**, (2019).
749 45. Madhi, S. A. *et al.* Potential for Maternally Administered Vaccine for Infant Group B
750 *Streptococcus*. *N Engl J Med* **389**, (2023).
751 46. Brueggemann, A. B., Pai, R., Crook, D. W. & Beall, B. Vaccine escape recombinants
752 emerge after pneumococcal vaccination in the United States. *PLoS Pathog* **3**, 1628–1636
753 (2007).
754 47. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina
755 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
756 48. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to
757 single-cell sequencing. *J Comput Biol* **19**, 455–477 (2012).
758 49. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for
759 genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
760 50. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results
761 for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
762 51. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
763 2069 (2014).
764 52. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis.
765 *Bioinformatics* **31**, 3691–3693 (2015).
766 53. Stamatakis, A., Ludwig, T. & Meier, H. RAxML-III: A fast program for maximum likelihood-
767 based inference of large phylogenetic trees. *Bioinformatics* **21**, 456–463 (2005).
768 54. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary
769 genetics analysis across computing platforms. *Mol Biol Evol* **35**, 1547–1549 (2018).
770 55. Huson, D. H. SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14**,
771 68–73 (1998).
772 56. Jia, B. *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive
773 antibiotic resistance database. *Nucleic Acids Res* **45**, D566–D573 (2017).

774 57. Gupta, S. K. *et al.* ARG-annot, a new bioinformatic tool to discover antibiotic resistance
775 genes in bacterial genomes. *Antimicrob Agents Chemother* **58**, 212–220 (2014).
776 58. Doster, E. *et al.* MEGARes 2.0: A database for classification of antimicrobial drug, biocide
777 and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res*
778 **48**, D561–D569 (2020).
779 59. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob*
780 *Chemother* **67**, 2640–2644 (2012).
781 60. Feldgarden, M. *et al.* Validating the AMRFINder tool and resistance gene database by
782 using antimicrobial resistance genotype-phenotype correlations in a collection of isolates.
783 *Antimicrob Agents Chemother* **63**, 1–19 (2019).
784 61. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: Hierarchical and refined
785 dataset for big data analysis - 10 years on. *Nucleic Acids Res* **44**, D694–D697 (2016).
786