Genomic adaptation in group B *Streptococcus* following intrapartum antibiotic prophylaxis and childbirth

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

SUMMARY

Through vaginal colonization, GBS causes severe outcomes including neonatal sepsis and meningitis. Although intrapartum antibiotic prophylaxis (IAP) has reduced neonatal 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, a pangenome analysis was performed on 97 isolates from 58 participants before (prenatal) and after (postpartum) IAP/childbirth. Thirty-one of the 34 paired strains from participants with persistent colonization clustered together in the core gene phylogeny, suggesting that the strains recovered at the postpartum sampling were highly similar to those recovered at the prenatal visit. A core-gene mutation analysis identified mutations in 74% (n=23) of the 31 postpartum genomes when compared to the prenatal strains of the same multilocus sequence type recovered from the same individuals. Several strains had acquired mutations in the same colonization-associated genes, though two postpartum strains accounted for most of the mutations. These two outliers were classified as mutators based on high mutation rates and mutations within DNA repair system genes. Increased biofilm production was observed in half of the postpartum strains relative to the prenatal strains, which is supported by the presence of point mutations in genes associated with adherence. Together, these findings suggest that antibiotics may impose a selective pressure on GBS that selects for mutations and phenotypes that promote adaptation and survival *in vivo*. Enhanced survival in the genitourinary tract can lead to persistent colonization, increasing the likelihood of invasive disease in subsequent pregnancies and in newborns following IAP.

INTRODUCTION

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

In the U.S., screening for vaginal-rectal colonization is recommended between 36 0/7 and 37 6/7 weeks' gestation[7]. To prevent neonatal infections, GBS-positive patients are given intrapartum antibiotic prophylaxis (IAP) during labor, for which penicillin is the most effective[3]. While IAP has successfully reduced the incidence of EOD, it has not impacted LOD rates nor is it useful for preventing preterm births or stillbirths.[8] It can also negatively impact the microbiome[9,10]. Importantly, GBS colonization was found to be intermittent during pregnancy with variation in colonization frequencies across individuals and geographic locations[11,12]. One study of EOD cases, for example, showed that 60% of cases were born to individuals whose GBS prenatal screening test was negative[13], thereby highlighting the need for improved treatment and prevention protocols.

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

53 (e.g., STs 19 and 23) despite their high prevalence globally[16]. The effectiveness of IAP was also
54 found to vary across STs in our prior study[22]. Relative to other STs, STs 17 and 19 more
55 commonly colonized patients before and after IAP while persisting for up to six weeks
56 postpartum. ST-12 strains, however, were more frequently eradicated after IAP, which was also
57 independently associated with persistent colonization. Despite these associations, the
58 mechanisms that promote GBS persistence in the vaginal tract and evasion of antibiotic-
59 mediated killing remain elusive. This is a serious concern since persistent colonization
60 enhances the risk of invasive disease, especially for subsequent pregnancies and babies after
61 IAP cessation[10,22], and may partly explain why GBS LOD rates have not decreased in the
62 U.S.[8,10].
63
64 Altogether, these prior findings led us to hypothesize that IAP imposes a selective pressure on
65 GBS, such that genomic features that promote colonization are selected for, resulting in
66 enhanced fitness and survival in the genitourinary tract. We used whole-genome sequencing
67 (WGS) to compare GBS isolates recovered from pregnant participants sampled both before
68 (prenatal) and 6 weeks after (postpartum) IAP and childbirth[23]. This comprehensive genomics
69 analysis has enhanced understanding of evolution and adaptation among colonizing GBS
70 populations, concurrently providing insight for treatment and prevention improvements.
71
72 **RESULTS**
73
74 **Antibiotic use and GBS colonization dynamics differ across participants**
75 Isolates (n=97) were previously collected[23] from 58 participants at two time points during
76 pregnancy and after childbirth. The first sampling took place at a prenatal visit between 35-37
77 weeks gestation, which was the previously recommended timing for GBS screening but differs
78 from recent guidelines of 36-37 weeks gestation[7]. The second sampling took place up to 11

weeks later at a postpartum visit. Among the 58 participants included in this follow-up study, 53 (91.4%) were given IAP during childbirth, with most receiving ampicillin (n=26; 49.1%) or penicillin (n=17; 32.1%) (**Table S1**).

Forty-one of the 58 (70.7%) participants had persistent colonization, meaning they were positive for GBS at both the prenatal and postpartum samplings even though IAP was given to all but three individuals (n=38/41; 92.7%). Among these 41 persistently colonized participants, 35 received a β-lactam antibiotic (ampicillin, n=21; penicillin, n=13; or cefazolin, n=1) and three were given clindamycin. Four of the participants also reported taking antibiotics after childbirth. One was treated with cefazolin and three received either ampicillin (n=1), a combination of trimethoprim and sulfamethoxazole (Septra) with clindamycin and nitrofurantoin (n=1), or metronidazole and cephalexin (n=1) prior to the postpartum sampling. The remaining 17 (29.3%) participants were colonized with GBS only at the prenatal sampling, indicating loss of GBS colonization by the postpartum visit; two of these participants did not receive IAP.

As determined previously[22], the 97 GBS strains belong to 20 STs comprising clonal complex (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 CC-26 (n=1) plus one singleton (ST-67). Eight distinct cps types were represented, with cps III (n=27) predominating followed by Ia (n=21) and V (n=17); four isolates were non-typeable (NT).

Pangenome analysis reveals a large accessory genome and clustering by CCs

Ninety-two of the 97 genomes were included in the genomics analysis. Three prenatal and two postpartum genomes from five participants were excluded due to poor assembly quality (**Table S2**), whereas two postpartum strains from two persistently colonized participants (IDs 11 and 14) were not available for WGS. Among the 92 high-quality genomes, 68 represented pairs of strains recovered from 34 participants with persistent colonization (i.e., participants were

105 positive for GBS at both the prenatal and postpartum samplings). Seven unpaired strains from
106 seven persistently colonized participants collected at the prenatal (n=4) or postpartum (n=3)
107 sampling were available for inclusion as well as the 17 prenatal strains from participants who
108 had lost GBS by the postpartum sampling.
109
110 The GBS pangenome consists of 5,054 unique genes comprising a small core-genome with
111 1,368 core and 213 soft-core genes (**Table S3**). It has a large accessory genome containing
112 945 shell genes and 2,528 cloud genes, which accounts for 68.7% of the total genes. The
113 evolutionary history of all strains was elucidated based on the 1,368 core genes using the
114 Maximum Likelihood method and Tamura-Nei model of nucleotide substitutions in MEGA12,[24]
115 which produced a phylogenetic tree with the highest log likelihood (**Figure S1**). Five sequence
116 clusters were resolved, grouping together with 100% bootstrap support. Each of the five clusters
117 was associated with a previously defined CC (**Figure 1**). The 17 CC-19 strains and 16 CC-17
118 strains are grouped into two separate clusters as are the CC-1 (n=23), CC-12 (n=14), and CC-
119 23 (n=19) strains. Several singletons were detected within each cluster. Construction of a
120 neighbor-net tree in SplitsTree v.19.2,[25] which was based on 17,825 parsimony-informative (PI)
121 sites within the 1,368 core genes, confirmed the clustering observed in the maximum likelihood
122 phylogeny and uncovered evidence for recombination (pairwise homoplasy index (PHI) p-value
123 = 0.00; **Figure S2**).
124
125 Among the 34 pairs of strains from participants with persistent colonization, 30 (88.2%)
126 clustered together in the phylogenetic tree (Figures 1 and S1). The paired strains from the three
127 participants (IDs 50, 55, and 56) who did not receive IAP also grouped together. Of the four
128 pairs that did not group together, two (IDs 3 and 8) had prenatal and postpartum strains with
129 different STs, suggesting acquisition of a new strain type following childbirth and IAP. Although
130 the remaining two participants (IDs 22 and 43) had prenatal and postpartum strains with the

same ST, only the prenatal and postpartum strains from participant 43 clustered together. Inclusion of only the paired strains in the consensus tree better illustrates that prenatal strain 43.1 is ancestral to postpartum strain 43.2 (**Figure S3**).

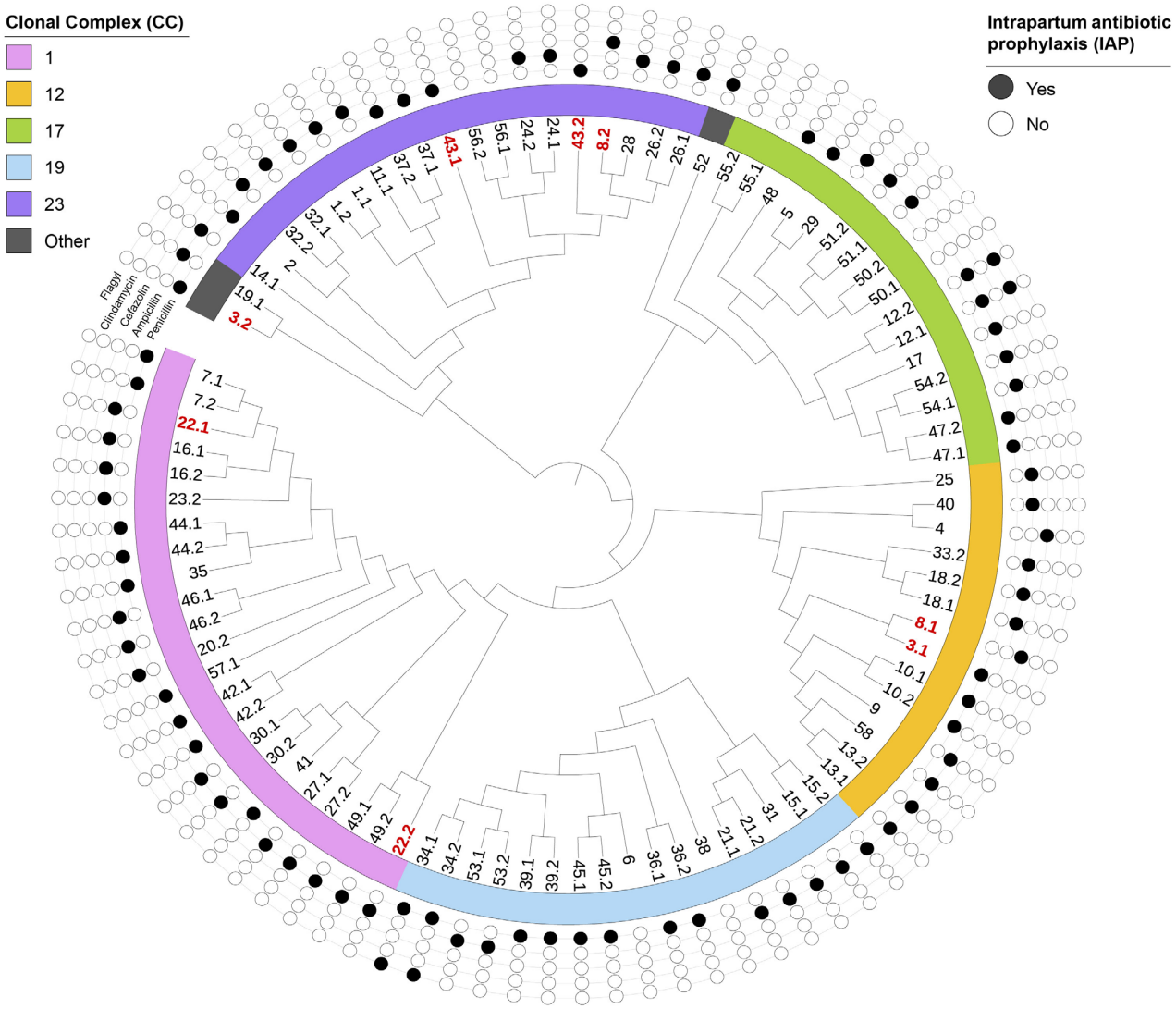


Figure 1: Evolutionary history of 92 GBS strains isolated from participants before and after IAP and childbirth. The bootstrap consensus tree based on 1,368 core genes (1,268,723 positions) was inferred from 88 replicates[26] with the collapsed branches corresponding to partitions reproduced in less than 50% of replicate trees. The original maximum likelihood tree (Figure S1) represents that with the highest log-likelihood between a Neighbor-Joining (NJ) tree[27] and Maximum Parsimony (MP) tree. The phylogeny was constructed in MEGA12[24] and visualized with iTOL[28] while rooting at the midpoint. Participant strain IDs are noted at the end of each branch; persistent strains contain a ".1" (prenatal) or ".2" (postpartum) indicating the

144 sampling timepoint. The clonal complex (CC) is indicated by colored strips and the circles
145 correspond to the antibiotic(s) used for IAP. Paired strains from the same participant have the
146 same antibiotic usage profiles. IDs in bold red font are the paired prenatal and postpartum
147 strains from four participants (IDs 3, 8, 22, and 43) that do not cluster together in the tree.
148
149
150
151
152
153 **Acquisition of antibiotic resistance genes is not important for persistent colonization**

154 To determine whether GBS residing in the vaginal tract acquired antibiotic resistance genes
155 (ARGs) to promote colonization, an assembly-based analysis was performed to extract these
156 genes from the 92 genomes. Nine distinct ARGs were identified conferring resistance to
157 tetracyclines (*tetL, tetM, tetO, tetW, lmrP*), macrolides-lincosamides-streptogramins (MLS)
158 (*ermA, lmrP, mreA*), fluoroquinolones (*norB*), and cationic peptides (*mprF*) (**Figure 2**). Both
159 *lmrP* and *mprF* were found in all strains, while *norB* and *mreA* were found in all but one.
160

161 Among the variably present ARGs, *tetM* predominated and was detected in 63 (68.5%)
162 genomes including 47 of the 68 (69.1%) genomes from the 34 participants with persistent
163 colonization. A similar proportion (n=11; 64.7%) was detected in the 17 prenatal strains that
164 were lost following IAP. Although the prenatal and postpartum strains from participant 22 had
165 the same ST, *tetM* was only detected in the prenatal strain (22.1). A discrepancy was also
166 observed between the paired genomes from participant 3, though this finding is less surprising
167 in that the two strains had different STs, strongly suggesting the acquisition of a new strain by
168 the postpartum visit.
169

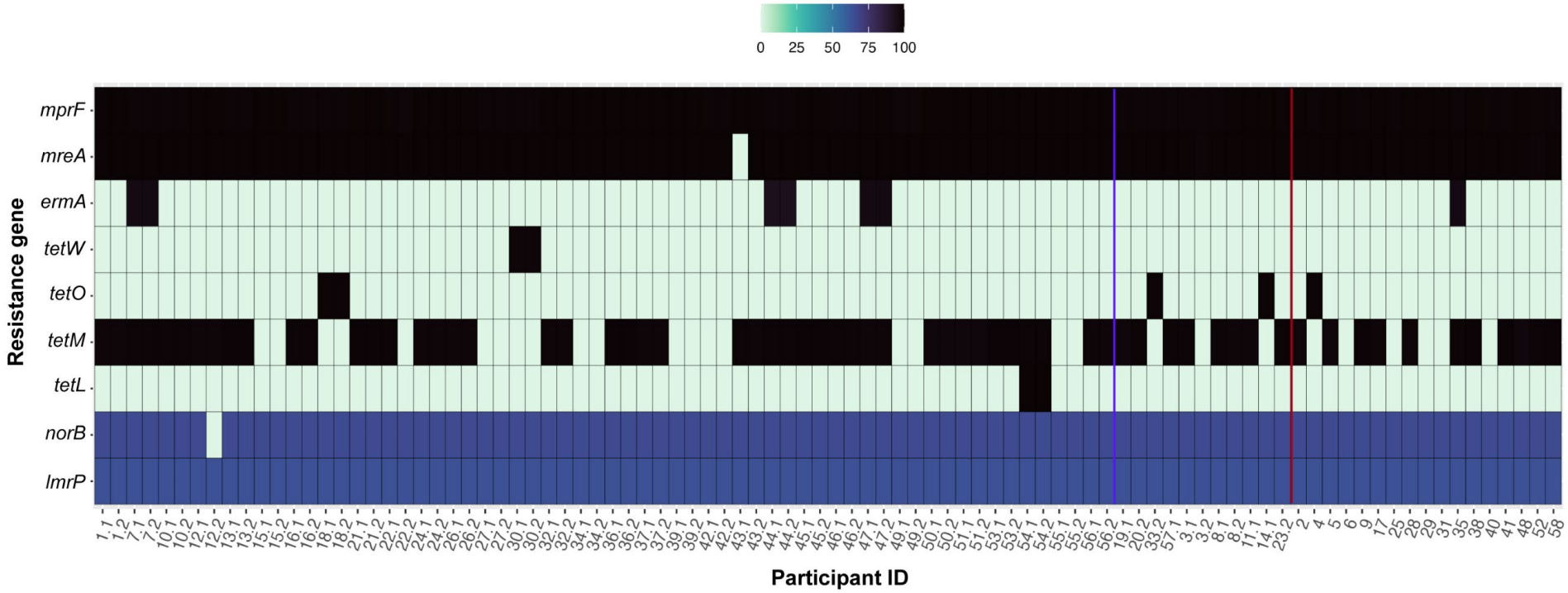


Figure 2: Antibiotic resistance genes (ARGs) detected in the 92 genomes. ARGs were extracted from the 92 high-quality genome assemblies using five different ARG databases. The x-axis shows the patient strains with a ".1" (prenatal) and ".2" (postpartum) after the "Participant ID" indicating the sampling timepoint. The gene profiles for the paired isolates from the same 32 participants with isolates of the same sequence type (ST) are shown next to each other on the left side of the vertical blue line. The two sets of paired strains (IDs 3 and 8) with different STs along with seven unpaired strains are shown between the blue 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. ARGs are indicated on the y-axis that confer resistance to tetracyclines (*tetL, tetM, tetO, tetW, lmrP*), macrolides-lincosamides-streptogramins (*ermA, lmrP, mreA*), fluoroquinolones (*norB*), and cationic peptides (*mprF*). The color gradient shows the average percent identity of a gene across the five databases, with black representing the highest (100%) identity and light green representing the lowest (<10%) identity.

182 **Specific virulence genes or gene profiles are not linked to persistent colonization**

183 The 92 genome assemblies were also interrogated for known virulence genes; 50 distinct genes
184 were detected with 65-100% identity and classified based on function (**Table S4**). Genes that
185 are important for adherence and invasion of host cells (n=18, 36%), immune modulation (n=17,
186 34%), metabolism (n=1, 2%), toxin production (n=13, 26%), and dissemination (n=1, 2%) were
187 found. The most variable genes were those that are important for capsule (*cps*) production and
188 pili (*pil*) formation (**Figure 3**).

189

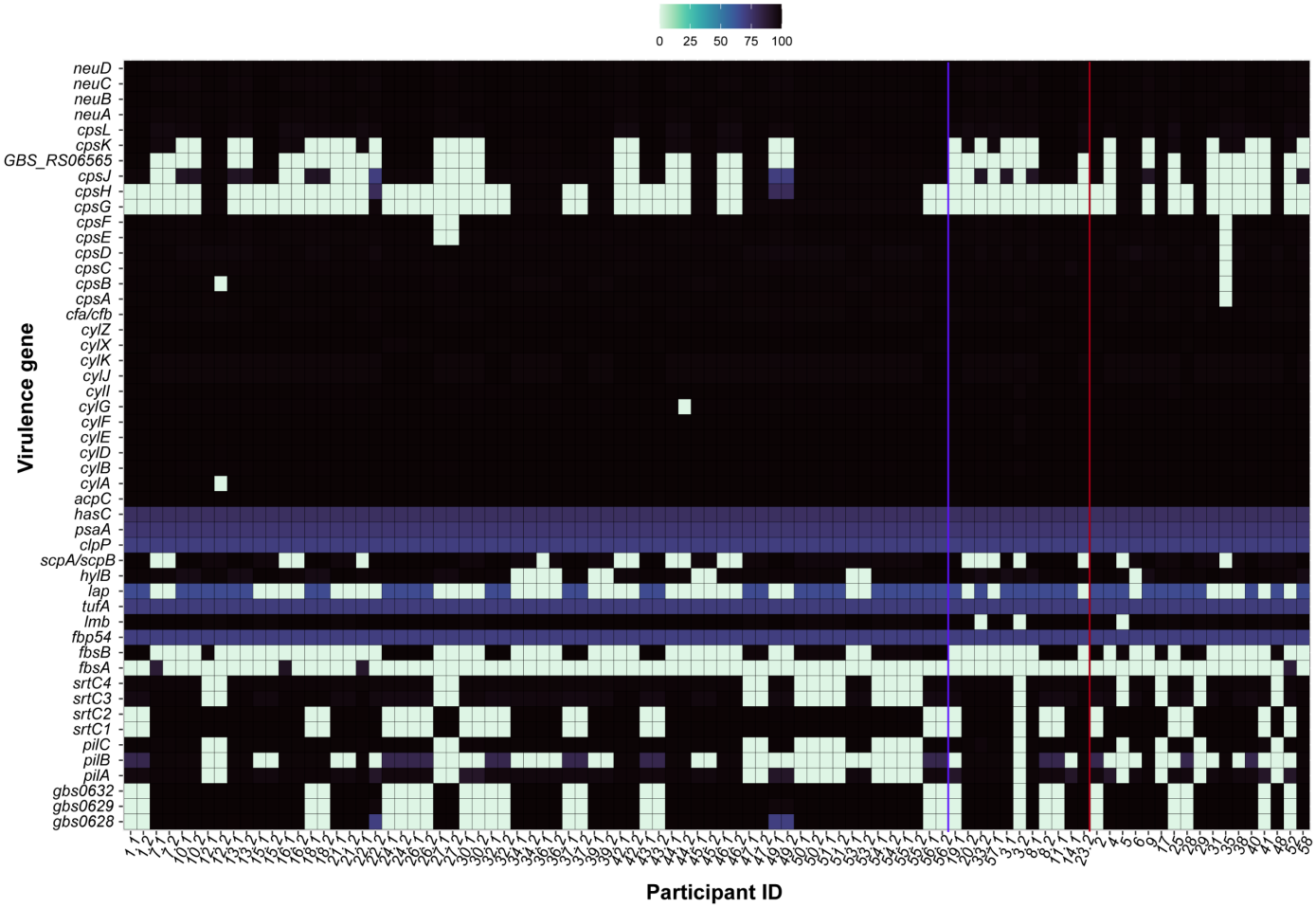
190 Among 32 of the 34 participants persistently colonized with strains of the same ST, several
191 discrepancies in gene presence were observed between the prenatal and postpartum strains
192 per pair. In most cases, the differences involved genes with values that fell below the identity
193 and coverage thresholds, indicating that they are present but may be variable. The paired
194 strains from participant 22, however, had multiple differences. Most notably, distinct capsule
195 (cps) types (serotype) were observed in the prenatal (*cpsV*) and postpartum (*cpsVI*) strains,
196 which was demonstrated previously,[22] and could indicate capsular switching or acquisition of a
197 new strain by the postpartum sampling. Not surprisingly, participants 3 and 8 with prenatal and
198 postpartum strains belonging to different STs had distinct virulence gene profiles.

199

200 To further compare the virulence gene profiles among the paired strains from the participants
201 with persistent colonization, a principal component analysis was performed. No distinct
202 clustering patterns, however, were observed for the prenatal versus postpartum strains (**Figure**
203 **S4A**) or after stratifying by antibiotic, *cps* type, or colonization phenotype (persistent or lost)
204 (**Figure S4B-E**). In addition, no difference was observed between colonization phenotype and
205 the number of adherence (n=1,031, Chi-square p = 0.10) or immune modulatory (n=1,231, Chi-
206 square p=0.52) genes. These data suggest that neither possession or acquisition of specific

207 combinations of virulence genes were linked to persistent colonization of the vaginal tract after

208 IAP and childbirth.



209 **Figure 3: Virulence genes detected in 92 GBS genomes.** The x-axis shows the participant ID
210 with the ".1" (prenatal) and ".2" (post-partum) indicating the sampling timepoint. Gene names
211 are displayed on the y-axis. The gene profiles for the paired isolates from 32 participants with
212 isolates of the same sequence type (ST) are next to each other, left of the purple line. The
213 seven unpaired strains plus the two paired strains with different STs from participants 3 and 8
214 are shown between the purple and red vertical lines. Profiles for the 17 "lost" strains are to the
215 right of the red line. The color gradient displays the average percent identity of a gene, with
216 black representing 100% identity and mint green representing <10% identity.
217

218
219 **The distribution of point mutations varies across persistent strains**

220 To detect nucleotide-level variation, we conducted a reads-based analysis of the 1,368 core

221 genes in the 64 paired prenatal and postpartum strains that shared STs. In this analysis, each

222 prenatal genome was used as the reference genome for its paired postpartum genome and the

223 number and type of mutations in the postpartum genomes were determined. Alarmingly, 7,025
224 mutations were identified in 24 (75.0%) of the 32 postpartum genomes, ranging from 1 to 6,564
225 mutations per genome with an average of 292 (**Table S5**). No mutations were detected in eight
226 postpartum genomes, which have identical core gene sequences as their paired prenatal
227 genomes despite IAP exposure. Of the three participants who did not receive IAP, the
228 postpartum isolates had only 2, 3, or 4 total mutations when compared to their respective
229 prenatal isolates.
230
231 Among the 24 postpartum genomes with at least one mutation, a skewed distribution of
232 mutations was observed with three outliers (**Figure 4A**). These outlier genomes contained 204,
233 262, and 6,454 total mutations as opposed to an average of five mutations per genome and
234 range of 1-16 mutations for the remaining 21 postpartum genomes (**Figure 4B**). Most (n=6,065;
235 86.3%) of the mutations were in coding regions, though an additional 958 (13.6%) were in
236 uncharacterized regions and two were in rRNAs (**Table S6**). The type of mutations varied
237 across genomes (**Figure S5A**) with the majority (n=5,982; 85.2%) representing single
238 nucleotide polymorphisms (SNPs) including synonymous (n=3,469; 58.0%), missense (n=1,771;
239 29.6%), or nonsense (n=26; 0.4%) mutations (**Figure S5B**).
240
241 To define which genes possessed SNPs within coding sequences, we mapped the 5,291
242 characterized SNPs to specific genes. Most of these SNPs (n=5,233; 98.9%) were detected in
243 the three outlier genomes, mapping to 1,167 genes and yielding an average of 2 SNPs/gene
244 (**Figure S6A**). After excluding all three outliers from the analysis, 58 SNPs were found within 56
245 genes in the remaining genomes (**Figure S6B**). Within the outlier genomes, most genes
246 (n=683) acquired less than four SNPs relative to the respective prenatal genomes with 390,
247 142, 86, and 65 genes having 1, 2, 3, and 4 SNPs, respectively. Four or more SNPs, however,

were detected within 366 genes. Among these, 91 genes had 662 SNPs, which were mostly classified (98.3%, n=651) as missense (nonsynonymous) mutations. A subset of these missense mutations was found in the same genes in more than one of the three outlier genomes and encode for proteins involved in type VII and Sec-system secretion as well as pilus assembly (**Figure S6C**). Although most (77.6%, n=284) of these 366 genes have evidence of negative selection with dN/dS values < 1.0, 61 (16.7%) have evidence of positive selection with dN/dS values > 1.0 (**Table S7**). Despite the range in type and distribution of mutations that could be classified, the high frequency of mutations within postpartum strain 22.2 skewed the distribution of each type of mutation identified.

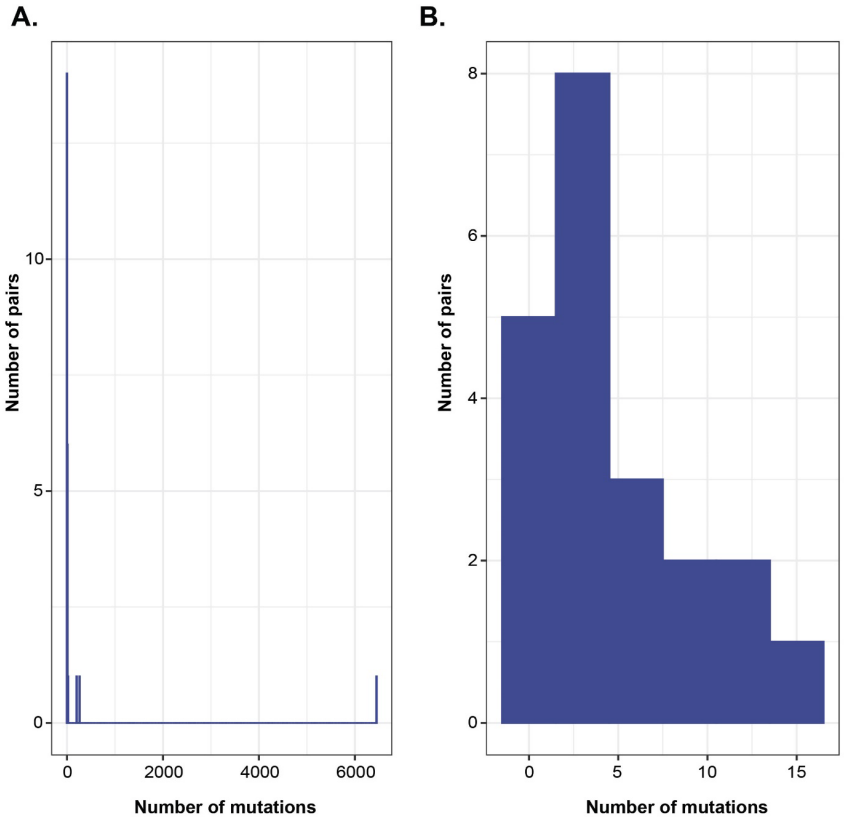


Figure 4. Number of core gene mutations in the postpartum GBS genomes with at least one mutation. Both plots show the number of mutations (x-axis) by the number of genomes (y-axis) for **A)** all 24 postpartum genomes, using a bin width, or interval size, of 10; and **B**) the subset of 21 postpartum genomes after removing three outliers with the greatest number of mutations with a bin width of 3. The eight genomes without any mutations were omitted from these plots.

277 **Increased mutation rates and mismatch repair mutations uncover "mutator" strains**

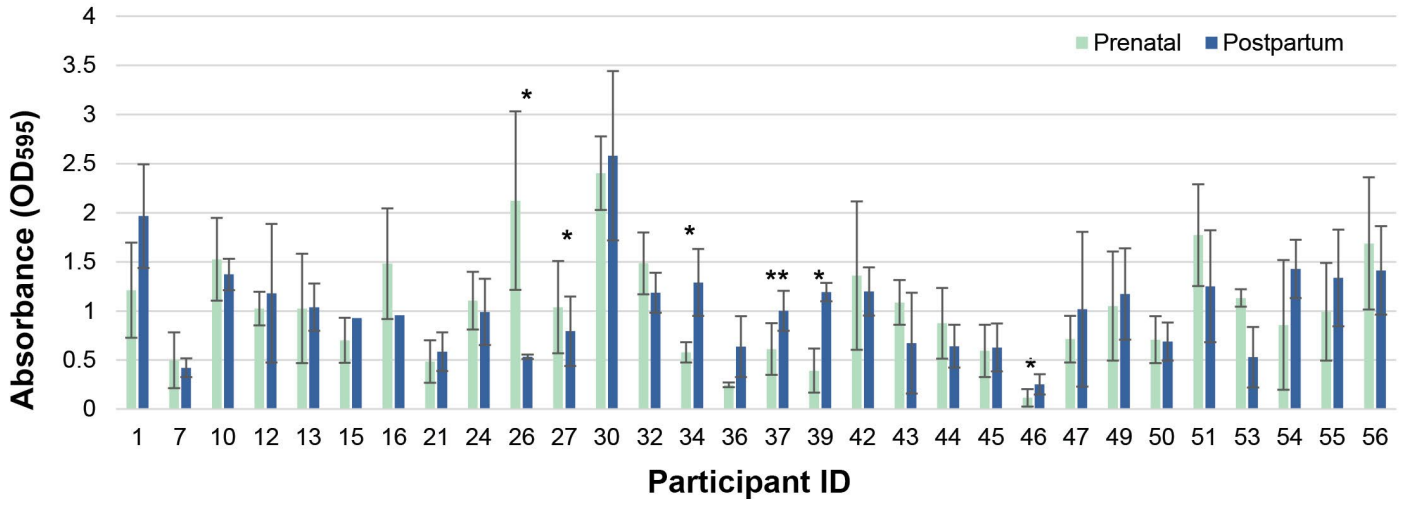
278 Given the wide range and frequency of mutations observed, we calculated the mutation rates
279 based on the total assembly length per genome for comparison. The mutation rate, which
280 represents the number of mutations per number of base pairs (bp), was determined for each
281 postpartum genome assembly relative to its respective paired prenatal genome assembly.
282 Notably, the average mutation rates for the three outlier genomes were 9.37×10^{-5} (ID=36.2),
283 1.32×10^{-4} (ID=43.2) and 0.003 (ID=22.2) mutations/bp, which is 53-, 75-, and 1,747-fold higher,
284 respectively, than the average mutation rate (1.77×10^{-6} mutations/bp) for the rest of the 29
285 genomes including those that lacked any detectable mutations (**Table S5**).
286
287 Among the three persistently colonized participants (IDs 50, 55, and 56) who did not receive
288 IAP, the average mutation rate was 1.50×10^{-6} mutations/bp, which is 77.3 times lower than the
289 average rate (1.16×10^{-4}) for the 29 participants who were given IAP, and 6.5 times lower after
290 omitting the outlier postpartum genome (22.2) with 6,454 mutations. Although the difference in
291 means was not significant (Student's t-test p=0.13), the mutation rate for strains without IAP
292 exposure includes the postpartum strain from participant 56 (n=4 mutations), who was one of
293 the four individuals reporting antibiotic use (cefazolin) after childbirth. Nonetheless, these results
294 demonstrate variable mutation rates across strains and suggest that the higher-than-average
295 mutation rates observed in the three outlier postpartum genomes could be indicative of
296 "mutators" that arose following IAP and childbirth.
297
298 Indeed, the maximum likelihood core gene phylogeny provides support for mutators in two of
299 the three postpartum strains (Figure 1). Except for strain 22.2, which had the greatest number of
300 mutations, the other two outlier strains, 36.2 and 43.2, grouped together with their respective
301 prenatal genomes. The pair of strains from participant 36 share a node in the phylogeny, while
302 the prenatal strain 43.1 is on an adjacent ancestral branch to its paired postpartum strain 43.2.

303 These data suggest that the genetic backbones of these two outlier postpartum genomes are
304 highly similar to their respective prenatal genomes despite the high number (>200) of core gene
305 mutations detected. By contrast, the placement of the 22.2 postpartum genome in the
306 phylogeny relative to the prenatal 22.1 genome along with different resistance and virulence
307 gene profiles (Figures 2 and 3), and an unusually large number of mutations, suggests that
308 strain 22.2 is not a mutator. Rather, these data indicate that participant 22 likely acquired a new
309 strain of the same CC with a distinct *cps* type by the postpartum visit.
310
311 Since mutators arise at a low frequency in the population due to mutations within genes
312 encoding DNA replication and proofreading machinery, mismatch repair (MMR) pathways, and
313 the 8-oxo-dG (GO) system[29,30], we compared the genes involved in these pathways among the
314 prenatal and potential mutator strains from participants 36 and 43. The six MMR pathway genes
315 (7,467 bp region) were extracted from each genome and aligned using MUSCLE[31]. These
316 genes encode DNA mismatch repair proteins, MutS (*mutS*) and MutL (GBS_230), a cold shock
317 protein, an MFS transporter/LmrP homolog (GBS_436), the Holliday junction branch migration
318 protein (RuvA), and DNA-3-methyladenine glycosylase I (GBS_235). The postpartum 36.2
319 genome has a 33 bp insertion in *mutS* relative to its paired 36.1 prenatal genome as well as a
320 deletion in the DNA polymerase IV gene and a missense SNP in *mutY* (A/G-specific adenine
321 glycosylase) of the GO system. Comparatively, genome 43.2 has a missense SNP in the DNA
322 topoisomerase gene and a 1 bp deletion downstream of the MMR loci relative to its paired 43.1
323 strain. The MMR loci were identical in the remainder of the paired prenatal and postpartum
324 genomes. The increased mutation rates along with mutations in these key regions meet the
325 criteria[29,32] for classifying these two strains as “mutators”.
326
327 **Specific lineages are more likely to acquire mutations *in vivo***
328 To better understand which point mutations could alter protein functionality, we examined the

329 distribution of the 350 characterized SNPs in the 31 postpartum strains after omitting outlier
330 strain 22.2 from the analysis. While specific mutations were not associated with individual
331 lineages, CCs 19 (n=121) and 23 (n=199) had the greatest number of SNPs, though these
332 numbers are skewed by inclusion of the two mutator strains. After excluding the mutators from
333 the analysis, the CC-23 strains still had the most SNPs (n=21) followed by CCs 12 (n=15) and 1
334 (n=12), ranging from 0-10, 0-9, and 0-7 SNPs per strain, respectively (**Table S8**). By
335 comparison, the postpartum non-mutator strains of CCs 17 and 19 acquired fewer SNPs overall
336 with 0-1 and 0-3 SNPs per strain, respectively. Among the three SNPs identified in the six CC-
337 17 postpartum strains, only one was nonsynonymous, indicating a pattern of negative or
338 purifying selection for strains in this lineage. The non-mutator strains of CCs 23, 12, and 1,
339 however, had twice as many nonsynonymous (n=32) SNPs as synonymous (n=16) SNPs,
340 suggesting a pattern of positive/diversifying selection and a trend of adaptive evolution for these
341 lineages.
342
343 Based on prior medical record reviews[23], the two mutator strains had been exposed to
344 penicillin, which was used for IAP in participants 36 and 43 during childbirth (**Table S8**). Among
345 the 16 participants given ampicillin and the 10 receiving penicillin, 28 and 310 SNPs were
346 detected in the postpartum strains, respectively. After excluding the two mutators, the number of
347 SNPs detected among the 8 participants who received penicillin decreased from 310 to 18
348 SNPs. While comparisons could not be made for the other antibiotics due to low numbers, the
349 three genomes from patients who did not receive IAP only had 1, 3, and 4 total mutations with
350 an average of 1 SNP/strain. The strain with 4 mutations, however, was exposed to cefazolin
351 after childbirth and prior to the postpartum sampling. Collectively, these data demonstrate that
352 certain strain types may be more susceptible to evolutionary pressures, which likely varies
353 depending on the antibiotic used for IAP.
354

355 **Persistent strains have enhanced biofilm production relative to prenatal strains**

356 Because biofilm formation is a common colonization strategy and mutations were found in
357 genes important for adherence, we examined this phenotype in the paired strains. An increase
358 in biofilm production was observed in slightly over half (54.8%, n=17) of the 31 postpartum
359 strains relative to each paired prenatal strain (**Figure 5**); four of these increases were
360 significant. Among the two mutators, one (36.2) had increased biofilm production and one (43.2)
361 had decreased production relative to the respective prenatal strains (**Figure S7**), though the
362 differences between both pairs were not significant. It is noteworthy that two postpartum isolates
363 from participants 26 and 27 had a significant decrease in biofilm formation relative to the
364 prenatal isolates. These isolates had acquired 16 and 2 mutations, respectively, with some
365 occurring in genes encoding binding and surface proteins such as penicillin-binding protein
366 Pbp2B, capsule protein CpsC, and an LPXTG anchor domain surface protein.



367
368
369 **Figure 5. Biofilm formation between prenatal and postpartum strains from 30 participants**
370 **with persistent colonization.** Biofilm formation was assessed across 30 paired strains by
371 measuring the absorbance (OD_{595}) values (y-axis) and averaging the biological replicates (n=3).
372 Prenatal (mint green) and postpartum (dark blue) isolates are shown by participant ID (x-axis).
373 One-tailed, paired t-tests were calculated to identify significant differences in biofilm formation
374 between the paired strains (*=p <0.05, **=p<0.01). Each measurement was normalized to a
375 negative media control. Error bars display standard deviation.
376

377 **DISCUSSION**

378

379 Through this analysis, we have identified a range of genomic alterations in GBS strains
380 recovered after IAP and childbirth as compared to strains isolated during prenatal visits from the
381 same individuals. Although the pangenome analysis of 1,368 core genes uncovered five distinct
382 sequence clusters, most of the paired prenatal and postpartum strains with identical STs
383 clustered together in the phylogeny. This finding along with a comprehensive evaluation of point
384 mutations indicates that in most cases, the strains that persisted in each participant are highly
385 similar to those recovered prior to IAP and childbirth.

386

387 Although various antibiotic resistance mechanisms are known to enhance bacterial survival in
388 the presence of antibiotics, our assembly-based genomic analysis failed to detect newly
389 acquired resistance genes in the postpartum strains relative to each respective prenatal strain.
390 Hence, resistance to the antibiotics used for IAP is not a mechanism of persistent colonization in
391 this study population. Among the paired postpartum genomes from the 26 (83.9%) participants
392 who received β-lactam antibiotics for IAP, none had β-lactam resistance genes, which is not
393 surprising since GBS remains universally susceptible to β-lactam antibiotics[10]. Resistance to
394 MLS antibiotics is more common in GBS[10], yet the three (9.7%%) participants given clindamycin
395 for IAP had postpartum strains that were neither phenotypically resistant[23] or possessed
396 macrolide resistance genes. In a similar analysis, acquisition of specific virulence genes was
397 also not associated with persistence since distinct variably present genes were not detected in
398 the postpartum versus prenatal strains. While these findings indicate that horizontally acquired
399 resistance and virulence genes are not a mechanism of GBS adaptation in this patient
400 population, it is important to note that there are limitations associated with assembly-based
401 analyses. Genes labeled as absent, for instance, may have been present below the detection
402 threshold due to poor coverage or allelic variation, or may exist in a region that is more difficult

403 to sequence. Since the only discrepancies observed were in three pairs of strains that did not
404 cluster together in the core gene phylogeny (Figure 1), we expect that the determination of
405 presence versus absence of each gene was accurate for the remaining pairs.
406
407 The accumulation of point mutations appears to be more important for GBS adaptation than
408 horizontal gene transfer in this study population, as was indicated in our reads-based analysis of
409 the 1,368 core genes. Genomic mutations were identified in 74% (n=23) of the 31 postpartum
410 genomes that shared an ST and clustered together with their respective prenatal genomes in
411 the core gene phylogeny (Figure 1). While this core gene mutation analysis has limitations
412 including the possibility of sequencing errors, each mutation was manually confirmed to rule out
413 sequencing artifacts and was determined to be in an area of >10x read coverage. In all, 0 to 262
414 mutations were detected in the 31 postpartum genomes, though two genomes are outliers that
415 harbor 81.6% (n=466) of the 571 mutations detected (Table S5). Exclusion of these two outliers
416 uncovered 105 mutations in 29 genomes with a range of 0 to 16 mutations per genome. This
417 number is slightly larger than the total number of mutations (n=29) detected among paired
418 strains from 19 pregnant patients and their sick neonates in a smaller WGS study[33]. In the
419 paired neonatal strains, mutations were detected within *covR*, which encodes a two-component
420 response regulator linked to virulence[34], and other virulence-associated genes. Although *covR*
421 mutations were not observed in our postpartum genomes, several strains had mutations in
422 genes encoding transcriptional regulators and ABC-transporters as well as colonization factors
423 like type VII secretion proteins (*essC*)[35], serine-rich repeat (srr) glycoprotein adhesins[36], and
424 sortases (*srtC)* used for pilus formation[37]. It is therefore possible that these mutations enhance
425 fitness *in vivo* while promoting adaptation to the genitourinary tract. Although epistasis, or the
426 impact of bacterial genotype on the fitness effects of a mutation, has been described[38] and CCs
427 commonly linked to maternal colonization[16] had the most mutations (Table S8), future

428 experiments are needed to confirm the impact of each mutation on GBS fitness and survival in
429 different strain backgrounds.
430
431 Even though we did not test the functional impact of all mutations identified, we did observe
432 functional changes in biofilm production that could be linked to the mutation data. Despite the
433 wide range in biofilm production, which is consistent with studies showing differences across
434 lineages, pilus types[39], and culture conditions[40], half of the postpartum strains had an increase
435 in biofilm formation relative to the prenatal strains. In some strains with enhanced biofilm
436 production, point mutations were identified in genes linked to adherence or colonization. Strain
437 36.2, for instance, had mutations in *srtC* (class C sortase), Sec-system proteins (GtfA, Asp2),
438 and a serine-rich glycoprotein adhesin (*srr1*), which was associated with persistent vaginal
439 colonization in mice[41]. Mutations were also discovered in some isolates with decreased biofilm
440 production. The postpartum isolate with the largest significant decrease in biofilm production
441 has a missense SNP in the capsular polysaccharide biosynthesis protein gene, *cpsC*. Given the
442 tradeoff between capsule production and adherence, which has been described in *S.*
443 *pneumoniae*,[42] it is reasonable to speculate that this mutation may promote capsule production
444 while simultaneously blocking adhesins that may be required for aggregation. Additional
445 mutagenesis and functional studies as well as use of more sophisticated biofilm quantification
446 methods, however, are needed to confirm the impact of this and other mutations
447
448 The high frequency of mutations observed in the two outlier postpartum genomes, which have
449 204 and 262 mutations relative to their respective prenatal genomes, led to their classification
450 as mutators. Support for this classification comes from the placement of each pair in the core
451 gene phylogeny (Figure 1) as well as the increased mutation rates in both postpartum genomes
452 and the identification of mutations in genes encoding DNA replication and proofreading
453 machinery[29,32]. Mutators represent a subpopulation of bacterial cells with a higher-than-average

454 mutation rate[29] that typically arise in populations undergoing adaptation to stressful
455 conditions[43,44]. They serve as a reservoir of mutations that can enhance fitness along with the
456 genetic diversity of bacterial populations[46]. A study of *S. pneumoniae*, for instance, described
457 increased mutation rates after exposure to subinhibitory concentrations of penicillin[45], which
458 facilitated adaptation. While mutators have been described in *S. pneumoniae*[47] and other
459 streptococcal species (e.g., *Streptococcus pyogenes*[48], *S. mutans*[30], and *S. iniae*[49]), this study
460 is the first to describe mutators in GBS.
461
462 Importantly, mutators have been linked to persistent bacterial infections, such as those caused
463 by *Pseudomonas aeruginosa*[50] and *Staphylococcus aureus*[51] in the lungs of cystic fibrosis (CF)
464 patients. They arise at low frequencies due to DNA polymerase errors, dysfunctional
465 proofreading mechanisms, and/or failure of mismatch/error correcting machinery, including the
466 MMR pathway[29,32]. A prior study of pathogenic *Escherichia coli* and *Salmonella enterica*, for
467 example, described the incidence of mutators with defective *mutS* alleles to be 1%[52]; MutS is a
468 critical protein that recognizes mismatches[53]. Impairment of genes in the MMR pathway can
469 contribute to increased mutation rates, recombination,[52] and horizontal gene transfer, which can
470 subsequently enhance the spread of antibiotic resistance determinants[32]. In *S. aureus*,
471 hypermutable strains recovered from the CF lung had high frequencies of mutations in
472 macrolide target sites[51], highlighting their role in the emergence of macrolide resistance.
473 Besides promoting adaptation in the presence of antibiotics, many of the *S. aureus*
474 hypermutators also had mutations in *mutS*. Herein, both GBS mutators had mutations within or
475 near the MMR-encoding region, which could cause disruptions in the functionality of this system
476 as was suggested for *S. aureus*[51]. It is notable that both mutators belong to two common
477 lineages (CCs 19 and 23) associated with colonization during pregnancy[16]. Within the non-
478 mutator postpartum strains, those belonging to CCs 1 and 23 along with CC-12 had the most
479 SNPs overall. Nevertheless, we cannot rule out the possibility that the prenatal strains have

480 other unique traits that may have made them more prone to mutations in DNA repair systems.
481 Because only two mutators were identified in our study, more *in vitro* experiments and cohort
482 studies are needed to determine how GBS mutators emerge in different populations. GBS strain
483 traits (e.g., serotypes, CCs) are variable across geographic locations[16] while many countries
484 lack IAP protocols altogether[54]. Therefore, different selection pressures in distinct locations will
485 likely have an impact on the type and number of mutations identified in specific strain
486 populations.
487
488 Since a high frequency of nonsynonymous SNPs was detected in the postpartum genomes,
489 which could influence protein function, and multiple strains had mutations in the same genes, it
490 is likely that some genes evolve in parallel under selection, leading to enhanced fitness and
491 similar adaptations[57]. We cannot conclude, however, that the mutations arose solely due to IAP
492 exposure, as we are unable to compare the impact of IAP versus no IAP on genomic changes in
493 this study. IAP was the standard protocol for GBS-positive participants and hence, only three
494 participants with persistent colonization were not given IAP. Although the postpartum strains
495 from these three participants had a slightly lower average mutation rate (1.50×10^{-6}) than the
496 average of the 19 strains with IAP exposure (9.76×10^{-6}), the difference was not significant (t-
497 test, $p=0.24$). Since all participants did not have the same antibiotic exposures, this variation
498 could impact our selection measures and mutation rate estimates. Indeed, not all participants
499 were given the same antibiotic or received the same number of doses, and some reported
500 postpartum antibiotic use prior to the second sampling. We also could not account for other
501 factors (e.g., changes in the vaginal microbiome or host susceptibility, over-the-counter
502 medication use, complications during childbirth, diet, etc.) that may impact selection and
503 evolution *in vivo*. These limitations highlight the need for long-term studies with a larger and
504 more diverse patient population as well as more serially collected samples from participants
505 who receive the same IAP exposure (e.g., concentration, doses and type of antibiotic). This type

506 of sampling scheme could have helped determine whether ancestral intermediaries exist in the
507 evolution of strain 22.2, for instance, which shares an ST with the prenatal 22.1 strain but differs
508 in gene content. Alternatively, it could pinpoint the timing when a new strain type was acquired
509 after childbirth.
510
511 In all, this analysis highlights the need for future studies that will examine potential mechanisms
512 that could be linked to persistent colonization and to understand how mutators emerge in
513 populations under selection. Even though the biofilm assays revealed functional differences in
514 strains collected before and after IAP and childbirth, additional experiments are needed to
515 determine if the mutations identified in this study can impact this phenotype or others. The lack
516 of mechanistic confirmation is a limitation of our descriptive study and hence, we cannot confirm
517 a causal relationship between IAP and mutation rates. Larger prospective clinical studies are
518 critically important to determine how IAP and other selective pressures, which are encountered
519 by GBS during childbirth, contribute to mutation rates and persistence *in vivo* and to further
520 define their role in colonization and disease. Despite our findings, it is important to note that the
521 prophylactic use of antibiotics to protect newborns from invasive GBS infections far outweighs
522 the risks imposed by evolutionary events that may enhance colonization.
523
524
525 **METHODS**
526 **Bacterial isolate selection and characterization**
527 A subset of 97 GBS isolates recovered from vaginal-rectal swabs from 58 pregnant
528 participants[23] were included in this analysis. Isolates were recovered from the same participant
529 during a prenatal (35-37 weeks gestation) visit and during a postpartum visit, which took place
530 six weeks after childbirth[23]. Isolates were previously characterized by MLST, serotyping, and
531 cps typing[22]. Participants with GBS at both samplings were considered to have persistent

532 colonization resulting in 80 strains from 41 participants. Seventeen participants had GBS only at
533 the prenatal sampling and were considered to have lost the pathogen by the postpartum visit.
534
535 **Whole-genome sequencing (WGS)**
536 A set of 41 of the 97 isolates was sequenced previously and the raw reads were downloaded
537 from the NCBI Sequence Read Archive (SRA); the remaining 56 isolates were sequenced for
538 this study (**Table S2**). Following DNA extraction of GBS grown overnight in Todd-Hewitt broth
539 (THB) at 37°C + 5% CO_2 using the E.Z.N.A. (Omega Bio-tek Inc., Norcross, GA, USA) or
540 Wizard HMW (Promega, Madison, WI, USA) kits, libraries were constructed at the Michigan
541 Department of Health and Human Services using the Nextera XT library prep kit (Illumina, San
542 Diego, CA, USA). Sequencing was performed using the MiSeq (Illumina) with 2x250 bp paired
543 end reads. Raw reads were trimmed with Trimmomatic v0.39[55] using "gentle trimming"
544 parameters to trim adapters and remove sequences with average quality scores <15 or that
545 were <36 nucleotides long. The paired-end reads, or single-end reads for SRR517012 that was
546 sequenced using an unpaired method, were assessed for quality using FastQC v0.11.9
547 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/); high-quality genomes were
548 defined as those that failed <10% of quality modules.
549
550 Sequences were assembled *de novo* using SPAdes v3.13.1 (kmers 21, 33, 55, 77, 99, 127) with
551 mismatch error correction[56], and quality checked using QUAST v5.0.2[58] and MultiQC v1.15[59]
552 with default parameters. High-quality assemblies, which were considered if the N50 and L50
553 scores were >15,000 and <50, respectively, were annotated using Prokka v1.14.6 with the –
554 proteins option that allows for the use of a custom database[60]. This database was curated from
555 25 closed *S. agalactiae* genomes downloaded from NCBI to ensure proper annotations (**Table**
556 **S9**). STs were assigned using PubMLST[61], though previously determined STs[22] were used for

557 three strains with poor assembly quality for one of the seven MLST loci.

558

559 **Pangenome and phylogenetic analysis**

560 Roary v3.11.2 was used to create a multiFASTA alignment of core genes from the 92 high-
561 quality assemblies using the -i (blastp of 95%) and -e (PRANK aligner 170427) parameters[62].
562 Maximum likelihood phylogenies were generated with RAxML v8.2.12 (parameters -m
563 GTRGAMMA, 500 bootstrap replicates)[63]. The core-gene alignment was input into MEGA12,[24]
564 to construct a phylogeny using the Maximum Likelihood method and Tamura-Nei model[64] of
565 nucleotide substitutions, which used up to 4 parallel computing threads. The tree with the
566 highest log likelihood was used and the percentage of replicate trees was determined
567 adaptively. The initial tree was selected by choosing the tree with the superior log-likelihood
568 between a Neighbor-Joining (NJ) tree,[27] which uses a matrix of pairwise distances, and a
569 Maximum Parsimony (MP) tree with the shortest length among 10 MP tree searches.
570 Phylogenies were visualized and annotated using the Interactive Tree of Life v7 (IToL,
571 https://itol.embl.de/)[28]. Neighbor-net trees were generated in SplitsTree v4.19.0[25] and the PHI
572 test was used to detect recombination with a window size of 100 and a k of 1; a p-value of <0.05
573 was considered significant. Analysis of the six MMR loci was performed by manually extracting
574 them from the assemblies using Geneious v2022.2.2 (https://www.geneious.com/) followed by a
575 MUSCLE[31] nucleotide alignment.

576

577 **Gene and nucleotide-level analyses**

578 ARGs and virulence genes were extracted from the 92 assemblies with ABRicate
579 (https://github.com/tseemann/abricate) using five ARG databases (CARD[65], ARG-annot[66],
580 MEGARes[67], Resfinder[68], NCBI AMRFinderPlus[69]) and the virulence-finder-database (VFDB)[70],

581 respectively. A gene was considered present if it had ≥10% identity and ≥80% coverage and for
582 the ARGs, if it was detected in two or more databases.
583
584 Core gene mutations were extracted from the assemblies of each pair of strains recovered from
585 participants with persistent colonization (n=64, 32 pairs). For each pair, the assembled prenatal
586 genome was used as the reference, while the trimmed reads from the postpartum isolate were
587 mapped against the reference. Each core genome was interrogated for mutations including
588 SNPs, insertions, deletions, MNPs, and complex mutations using Snippy v4.6.0
589 (https://github.com/tseemann/snippy). For discrepancies in gene presence/absence within a pair
590 of strains, Geneious v2022.2.2 was used to confirm ≥15X read coverage and the up and
591 downstream regions were mapped with respect to a given gene to confirm location and
592 completeness of the gene within the contig. Core gene mutations were validated using a ≥10X
593 read coverage cutoff.
594
595 **Biofilm Assays**
596 Biofilm assays were performed in triplicate on 30 pairs of strains from persistently colonized
597 participants (n=60 isolates). Values were normalized to blank media controls as we described[39]
598 previously with the modification of stagnant growth for 24hrs. The magnitude (fold-change) of
599 the difference in biofilm formation was determined for each postpartum strain relative to each
600 respective prenatal strain.
601
602 **Data analysis and statistics**
603 Raw WGS output was managed in Microsoft Excel. R v4.1.2 (https://www.r-project.org/) and
604 RStudio v2022.07.1+554 (https://www.rstudio.com/) was used for importing data (readr v2.1.4),
605 data wrangling (devtools v2.4.5, dplyr v1.1.3, forcats v1.0.0, plyr v1.8.8, tidyr v1.3.0, tidyverse

606 v2.0.0), and visualization (ggplot2 v3.4.3, viridis v0.6.4) (https://cran.r-project.org/). R package
607 factoextra v1.0.7 was used to perform and visualize PCA with 95% confidence ellipses for
608 virulence genes using the prcomp() and fviz_pca() functions (https://cran.r-project.org/). Chi-
609 square tests were performed in Microsoft Excel or Epi Info™ v7 (Centers for Disease Control
610 and Prevention, Atlanta, GA, USA); statistically significant associations were classified at $p \leq$
611 0.05. For unnamed genes that acquired mutations, arbitrary "gene ids" using the GBS_00X
612 notation were assigned for easier reference (**Table S6**). For all SNP-only analyses, SNPs were
613 filtered to exclude those that occurred in rRNA and/or uncharacterized, non-coding regions.
614
615 As characterized by Snippy, nonsynonymous SNPs included those classified as "initiator codon
616 variant", "missense", "start lost", "stop lost", and "stop gained" (nonsense), while synonymous
617 SNPs included those referred to as "stop retained" and "synonymous variant". Measures of
618 evolutionary selection were calculated based on the nonsynonymous to synonymous SNP ratio
619 ($\pi N/\pi S$) across the genomes, as well as across each gene of interest to discern areas under
620 positive ($\pi N/\pi S > 1$), neutral ($\pi N/\pi S = 1$), or negative ($\pi N/\pi S < 1$) selection. Calculations for
621 evolutionary selection across genes were limited to genes that acquired four or more SNPs.
622 Student t-tests were used to analyze differences in biofilm formation between isolate pairs and
623 mutation rates across CCs; $p<0.05$ was deemed significant.
624
625 **Data availability**
626 All datasets are available in the supplementary tables and raw sequencing reads for each
627 genome have been deposited in GenBank under several BioProjects listed in **Table S2** unless
628 they were uploaded previously. R scripts are at: https://github.com/macyEpell and via Zenodo
629 (https://doi.org/10.5281/zenodo.16413547).
630

631

632 **Ethics statement**

633 The parent study was conducted in November 1998-April 2000 and involved obtaining informed
634 consent from each patient prior to enrollment. It was approved by the Conjoint Medical Ethics
635 Committee of the University of Calgary and the Ethics Board of the participating University of
636 Alberta Hospitals.

637

638 **Acknowledgements**

639 We thank the Research Technology Support Facility and Michigan Department of Agriculture
640 and Rural Development for DNA sequencing, as well as the National Center for *Streptococcus*
641 (Edmonton, Alberta), particularly Marguerite Lovgren, for strain collection and serotyping.
642 Financial support was provided by the Michigan State University (MSU) Research Foundation
643 (to SDM) and in part, by the Michigan Sequencing and Academic Partnerships for Public Health
644 Innovation and Response (MI-SAPPHIRE) initiative at the MDHHS via CDC through the
645 Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious
646 Diseases Enhancing Detection Expansion program (6NU50CK000510-02-07; to HMB and
647 SDM). Salary support was provided by the National Institutes of Health (R01HD090061 to JAG
648 and SDM, R01AI134036 to JAG), Department of Veterans Affairs Merit Award I01BX005352
649 (Office of Research) to JAG, and the U.S. Department of Agriculture (2019-67017-29112) to
650 SDM. Student support was provided to MEP by the MSU Department of MGI via the Philipp and
651 Vera Gerhardt Travel and Ralph Evans Awards and the MSU College of Natural Science via two
652 Outstanding Scholar Fellowships.

653

654 **Author contributions**

655 Conceptualization: MEP, SDM; methodology: HMB, MEP; validation and formal analysis: MEP,

SDM; resources: SDM, HMB; supervision and insight: SDM, HMB, JAG, HDD; data curation: MEP, SDM; writing—original draft preparation, MEP; writing—review and editing and data visualization: MEP, SDM; funding acquisition: SDM, HMB. All authors have read and agreed to the final version of the manuscript.

Competing interests

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

REFERENCES

1. Russell, N.J., Seale, A.C., O'Driscoll, M., O'Sullivan, C., Bianchi-Jassir, F., Gonzalez-Guarin, J., Lawn, J.E., Baker, C.J., Bartlett, L., Cutland, C., et al. (2017). Maternal colonization with group B *Streptococcus* and serotype distribution worldwide: Systematic review and meta-analyses. Clin Infect Dis *65 (suppl_2)*, S100-S111.
2. Boyer, K.M., and Gotoff, S.P. (1986). Prevention of early-onset neonatal group B streptococcal disease with selective intrapartum chemoprophylaxis. N Engl J Med *314*, 1665–1669.
3. Verani, J.R., McGee, L., Schrag, S.J. (2010). Prevention of perinatal group B streptococcal disease--revised guidelines from CDC, 2010. MMWR Recomm Rep *59*, 1–36.
4. Phares, C.R., Lynfield, R., Farley, M.M., Mohle-Boetani, J., Harrison, L.H., Petit, S., Craig, A.S., Schaffner, W., Zansky, S.M., Gershman, K., et al. (2008). Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005. JAMA *299*, 2056-65.
5. Seale, A.C., Blencowe, H., Bianchi-Jassir, F., Embleton, N., Bassat, Q., Ordi, J., MenCrossed, C., Cutland, C., Briner, C., Berkley, J.A., et al. (2017). Stillbirth with Group B *Streptococcus* disease worldwide: Systematic review and meta-analyses. Clin Infect Dis *65* suppl_2, S125–S132
6. Schuchat, A. (1998). Epidemiology of group B streptococcal disease in the United States: Shifting paradigms. Clin Microbiol Rev *11*, 497-513
7. King, T.L., Silverman, N.S., and Turrentine, M. (2020). Prevention of group B streptococcal early-onset disease in newborns: ACOG Committee Opinion, Number 797. Obstet Gynecol *135*, e51–e72.
8. Schrag, S.J., Zywicki, S., Farley, M.M., Reingold, A.L., Harrison, L.H., Lefkowitz, L.B., Hadler, J.L., Danila, R., Cieslak, P.R., and Schuchat, A. (2000). Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis. N Engl J Med *342*, 15–20.
9. Azad, M.B., Konya, T., Persaud, R.R., Guttman, D.S., Chari, R.S., Field, C.J., Sears, M.R., Mandhane, P.J., Turvey, S.E., Subbarao, P., et al. (2016). Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: A prospective cohort study. BJOG *123*, 983–993
10. Schrag, S.J., and Verani, J.R. (2013). Intrapartum antibiotic prophylaxis for the prevention of perinatal group B streptococcal disease: Experience in the United States and implications for a potential group B streptococcal vaccine. Vaccine *31*, D20–D26.
11. Anthony, B.F., Okada, D.M., and Hobel, C.J. (1979). Epidemiology of the group B *Streptococcus*: Maternal and nosocomial sources for infant acquisitions. J Pediatr *95*, 431-436
12. Kwatra, G., Cunnington, M.C., Merrall, E., Adrian, P. V., Ip, M., Klugman, K.P., Tam, W.H., and Madhi, S.A. (2016). Prevalence of maternal colonisation with group B *Streptococcus*: A systematic review and meta-analysis. Lancet Infect Dis *16*, 1076–1084.
13. Van Dyke, M.K., Phares, C.R., Lynfield, R., Thomas, A.R., Arnold, K.E., Craig, A.S., Mohle-Boetani, J., Gershman, K., Schaffner, W., Petit, S., et al. (2009). Evaluation of universal antenatal screening for group B *Streptococcus*. Obstet Gynecol Surv *64*, 703–704.

14. Jones, N., Bohnsack, J.F., Takahashi, S., Oliver KA, Chan MS, Kunst F, Glaser P, Rusniok C, Crook DW, Harding RM, et al. (2003). Multilocus sequence typing system for group B *Streptococcus*. J Clin Microbiol *41*, 2530–2536.
15. Manning, S.D., Springman, A.C., Lehotzky, E., Lewis, M.A., Whittam, T.S., and Davies, H.D. (2009). Multilocus sequence types associated with neonatal Group B streptococcal sepsis and meningitis in Canada. J Clin Microbiol *47*, 1143–1148..
16. Bianchi-Jassir, F., Paul, P., To, K.N., Carreras-Abad, C., Seale, A.C., Jauneikaite, E., Madhi, S.A., Russell, N.J., Hall, J., Madrid, L., et al. (2020). Systematic review of Group B Streptococcal capsular types, sequence types and surface proteins as potential vaccine candidates. Vaccine *38*, 6682-6694
17. Springman, A.C., Lacher, D.W., Wu, G., Milton, N., Whittam, T.S., Davies, H.D., and Manning, S.D. (2009). Selection, recombination, and virulence gene diversity among group B streptococcal genotypes. J Bacteriol *191*, 5419-27
18. Korir, M.L., Knupp, D., LeMerise, K., Boldenow, E., Loch-Caruso, R., Aronoff, D.M., and Manning, S.D. (2014). Association and virulence gene expression vary among serotype III group B *Streptococcus* isolates following exposure to decidual and lung epithelial cells. Infect Immun *82*, 4587-95
19. Korir, M.L., Laut, C., Rogers, L.M., Plemmons, J.A., Aronoff, D.M., and Manning, S.D. (2017). Differing mechanisms of surviving phagosomal stress among group B *Streptococcus* strains of varying genotypes. Virulence *8*, 924–937.
20. Korir, M.L., Doster, R.S., Lu, J., Guevara, M.A., Spicer, S.K., Moore, R.E., Francis, J.D., Rogers, L.M., Haley, K.P., Blackman, A., et al. (2022). *Streptococcus agalactiae cadD* alleviates metal stress and promotes intracellular survival in macrophages and ascending infection during pregnancy. Nat Commun *13*, 5392.
21. Flaherty, R.A., Aronoff, D.M., Gaddy, J.A., Petroff, M.G., and Manning, S.D. (2021). Distinct Group B *Streptococcus* sequence and capsule types differentially impact macrophage stress and inflammatory signaling responses. Infect Immun. *89*, e00647-20
22. Manning, S.D., Lewis, M.A., Springman, A.C., Lehotzky, E., Whittam, T.S., and Davies, H.D. (2008). Genotypic diversity and serotype distribution of group B *Streptococcus* isolated from women before and after delivery. Clin Infect Diseases *46*, 1829–1837.
23. Spaetgens, R., DeBella, K., Ma, D., Robertson, S., Mucenski, M., and Dele Davies, H. (2002). Perinatal antibiotic usage and changes in colonization and resistance rates of group B *Streptococcus* and other pathogens. Obstet Gynecol *100*, 525–533.
24. Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., and Tamura, K. (2024). MEGA12: Molecular Evolutionary Genetic Analysis version 12 for adaptive and green computing. Mol Biol Evol *41*, msae263
25. Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol Biol Evol *23*, 254-267
26. Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution *39*, 783-791
27. Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol *4*, 406-425.
28. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res *52*, W78–W82.
29. Miller, J.H. (1996). Spontaneous mutators in bacteria: Insights into pathways of mutagenesis and repair. Annu Rev Microbiol *50*, 625–643.

30. Gonzalez, K., Faustoferri, R.C., and Quivey, R.G. (2012). Role of DNA base excision repair in the mutability and virulence of *Streptococcus mutans*. Mol Microbiol *85*, 361–377.
31. Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792–1797.
32. Chopra, I., O'Neill, A.J., and Miller, K. (2003). The role of mutators in the emergence of antibiotic-resistant bacteria. Drug Resist Updat *6*, 137-45
33. Almeida, A., Villain, A., Joubrel, C., Touak, G., Sauvage, E., Rosinski-Chupin, I., Poyart, C., and Glaser, P. (2015). Whole-genome comparison uncovers genomic mutations between group B streptococci sampled from infected newborns and their mothers. J Bacteriol *197*, 3354–3366.
34. Lamy, M.C., Zouine, M., Fert, J., Vergassola, M., Couve, E., Pellegrini, E., Glaser, P., Kunst, F., Msadek, T., Trieu-Cuot, P., et al. (2004). CovS/CovR of group B *Streptococcus*: A two-component global regulatory system involved in virulence. Mol Microbiol *54*, 1250-1268
35. Job, A.M., Doran, K.S., and Spencer, B.L. (2024). A group B streptococcal type VII-secreted LXG toxin mediates interbacterial competition and colonization of the murine female genital tract. mBio *15*, e0208824
36. Samen, U., Eikmanns, B.J., Reinscheid, D.J., and Borges, F. (2007). The surface protein Srr-1 of *Streptococcus agalactiae* binds human keratin 4 and promotes adherence to epithelial HEp-2 cells. Infect Immun *75*, 5405-5414
37. Konto-Ghiorghi, Y., Mairey, E., Mallet, A., Duménil, G., Caliot, E., Trieu-Cuot, P., and Dramsi, S. (2009). Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*. PLoS Pathog *5*, e1000422
38. Wong, A. (2017). Epistasis and the evolution of antimicrobial resistance. Front Microbiol *8*, 246
39. Parker, R.E., Laut, C., Gaddy, J.A., Zadoks, R.N., Davies, H.D., and Manning, S.D. (2016). Association between genotypic diversity and biofilm production in group B *Streptococcus*. BMC Microbiol *16*, 1–12.
40. Silvestre, I., Borrego, M.J., and Jordão, L. (2020). Biofilm formation by ST17 and ST19 strains of *Streptococcus agalactiae*. Res Microbiol *171*, 311-318
41. Sheen, T.R., Jimenez, A., Wang, N.Y., Banerjee, A., van Sorge, N.M., and Doran, K.S. (2011). Serine-rich repeat proteins and pili promote *Streptococcus agalactiae* colonization of the vaginal tract. J Bacteriol *193*, 6834–6842.
42. Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Müller, E., and Rohde, M. (2005). Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infect Immun *73*, 4653-4667
43. Sniegowski, P.D., Gerrish, P.J., and Lenski, R.E. (1997). Evolution of high mutation rates in experimental populations of *E. coli*. Nature *387*, 703-705
44. Mehta, H.H., Prater, A.G., Beabout, K., Elworth, R.A.L., Karavis, M., Gibbons, H.S., and Shamoo, Y. (2019). The essential role of hypermutation in rapid adaptation to antibiotic stress. Antimicrob Agents Chemother *63*, e00744-19.
45. Henderson-Begg, S.K., Livermore, D.M., and Hall, L.M.C. (2006). Effect of subinhibitory concentrations of antibiotics on mutation frequency in *Streptococcus pneumoniae*. J Antimicrob Chemother *57*, 849–854.
46. Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., and Godelle, B. (1997). Role of mutator alleles in adaptive evolution. Nature *387*, 700-702
47. Samrakandi, M.M., and Pasta, F. (2000). Hyperrecombination in *Streptococcus pneumoniae* depends on an atypical *mutY* homologue. J Bacteriol *182*, 3353-3360

48. Scott, J., Thompson-Mayberry, P., Lahmamsi, S., King, C.J., and McShan, W.M. (2008). Phage-associated mutator phenotype in group A *Streptococcus*. J Bacteriol *190*, 6290–6301.
49. Rudenko, O., Engelstädter, J., and Barnes, A.C. (2020). Evolutionary epidemiology of *Streptococcus iniae*: Linking mutation rate dynamics with adaptation to novel immunological landscapes. Infect Genet Evol *85,* 104435
50. Oliver, A., Cantón, R., Campo, P., Baquero, F., and Blázquez, J. (2000). High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science *288*, 1251-1254
51. Prunier, A.L., Malbruny, B., Laurans, M., Brouard, J., Duhamel, J.F., and Leclercq, R. (2003). High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. J Infect Dis *187*, 1709-16
52. LeClerc, J.E., Li, B., Payne, W.L., and Cebula, T.A. (1996). High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science *274*, 1208-1211
53. Hsieh, P. (2001). Molecular mechanisms of DNA mismatch repair. Mutat Res *486*, 71-87
54. Le Doare, K., O'Driscoll, M., Turner, K., Seedat, F., Russell, N.J., Seale, A.C., Heath, P.T., Lawn, J.E., Baker, C.J., Bartlett, L., et al. (2017). Intrapartum antibiotic chemoprophylaxis policies for the prevention of group B streptococcal disease worldwide: Systematic review. Clin Infect Dis *65 (suppl_2)*, S143-S151
55. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
56. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol *19*, 455–477.
57. Bailey, S.F., Rodrigue, N., and Kassen, R. (2015). The effect of selection environment on the probability of parallel evolution. Mol Biol Evol *32*, 1436-1448
58. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.
59. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.
60. Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069
61. Jolley, K.A., Bray, J.E., and Maiden, M.C.J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res *3*, 1–20.
62. Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., and Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics *31*, 3691–3693.
63. Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics *21*, 456–463.
64. Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol *10*, 512-526

65. Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., et al. (2017). CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res *45*, D566–D573.
66. Gupta, S.K., Padmanabhan, B.R., Diene, S.M., Lopez-Rojas, R., Kempf, M., Landraud, L., and Rolain, J.M. (2014). ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother *58*, 212–220.
67. Doster, E., Lakin, S.M., Dean, C.J., Wolfe, C., Young, J.G., Boucher, C., Belk, K.E., Noyes, N.R., and Morley, P.S. (2020). MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res *48*, D561–D569.
68. Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., and Larsen, M.V. (2012). Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother *67*, 2640–2644.
69. Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.H., McDermott, P.F., et al. (2019). Validating the AMRFINder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob Agents Chemother *63*, 1–19.
70. Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: Hierarchical and refined dataset for big data analysis - 10 years on. Nucleic Acids Res *44*, D694–D697.