**A 15-year single-centre clinical and genomic analysis of late-onset Group B *Streptococcus* infection.**

Carl Britto MD PhD1, Lea Cavalli MSc2, Madikay Senghore PhD2, Alexander McAdam3, William Hanage2, Ying-Jie Lu4, Richard Malley4

1. Division of Critical Care, Department of Anaesthesiology, Critical Care and Pain Medicine

Boston Children's Hospital

2. Center for Communicable Disease Dynamics (CCDD), Harvard T.H. Chan School of Public Health

3. Infectious Diseases Diagnostic Laboratory, Department of Laboratory Medicine, Boston Children's Hospital

4. Division of Infectious Diseases, Boston Children's Hospital and Harvard Medical School

Keywords: Group B Streptococcus; GBS meningitis; Resistance genes; Surface protein; Virulence factors; Whole-genome sequencing.

ABSTRACT

**Background:** Group B Streptococcus (GBS) is a leading cause of infant meningitis and intensive care admission. The pathogenic determinants of severe disease, particularly in late-onset disease are poorly understood.

**Methods:** We used whole-genome sequencing to characterize invasive GBS isolates from patients admitted to Boston Children’s Hospital over a 15-year period and contextualized local isolates with globally representative isolates. The pathogen genomic determinants (surface proteins, virulence factors and AMR determinants) were then correlated with high resolution clinical, laboratory and demographic data from each patient using logistic regression using the Bonferroni correction to account for multiple testing..

**Results:** During the study period, GBS isolates and patient data from 87 patients. Overall, 43% of patients required ICU care and 29% had evidence of meningitis. Overall were Serotype III (N=48, 55.2% of all isolates) and CC17 (N=37, 47.4%) were the most predominant serotype and clonal complex, respectively. The distribution of serotypes was stable across the study period. There were 5 different serotypes (I-V), 6 clonal complexes (1,10,17,19,23 and 26), 13 sequence types and the molecular composition of these strains were similar in serotype distribution, clonal complex, surface proteins, virulence factors and antimicrobial resistance when with 600 strains from the US CDC and 1,200 other globally representative strains. Factors associated with ICU admission or meningitis included pathogen genes encoding pilus proteins pilA, pilB, pilC and pi2a, surface proteins srr1 and alp1 and srtc4 and laboratory parameters leukopenia and neutropenia. Genes conferring non-susceptibility to tetracyclines and aminoglycosides were identified in 86% and 5% of isolates respectively and genes encoding virulence factors varied from 100% (*acpC* and *cfb*) to 2% (*pi1B*)

**Conclusions:** The findings from this large whole-genome sequence of GBS isolates establish important baseline data required for further surveillance. Infants with this infection have high rates of ICU admissions. The pilus and surface proteins were significantly associated with severe clinical manifestations and thus may be potential targets for therapeutic and immunological interventions.

**Introduction**

*Streptococcus agalactiae* (Group B *Streptococcus* (GBS)) is a pathobiont that transitions from an asymptomatic vaginal or gastrointestinal carriage state to a virulent pathogen causing severe invasive infections.(1) It is a common etiological agent of sepsis and meningitis in neonates and infants globally.(1, 2) According to the US Centers for Disease Control and Prevention (CDC), 1 in 2,000 newborns in the US develop GBS infections, which can be associated with mortality or long-term health problems such as developmental delays, hearing loss, or cerebral palsy.(1–4)

Current global trends in the molecular epidemiology of GBS suggest that serotype III is the most common serotype associated with LOD, which often presents as meningitis. (1, 3, 5–7). However, higher-resolution associations between clinical features and infecting serotype are lacking. While epidemiological associations between serotype III and meningitis from a 30-year nationwide analysis in the Netherlands have been convincing, the reasons for neurotropism have not been clearly established (7, 8). From the pathogen perspective, many virulence factors have been described and hypervirulent clones have been discovered. Delineating determinants of GBS mucosal colonization and invasion is important for the development of immunopreventive strategies to control invasive GBS disease. Additionally, there is a need to identify and characterize novel virulence factors that may be involved in GBS pathogenesis and delineate clear associations between genomic features and clinical outcomes.

There is a substantial body of data from various regions describing long-term neurodevelopmental outcomes, highlighting the clinical burden of this disease.(1, 7, 9–16) However, there is a dearth of data detailing clinical characteristics in infants during the acute phase of infection. GBS infections in newborns can present as early-onset disease (EOD, defined as occurring within 7 days of birth) or late-onset disease (LOD) in those between a week and 90 days of age. Both conditions are associated with significant morbidity and mortality.

GBS is likely vertically acquired in EOD but horizontally acquired in other disease forms.(17) Globally, the rates of EOD account for 60-70% of neonatal GBS illness, however, in a few settings like the US where GBS screening and intrapartum antibiotic prophylaxis (IAP) are standard of care, the rates of EOD (incidence rates between 0.0–0.6%), are lower than that of LOD (incidence rate > 0.7%). (1, 5, 18) While there is currently no effective preventive strategy against LOD, there are a number of vaccine candidates being developed, and a correlate of protection is actively being investigated.

This study aimed to analyze the association of well-characterized clinical outcomes, including risk of ICU admission and meningitis, with high-resolution molecular characteristics of the pathogen.

**Methods**

***Ethics statement***

Ethical approval was obtained from the BCH IRB (R31579/CN007). The study adhered to the STROBE guidelines as well as ethical principles of the Helsinki declaration.(19)

***Data Collection***

The bacterial strains isolated from blood or CSF of patients for this study were obtained between 2007 and 2021 from Boston Children’s Hospital (BCH), a tertiary referral pediatric hospital in Boston, Massachusetts, with a catchment area comprising the majority of 72,000 sq miles of New England. The hospital receives a wide range of patients including neonates with suspected sepsis. A paucity of EOD cases was expected given the general epidemiology of EOD in the US and high rates of intrapartum screening and peripartum prophylaxis. Moreover, BCH is a stand-alone pediatric hospital with no labor and delivery services. Therefore this analysis focuses mainly on LOD and infections during later infancy. Each isolate was stored for whole-genome sequencing and the corresponding demographic, antimicrobial sensitivity profile, laboratory and clinical information of the patient was extracted from hospital records.

***Clinical data analysis***

Clinical and demographic variables of each patient were extracted from hospital electronic health records. Age, sex, length of stay, vitals, clinical features on presentation, hematological indices, biochemistry variables, medications, interventions and microbiology variables were recorded. Outcomes of interest included level of care (ICU vs. inpatient pediatric unit) and meningitis (diagnosed via CSF culture or pleocytosis). Because some children may have been admitted to the ICU as a precaution, we defined children as having been in intensive care if their stay in the ICU was over 48 hours. Data were then linked to pathogen genomic data for further downstream analysis.

***Pathogen genomic analysis***

By contextualising GBS genomes on a global scale, we attempted to ascertain whether there was a dominance of certain known virulent clones that could be associated with severity of clinical outcomes in our cohort. Kraken (v.1.0)(20) was first used to identify potential contamination with the miniKraken database (minikraken\_20171019\_8GB). Contamination was filtered out from samples with less than 90% of reads classifying as ‘*Streptococcus agalactiae’* (GBS). We eliminated 4 sequences since they had less than 50% of reads classified as ‘*Streptococcus agalactiae’* (GBS). GBS was isolated from both the blood and CSF in 8 patients, which yielded identical genome sequences. We therefore removed duplicate strains from the genomic analysis, leaving us with a total of 87 samples for analysis. (Figure 1)

For analysis of SNPs in GBS and to obtain a multi-sequence alignment, Illumina reads were mapped to the reference genome sequence of the GBS reference genome for an invasive human strain HU-GS5823 (GenBank accession: AP018935.1) using the RedDog (V1beta.10.3) mapping pipeline, available at [https://github.com/katholt/RedDog](about:blank). RedDog uses Bowtie (v2.2.9)(21) to map reads to the reference sequence; uses SAMtools (v1.3.1)(22) to identify SNPs with phred quality scores above 30; filters out those supported by <5 reads or with >2.5 times the average read depth (representing putative repeated sequences), or with ambiguous consensus base calls. For each SNP that passed these criteria in any one isolate, consensus base calls for the SNP locus were extracted from all genomes (ambiguous base calls and those with phred quality scores less than 20 were treated as unknowns and represented with a gap character). Unicycler (v0.4.8)(23) was used for genome assembly and the assembled genomes were then used to determine sequence types using the mlst tool ([v2.19.0](about:blank))(24) and serotypes using GBS-SBG.(25) Clonal complexes were identified from mlst profiles using the goeBURST tool.(26) Abricate (27) was used on assembled genomes to detect antimicrobial resistance (AMR) and virulence factor genes with the ARG-Annot(22), ResFinder(28) and vfdb(29) databases, respectively. The presence of the *sip* gene was identified using the Basic Local Alignment Search Tool (BLAST).(22) The presence or absence of surface proteins was detected using the *GBS\_Surface\_Typer.pl* script.(30) The BCH strains were then contextualized on a regional and global level using genome data of previously published strains from the US CDC (n=453) and other global cohorts (Netherlands (n=415), Kenya (n=69), Malawi (n=65), China (n=29), India (n=4), Southeast Asia (n=11), Canada (n=44)) to build a global phylogeny for phylogeographic comparisons (Supplementary Table 1). Maximum likelihood (ML) phylogenetic trees were inferred in RAxML (v8.2.12).(31) from the genome alignment. Inference was conducted under the generalized time-reversible model with a Gamma distribution (GTR+ Γ) to model site-specific rate variation and with 100 bootstrap replicates to support results. A *Streptococcus* pneumoniae strain (ENA accession: ERS812015) was used as an outgroup to root the inferred phylogenetic tree of the GBS isolates. The resulting tree was visualized and annotated using the R package ggtree.(32)

**Statistical analysis**

To identify correlates of severe manifestations of disease (ICU admission > 48 hours and/or meningitis), demographic data such as age and gender, clinical data such as presenting complaint and vitals on presentation, laboratory data such as hematology and inflammatory parameters, pathogen virulence factors, surface proteins, serotype, clonal complex and accessory genes were included as independent variables into the models. Meningitis (defined as GBS presence on CSF culture or pleocytosis on CSF cytology in the presence of GBS bacteremia) and ICU admission were the dependent variables. Logistic regression was applied as a statistical modeling technique for binary classification. By assuming a binomial distribution, logistic regression estimated the probabilities of the binary outcomes and utilized a logit transformation to linearly combine the predictors. We also accounted for multiple testing and used the Bonferroni correction where appropriate.

**Results**

***Invasive GBS surveillance***

Between 2007 and 2021 there were 96 clinical isolates collected from 87 patients with invasive GBS infection. Of these, 72 were isolated from blood, 6 from CSF and 9 from both. As seen in Figure 1A, there were 2 patients diagnosed within 7 days of birth (i.e. EOD), 48 diagnosed between 7 days and 3 months after birth (i.e. LOD), 20 between 3 and 12 months (also referred to as “Very Late Onset Disease”, or VLOD).Therefore, there was a total of 70 infants who had an invasive GBS infection. The other 17 clinical isolates were from older children (n=10) and adults (n=7). Over the duration of the study period, Serotype III was the predominant serotype in infants (n=45, 64.3%) followed by serotype Ia (n=15, 21.4%), hich was the most predominant among the older children (n = 5, 51%) (Figure 1). The distribution of serotypes across time between patients admitted to the ICU vs those not admitted to the ICU was similar as seen in Figure 1B. There were no isolates with capsular serotypes VI –IX and no non-typeable (NT) isolates.

***Clinical Characteristics***

There were no meaningful differences in clinical (age-corrected heart rate, respiratory rate, oxygen saturation or blood pressure on admission) or laboratory parameters between infants and older individuals on admission. (Table 1). The most common presenting complaint was “fussiness” and the most common clinical sign was fever. There were no deaths due to GBS infection during hospitalisation although there was significant morbidity based on 6 months of medical record data post-discharge associated with invasive GBS disease.

Among infants, 31 (44.3%) required ICU level of care due to risk of clinical decompensation, mental status changes or escalating respiratory support. Infants who required ICU stay were more likely to have leukopenia on admission as well as a lower absolute neutrophil count. In those requiring ICU care, the length of hospital stay was significantly longer with evidence of neurodevelopmental deficits/delays 6 months after initial diagnosis. Eighteen (25.7%) infants had meningitis.

***Molecular characteristics of the bacterial population***

The strains isolated from this characterised cohort of patients were contextualised on a national and global level as mentioned above. The SNP-based phylogenetic global tree in figure 2A highlights the interspersion of isolates from BCH with other isolates from USA and between geographies indicating a high degree of homogeneity within CCs. There was no clustering among the BCH isolates suggesting there was no unique expansion of virulent clones. In keeping with previous reports, (5, 7, 18, 33, 34) CC17 and CC23 were dominant (Supplementary Figure 1).

There were 6 distinct CCs that were apparent from the pathogen population in this cohort: CC1 (N=4, 4.6 %), CC10 (N=8, 9.2%), CC17 (N=41, 47.1%), CC19 (N=9, 10.3%), CC23 (N=20, 23.0%) and C26 (N=5, 5.8%). (Figure 2B). When grouped by Sequence Types (STs), 13 STs became apparent from this cohort, with ST-17 being the most common (n=36, 41.4%), followed by ST-23 (n=28, 32.2%) and ST-19 (n=6, 6.9%). Additionally, one isolate had a new MLST profile that was not reported in the pubMLST database: *adhP*(4) *pheS*(1) *atr*(4) *glnA*(4) *sdhA*(3) *glcK*(3) *tkt*(2). The isolate with the new ST differed from ST-8 by one housekeeping gene *glnA*, and from ST-10 by two housekeeping genes: *adhP* and *glnA*. The age distribution of patients infected with the various STs and unique genes in each ST are in supplementary table 3 and 4 respectively

Aside from ST-10, every ST was associated with a single serotype, and each CC was represented by a dominant serotype (Fig. 2B). CC17 was exclusively serotype III and CC23 was exclusively serotype Ia. Within CC19, ST-19 and ST-335 were serotype III while ST-28 was exclusively serotype II. Within CC1,  ST-1 was exclusively serotype V and ST-459 was exclusively serotype IV. Aside from one ST-10/cps II isolate, all CC10/12 isolates were serotype Ib. Despite being grouped with CC10/12 isolates based on e-burst clustering as well as phylogenetically, the isolate with the new ST was serotype Ia. As such, CC10/12 was the lineage exhibiting the most serotype diversity within our sample. A phylogenetic tree encompassing only strains from patients admitted to the ICU revealed a similar representation of CCs (Figure 2C).

***Coverage of genes encoding proteins of vaccine targets***

Next, we were interested in estimating the potential coverage of the two most advanced candidate GBS vaccines. One of these, a hexavalent vaccine targeting serotypes Ia, Ib, II, III, IV, and V (developed by Pfizer)(35) would theoretically cover 100% of cases (Figure 1). Another, being developed by Minervax (36) contains fusion proteins from the alpha-like protein (Alp) family of GBS. All isolates from infants in our study contained the genes encoding at least one Alp proteins:*rib* (n=46/70, 66%), *alphaC* (n=8, 9%), *alp1* (n=22, 25%) and *alp23* (n=7, 8%) (Figure 2B and 2C). The *rib* gene was exclusively present in CC17 (n=39/39) and CC19 (n=5/5) isolates, *alpha C* of CC12 (n=7/7), and *alp1* and *alp23* of CC23 (n=13/14 and n=1/14, respectively) and CC1 (n=2/3 and n=1/3, respectively). This protein-based vaccine could potentially cover all the strains isolated in our study, assuming that the vaccine technology used can stimulate a robust, protective antibody response. The distribution genes of potential vaccine antigens identified from isolates are stratified by age of patients they were identified in are in supplementary table 5.

All isolates cultured from infants encoded at least one of three pilus islands *PI-1* (n=52), *PI-2a* (n=69) and *PI-2b*  (n=38), as well as the *sip* gene. Most isolates possessed the genes encoding C5a peptidase (n=65), the Lmb protein (n=66) and genes for at least one of two serine rich proteins *srr1* and *srr2* (n=69). Finally, *fbsB* was encoded by all CC17 and CC23 isolates (n=54, 77%) but was lacking from all CC1, CC12, and CC19 isolates.

***Correlation between phenotypic antibiotic resistance and presence of known genotypic markers***

Phenotypic AMR testing revealed that all isolates were susceptible to penicillin, the current antimicrobial of choice for GBS infection, and vancomycin, the recommended antibiotic in cases of penicillin allergy and clindamycin-resistance. Thirty-three (38%) isolates were erythromycin-resistant, twenty-three (26%) isolates were clindamycin-resistant and three had intermediate clindamycin resistance (3%). Genes encoding resistance to macrolides (*ermA*, *ermB*, *ermT*, and *mefA-msrD*) were detected in all (33/33) phenotypically erythromycin-resistant isolates and 4/54 (7%) of  phenotypically erythromycin-susceptible isolates. Genes encoding resistance to lincosamides (*ermA*, *ermB*, and *ermT*) were detected in 23/23 (100%) , 1/3 (33%) and 7/61 (11%) phenotypically clindamycin-resistant, -intermediate and -susceptible isolates, respectively (Figure 3). Despite the overlap of phenotypic and genotypic resistance, the presence of a few false negative suggests a lack of sensitivity of phenotypic resistance testing. This limitations may notably be related to the subjectivity of Minimum Inhibitory Concentration (MIC) Breakpoints. This observation highlight the strengths of in-silico resistance gene detection, and emphasizes the ongoing necessity of performing phenotypic resistance typing and regularly updating online databases. to .

We further evaluated resistance to other antibiotics based on the detection of resistance genes or mutations only. Tetracycline resistance, the most commonly reported resistance phenotype for GBS, was predicted in 78 isolates (90%) from the presence of *tet-M* (n =76, 87%), *tet-O* (n =2, 2%) or both (n=6, 7%). The highest rate of tetracycline resistance was detected in CC1 (n=4/4, 100%), CC17 (n=39/40, 98%) and CC23 (n=17820, 90%) compared with other CCs (CC12: n=5/7, 71%,  CC459: n=2/4, 50%,  CC19: n=7/8, 88%). Aminoglycoside resistance was predicted from the presence of the *aph(3′)-III and ant(6)* resistance genes in 4 isolates (5%), all encoding both genes and belonging to CC17/ST-17. The distribution of resistance genes and corresponding resistance phenotypes by CC are described in Supplementary Table 5.  No SNP variants conferring reduced penicillin susceptibility were detected the rRNA 23-S1 and 23S-3 genes across our isolates..However, we detected*gyrA* or *parC* SNP variants typically conferring fluoroquinolone resistance in threeisolates: two CC12 isolates carried mutation S81L in GyrA and S79F in ParC and one CC17 isolate carried mutation S79F in ParC only.

It may be interesting to additionally mention whether the distribution of resistance phenotypes/genes changes over time. Notably since the antibiotic prophylaxis guidelines were updated in 2010, removing the recommendation to use Erythromycin in response to increasing resistance rates. It may additionally be interesting to check whether the distribution of resistance genes differs across disease severity groups.

***Identifying correlates of severe clinical outcomes***

Among infants, 32 (46.3 %) required ICU level due to risk of clinical decompensation, central nervous system changes and/or escalating respiratory support. As shown in Table 1, infants who required ICU stay were more likely to have leukopenia (X2=11.41 p <0.001) and neutropenia (X2=4.6, 0.03) while leukocytosis was associated with a lower risk of ICU admission (X2 = 8.8721, p =0.002). The length of stay was significantly longer in those needing ICU level of care (median duration of 14.7 vs 27.1 days p=0.003) and these children had developmental delay 6 months after initial diagnosis. Eighteen (29%) infants had evidence of meningitis which was associated with a significantly higher likelihood of neutropenia compared with those without meningitis. Risk of PICU admission was highly correlated with infection with isolates carrying pilus island genes (*pilB, pi2a, pi2b* *pilA*, *pilC*). Other genes significantly associated with ICU admission in a univariate and multivariate model included *alp1*, *srr1*, and *srtc4*.

**Discussion**

Based on epidemiologic evidence of neurotropism and pathogenicity of serotype III, we hypothesized that unique virulence factors encoded by genes in the pathogen would be significantly associated with clinically relevant outcomes. Data from different geographies consistently report that serotypes III and I dominate GBS distribution and over 90% of infant infections are attributed to five serotypes (Ia, Ib, II, III, V).(1–3, 7, 8) In our study serotype III was notably associated with a previously described hypervirulent CC17 clone (CC17A), and was the most prevalent serotype that was significantly associated with infant meningitis, in corroboration with multiple reports.(1, 7, 8, 18, 37, 38). The phase transition from carriage to pathogenicity is not clearly understood although multiple pathogenic attributes have been characterized in animal models. The fibrinogen-binding proteins Fbs (comprising FbsA, FbsB, FbsC, or BsaB), serine-rich repeat glycoproteins Srr1 and Srr2, the laminin-binding protein (Lmb), Streptococcal C5a peptidase from group B (ScpB), streptococcal fibronectin-binding protein A (SfbA), and the GBS immunogenic bacterial adhesin BibA are some of the well-established entities. (1, 7, 8, 18, 37, 38).

Pathogen genomic analysis from our study revealed that the overall prevalence of genes encoding virulence factors varied from 100% (CAMP factor, *b-hemolysin-cytolysin*, *acpC* and *cfb*) to 2% (*pi1B*). However, the main virulence factors that were statistically different between those with clinically significant outcomes (meningitis and ICU level of care > 48 hours) were pilus island genes (*pilB, pi2a,* *pilA*, *pilC*), *alp1*, *srr1*, and *srtc4*. Pili in GBS are primarily involved in epithelial cell colonization, biofilm formation, translocation, and invasion, with specific pilus types implicated in certain mechanisms. Pi-1 pili are involved in evasion of innate immunity mechanism.(39–42)

The presence of PI-2B plays a critical role in infection and penetration of the blood-brain barrier by enhancing intracellular pathogen survival in macrophages through conferring the ability to colonize epithelial cells, form biofilms, translocate and by significantly promoting an organism's invasiveness.(37) The evolution of pilus islands and adhesion appendages to confer a survival advantage could be one explanation for the predominance serotype III and the niche selection of CC17A. The isolates in our study also had a high proportion of genes expressing adhesin proteins which are key drivers of both colonisation and invasion. Lmb adhesin and HvgA, a cell wall-anchored protein, are specific for the hypervirulent clone CC-17A and shown to mediate meningeal tropism in murine models(38, 43).

Additionally, all isolates in this study were encapsulated. The serotype III capsule is known to confer virulence through resistance to complement and platelet-mediated antimicrobial killing and capsule expression is linked to invasion capabilities of GBS. Teatero *et al* observed that isolates cultured from colonisation specimens were less likely to be encapsulated compared with invasive isolates, pointing to the role of the capsule in invasiveness. (44) The interest in the pathogenicity and immunogenicity of the capsule revolves around its potential as an antigen in vaccine candidates. Ten capsular polysaccharide serotypes of group B streptococcus have been identified (Ia, Ib, and II-IX). Numerous sero-epidemiological investigations have revealed a correlation between the presence of an adequate level of trans-placentally transferred serotype-specific anti-CPS IgG antibodies and a decreased likelihood of contracting invasive group B streptococcal disease from the same serotype. A classical Phase 3 efficacy trial of a maternal GBS vaccine would be extremely difficult to conduct, given the very large sample size that would be required and the relatively low incidence of disease in areas in which such studies would be conducted. An alternative approach, currently being evaluated, involves the development of serological correlates of protection, which may ultimately allow for the licensure of a vaccine on the basis of immunogenicity alone. However, heterogeneity in study designs and serologic assays of binding antibodies have hindered the establishment of a universally accepted threshold for protective antibody concentrations. The identification of a serological correlate will require rigorously profiled clinical cohorts of patients to identify immune correlates, and high resolution insight of the infecting serotype of GBS to appropriately parametrize serological responses to antigens of interest. Clinically relevant outcomes of interest should also be determined to set appropriate protective immunologic thresholds for the clinical outcome in question. For instance, if a vaccine or immunotherapy can provide an immunologic response that is correlated with prevention of an ICU admission even if not with sterilizing immunity, it will still be a useful intervention in decreasing the morbidity of illness. From an alternate perspective, given that GBS isolates are still largely susceptible to antibiotics, early identification of infection and prompt initiation of therapy and a better understanding of clinical trajectories are therefore crucial in identifying ways to reduce the burden of GBS sepsis in infants.

In this study, infants requiring ICU care exhibited distinct laboratory characteristics, including leukopenia and lower absolute neutrophil counts on admission. The association between ICU admission and prolonged hospital stays, as well as neurodevelopmental deficits/delays from multiple sources in literature as well as in our study, underscores the significant morbidity associated with invasive GBS disease in this population. The identification of distinct laboratory characteristics of infants in this study highlights the complex interplay between host factors and pathogen virulence in determining disease severity and outcomes.

The underlying immune status, possible delay in initiating antimicrobial therapy and other co-morbidities also play a role in clinical manifestations. There may also be particularly virulent strains that are responsible of worse clinical outcomes. As seen in Figure 2A, the isolates from this study which integrated well in the global phylogenetic tree, indicating strong homogeneity among the global cohort. While clinical data such as neutropenia were associated with higher risk of ICU admission and meningitis, the mechanistic host-pathogen interactions leading to neutropenia remain to be determined. Clinical markers such as neutropenia could be used for early triage and treatment decisions. Exploring maternal antibody transfer and transplacental antibody effects could uncover immunological targets for prevention and treatment. Studying disease clinical characteristics alongside pathogen genomic determinants opens avenues for testing hypotheses related to disease pathogenesis, immunotherapeutic targets, and measurable intervention outcomes.

There are several limitations to this study. This study was retrospective in nature and therefore clinical findings were gathered from chart review which is subject to significant heterogeneity in assessment and decision-making including disposition to the ICU vs the pediatric ward. The hospital is a tertiary medical center dedicated to the care of children with no attached obstetric services and therefore findings and standard of care here may not be a reliable reflection of other healthcare centers.

In conclusion, this study confirms the prevalence of serotype III in infants (LOD) and its association with severe disease manifestations. It also delves into the genomic diversity of GBS, highlighting key virulence factors and their contributions to the pathogenesis of meningitis. The findings underscore the importance of a GBS vaccine for infant protection and the potential of various vaccine candidates to address this public health concern and by extension decrease the risk of ICU admission.

**Supplementary and higher resolution figures**

<https://docs.google.com/document/d/1Y-mckZWiVTo5yWwCyq8CjiAGztBu9QLWcTMwgzQxMXQ/edit>

**References**

1. Shabayek S, Spellerberg B. Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology. *Front Microbiol* 2018;9:.

2. Group B Streptococcus infection causes an estimated 150,000 preventable stillbirths and infant deaths every year. at <https://www.who.int/news/item/05-11-2017-group-b-streptococcus-infection-causes-an-estimated-150-000-preventable-stillbirths-and-infant-deaths-every-year>.

3. Group B Strep: Fast Facts and Statistics | CDC. at <https://www.cdc.gov/groupbstrep/about/fast-facts.html>.

4. McGee L, Chochua S, Li Z, Mathis S, Rivers J, Metcalf B, *et al.* Multistate, Population-Based Distributions of Candidate Vaccine Targets, Clonal Complexes, and Resistance Features of Invasive Group B Streptococci within the United States, 2015-2017. *Clin Infect Dis* 2021;72:1004–1013.

5. Phares CR, Lynfield R, Farley MM, Mohle-Boetani J, Harrison LH, Petit S, *et al.* Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005. *JAMA* 2008;299:2056–2065.

6. Madhi SA, Izu A, Kwatra G, Jones S, Dangor Z, Wadula J, *et al.* Association of Group B Streptococcus (GBS) Serum Serotype-Specific Anticapsular Immunoglobulin G Concentration and Risk Reduction for Invasive GBS Disease in South African Infants: An Observational Birth-Cohort, Matched Case-Control Study. *Clin Infect Dis* 2021;73:E1170–E1180.

7. van Kassel MN, de Boer G, Teeri SAF, Jamrozy D, Bentley SD, Brouwer MC, *et al.* Molecular epidemiology and mortality of group B streptococcal meningitis and infant sepsis in the Netherlands: a 30-year nationwide surveillance study. *The Lancet Microbe* 2021;2:e32–e40.

8. Jamrozy D, Bijlsma MW, de Goffau MC, van de Beek D, Kuijpers TW, Parkhill J, *et al.* Increasing incidence of group B streptococcus neonatal infections in the Netherlands is associated with clonal expansion of CC17 and CC23. *Sci Reports 2020 101* 2020;10:1–13.

9. Horváth-Puhó E, van Kassel MN, Gonçalves BP, de Gier B, Procter SR, Paul P, *et al.* Mortality, neurodevelopmental impairments, and economic outcomes after invasive group B streptococcal disease in early infancy in Denmark and the Netherlands: a national matched cohort study. *Lancet Child Adolesc Heal* 2021;5:398–407.

10. Bramugy J, Mucasse H, Massora S, Vitorino P, Aerts C, Mandomando I, *et al.* Short- and Long-term Outcomes of Group B Streptococcus Invasive Disease in Mozambican Children: Results of a Matched Cohort and Retrospective Observational Study and Implications for Future Vaccine Introduction. *Clin Infect Dis* 2021;74:S14–S23.

11. Chandna J, Liu WH, Dangor Z, Leahy S, Sridhar S, John HB, *et al.* Emotional and Behavioral Outcomes in Childhood for Survivors of Invasive Group B Streptococcus Disease in Infancy: Findings From 5 Low- and Middle-Income Countries. *Clin Infect Dis* 2022;74:S35–S43.

12. Paul P, Chandna J, Procter SR, Dangor Z, Leahy S, Santhanam S, *et al.* Neurodevelopmental and growth outcomes after invasive Group B Streptococcus in early infancy: A multi-country matched cohort study in South Africa, Mozambique, India, Kenya, and Argentina. *EClinicalMedicine* 2022;47:101358.

13. Harden LM, Leahy S, Lala SG, Paul P, Chandna J, Lowick S, *et al.* South African Children: A Matched Cohort Study of Neurodevelopmental Impairment in Survivors of Invasive Group B Streptococcus Disease Aged 5 to 8 Years. *Clin Infect Dis* 2022;74:S5–S13.

14. John HB, Arumugam A, Priya M, Murugesan N, Rajendraprasad N, Rebekah G, *et al.* South Indian Children’s Neurodevelopmental Outcomes After Group B Streptococcus Invasive Disease: A Matched-Cohort Study. *Clin Infect Dis* 2022;74:S24–S34.

15. Gonçalves BP, Procter SR, Paul P, Chandna J, Lewin A, Seedat F, *et al.* Group B streptococcus infection during pregnancy and infancy: estimates of regional and global burden. *Lancet Glob Heal* 2022;10:e807–e819.

16. Bramugy J, Mucasse H, Massora S, Vitorino P, Aerts C, Mandomando I, *et al.* Short- and Long-term Outcomes of Group B Streptococcus Invasive Disease in Mozambican Children: Results of a Matched Cohort and Retrospective Observational Study and Implications for Future Vaccine Introduction. *Clin Infect Dis An Off Publ Infect Dis Soc Am* 2022;74:S14.

17. Furuta A, Brokaw A, Manuel G, Dacanay M, Marcell L, Seepersaud R, *et al.* Bacterial and Host Determinants of Group B Streptococcal Infection of the Neonate and Infant. *Front Microbiol* 2022;13:.

18. Nanduri SA, Petit S, Smelser C, Apostol M, Alden NB, Harrison LH, *et al.* Epidemiology of Invasive Early-Onset and Late-Onset Group B Streptococcal Disease in the United States, 2006 to 2015: Multistate Laboratory and Population-Based Surveillance. *JAMA Pediatr* 2019;173:224.

19. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–2194.

20. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014;15:1–12.

21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods 2012 94* 2012;9:357–359.

22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.

23. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:.

24. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;11:.

25. Tiruvayipati S, Tang WY, Barkham TMS, Chen SL. GBS-SBG - GBS Serotyping by Genome Sequencing. *Microb Genomics* 2021;7:688.

26. Francisco AP, Bugalho M, Ramirez M, Carriço JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 2009;10:1–15.

27. GitHub - tseemann/abricate: :mag\_right: Mass screening of contigs for antimicrobial and virulence genes. at <https://github.com/tseemann/abricate>.

28. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67:2640.

29. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. *Nucleic Acids Res* 2016;44:D694–D697.

30. GBS\_Scripts\_Reference/GBS\_Surface\_Typer.pl at master · BenJamesMetcalf/GBS\_Scripts\_Reference · GitHub. at <https://github.com/BenJamesMetcalf/GBS\_Scripts\_Reference/blob/master/GBS\_Surface\_Typer.pl>.

31. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.

32. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36.

33. Ji W, Zhou H, Li J, Britto CD, Liu Z, Zhang W, *et al.* Distributions of candidate vaccine Targets, virulence Factors, and resistance features of invasive group B Streptococcus using Whole-Genome Sequencing: A Multicenter, population-based surveillance study. *Vaccine* 2024;42:3564–3571.

34. Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, Davies HD. Multilocus sequence types associated with neonatal Group B streptococcal sepsis and meningitis in Canada. *J Clin Microbiol* 2009;47:1143–1148.

35. Absalon J, Segall N, Block SL, Center KJ, Scully IL, Giardina PC, *et al.* Safety and immunogenicity of a novel hexavalent group B streptococcus conjugate vaccine in healthy, non-pregnant adults: a phase 1/2, randomised, placebo-controlled, observer-blinded, dose-escalation trial. *Lancet Infect Dis* 2021;21:263–274.

36. Banks C, Lindbom BJ, Kitson G, Darsley M, Fischer PB. Preclinical development of a novel Group B Streptococcus (GBS) vaccine candidate for maternal immunization based upon the alpha-like protein family of GBS surface proteins (Alp). *Birth defects Res* 2023;115:933–944.

37. Springman AC, Lacher DW, Waymire EA, Wengert SL, Singh P, Zadoks RN, *et al.* Pilus distribution among lineages of group b streptococcus: an evolutionary and clinical perspective. *BMC Microbiol* 2014;14:.

38. Spellerberg B, Rozdzinski E, Martin S, Weber-Heynemann J, Schnitzler N, Lütticken R, *et al.* Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human laminin. *Infect Immun* 1999;67:871–878.

39. Dramsi S, Caliot E, Bonne I, Guadagnini S, Prévost MC, Kojadinovic M, *et al.* Assembly and role of pili in group B streptococci. *Mol Microbiol* 2006;60:1401–1413.

40. Cozzi R, Malito E, Lazzarin M, Nuccitelli A, Castagnetti A, Bottomley MJ, *et al.* Structure and assembly of group B streptococcus pilus 2b backbone protein. *PLoS One* 2015;10:.

41. Maisey HC, Hensler M, Nizet V, Doran KS. Group B Streptococcal Pilus Proteins Contribute to Adherence to and Invasion of Brain Microvascular Endothelial Cells. *J Bacteriol* 2007;189:1464.

42. Rubens CE, Wessels MR, Heggen LM, Kasper DL. Transposon mutagenesis of type III group B Streptococcus: correlation of capsule expression with virulence. *Proc Natl Acad Sci U S A* 1987;84:7208–7212.

43. Tazi A, Disson O, Bellais S, Bouaboud A, Dmytruk N, Dramsi S, *et al.* The surface protein HvgA mediates group B streptococcus hypervirulence and meningeal tropism in neonates. *J Exp Med* 2010;207:2313–2322.

44. Teatero S, McGeer A, Low DE, Li A, Demczuk W, Martin I, *et al.* Characterization of invasive group B streptococcus strains from the greater Toronto area, Canada. *J Clin Microbiol* 2014;52:1441–1447.