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

Carl Britto1\*, Léa Cavalli2\*, Madikay Senghore2, Alexander McAdam3, William P. Hanage2\*\*, Ying-Jie Lu4\*\*, and Richard Malley4\*\*

1Division of Critical Care, Department of Anaesthesiology, Critical Care and Pain Medicine Boston Children’s Hospital.

2Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

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

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

\*Contributed equally

\*\*Contributed equally

**Abstract**

**Background:** Group B Streptococcus (GBS) is a major cause of meningitis in infants, often prompting intensive care. The molecular factors leading to meningitis and other manifestations of severe infant disease are poorly understood. Furthermore, although Late-Onset Disease (LOD) and Very Late-Onset Disease (VLOD) are clinically equivalent, the genomic differences between the GBS strains in LOD and VLOD have yet to be extensively described.

**Methods:** We characterized invasive GBS isolates from patients at Boston Children’s Hospital over 15 years using whole-genome sequencing. We compared isolate diversity with samples representing both global and national contexts and examined vaccine coverage and antimicrobial resistance. Logistic regression and linear mixed models evaluated the relationships between clinical presentations or pathogen virulence factors and disease severity or age of onset, accounting for population structure and multiple testing.

**Results:** In the 87 patients studied, 44.3% needed ICU care, and 18.6% had meningi- tis. Among GBS, we identified 5 serotypes and 6 clonal complexes, with hypervirulent clones CC17/cpsIII and CC23/cpsIa being predominant. All isolates contained vaccine candidate targets and were susceptible to penicillin and vancomycin, though many were resistant to erythromycin and clindamycin. ICU admission correlated with specific hematological abnor- malities, but meningitis did not. No known virulence factors were associated with ICU or meningitis. LOD and VLOD did not differ bacteriologically, though the PI-2A1 pilus was linked with a higher likelihood of VLOD.

**Conclusions:** The frequent ICU admissions and cases of meningitis observed in infant GBS invasive disease underscore its significant morbidity. However, the lack of molecular risk factors associated with severe disease highlights the need for further research into the host-pathogen interactions underlying infection. While clinical similarities between LOD and VLOD are widely accepted, we find that the 3 month LOD cut-off is also arbitrary from a bacteriological perspective, with potential differences linked to varying immune status with age.

# Introduction

*Streptococcus agalactiae*, commonly known as Group B *Streptococcus* (GBS), is a pathobiont that can transition from an asymptomatic colonizer of the vaginal or gastrointestinal tract to a virulent pathogen, causing severe invasive infections. [[1](#_bookmark2)] It is a common cause of sepsis and meningitis among neonates and infants globally. [[1](#_bookmark2),[2](#_bookmark3)] In the United States, the Centers for Disease Control and Prevention (CDC) reports that GBS infections occur in approximately 0.4 out of every 1,000 infants under three months of age, with potential outcomes including mortality or long-term complications such as developmental delays, hearing loss, or cerebral palsy. [[3](#_bookmark4), [4](#_bookmark5)]

Neonatal GBS disease is classified as early-onset disease (EOD) when it is diagnosed within the first seven days of life. After this period, infant infections are categorized as late-onset disease (LOD) if diagnosed between seven days and three months, or as very late-onset disease (VLOD) , sometimes referred to as ultra late-onset disease (ULOD), when occurring between three months and one year of age. EOD results from vertical transmission during birth via a colonized vaginal canal and is effectively prevented through intrapartum antibiotic prophylaxis (IAP) for GBS-positive pregnant women. In contrast, LOD incidence has not been affected by IAP implementation in Europe and North America, allowing these later-onset infections to surpass EOD in frequency, despite historically being less common. [[7](#_bookmark8), [8](#_bookmark9)] This suggests that LOD might be acquired through different pathways besides vertical acquisition of commensal GBS with later translocation and invasion. For instance, post-partum acquisition from maternal sources (including mucosal colonization and breast milk), as well as hospital and community infections have been proposed [[9](#_bookmark10)]. The different routes of acquisitions of EOD and LOD, along with potential changes in immune status during infancy that may alter strain-specific susceptibilities, are reflected in the different strain distributions of GBS isolates from EOD and LOD, which have been extensively characterized. ~~These bacteriological differences, combined with the evolving patient characteristics as they age, lead to variations in the clinical presentation and management of EOD and LOD.~~ In contrast, it is widely accepted that the 3 month cut-off for LOD is clinically arbitrary, with LOD and VLOD cases often exhibiting similar clinical presentations and treatments. [5, 6] However, the genomic characteristics of GBS isolates from VLOD have not extensively been described, leaving uncertainty about whether this cut-off is equally arbitrary from a bacteriological perspective. A deeper understanding of the bacterial etiologies of LOD and VLOD is needed to assess the effectiveness of potential prevention and treatment strategies with specific bacteriological targets. For instance, it may inform the potential coverage of vaccine candidates under development for VLOD cases. [[11](#_bookmark12), [12](#_bookmark13)]

The substantial clinical burden of GBS meningitis, especially its long-term neurodevelopmental impacts, is well-documented. [[13](#_bookmark14)–[20](#_bookmark15)] Global trends in molecular epidemiology identify serotype III, which belongs to the hypervirulent clonal complex 17 (CC17), as the most prevalent serotype in cases of meningitis and LOD. [[7](#_bookmark8),[13](#_bookmark14),[21](#_bookmark16)] Despite this, the specific molecular mechanisms that underlie neurotropism remain unclear. Data scarcity also limits the understanding of the molecular basis of other clinical features of acute infection in infants beyond meningitis, such as ICU admission rates. Understanding these factors is essential for the development of effective immunoprevention strategies to manage invasive disease.

In this study, we characterize the clinical and molecular characteristics of the neonatal invasive GBS disease cases recorded at Boston Children’s Hospital (BCH), Massachusetts, between 2007 and 2021. Our dataset notably provides comprehensive clinical details for infant GBS invasive disease beyond EOD, encompassing both LOD and VLOD. We employ whole-genome sequencing to examine the diversity of GBS isolates, estimate the potential coverage of vaccine candidates, and assess antimicrobial resistance. We also compare the genomic characteristics of GBS isolates from LOD and VLOD, and examine a twin pair providing insights into possible VLOD acquisition pathways. Furthermore, we describe the clinical presentation of severe disease, as measured by ICU admission and meningitis presentation. Finally, we investigate the association of molecular factors with well-defined clinical outcomes among infants, such as risk of ICU admission, the occurrence of meningitis, and abnormal clinical laboratory data.

# 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. [[22](#_bookmark17)]

## Data Collection

We obtained 100 GBS isolates from patients at Boston Children’s Hospital (BCH) between 2007 and 2021. Located in Boston, Massachusetts, BCH is a tertiary referral pediatric hospital serving a catchment area encompassing the majority of New England’s 72,000 square miles. The hospital admits a diverse range of patients, including neonates with suspected sepsis. A scarcity of EOD was observed, which aligns with the absence of labor and delivery services at BCH. Consequently, our analysis primarily focuses on LOD and infections occurring later in infancy, notably VLOD. We also included older patients, including adults with congenital anomalies treated at BCH, in this analysis as a comparator population to the infant population. *[~~Carl/Rick to include statement about why we also observe some adult patients in our cohort~~]* Isolates were obtained from either the blood or cerebral spinal fluid (CSF) of patients. The presence of bacteria in the blood indicated bacteremia, while its presence in the CSF, characterized by pleocytosis, indicated meningitis. Each isolate was stored for whole-genome sequencing, and the associated demographic, antimicrobial sensitivity profile, laboratory, and clinical information were extracted from hospital records. The recorded variables included gestational age, sex, length of stay, vitals, clinical presentation, hematological laboratory data, medications, and interventions. *[Carl/Rick: length of stay, vitals, clinical presentation/symptoms, medications, and interventions are not in- cluded in the current data sheet for publication - would you like to include them, and describe them in the paper? Or exclude them ?]* The hematological data recorded included white blood cell (WBC) count, hemoglobin concentration, platelet count, and absolute neutrophil count (ANC). Hematological data were entirely missing for two patients, and ANC data were missing for an additional patient. We treated the missing data as random and excluded these patients in statistical analyses involving the missing variables. For all patients except one preterm infant (Patient#86), gestational age coincided with age in days since birth. We also documented the level of care received by each patient, distinguishing between Intensive Care Unit (ICU) and inpatient pediatric unit admissions. To exclude precautionary admissions, ICU admissions were defined as stays exceeding 48 hours.

## Whole-genome sequencing and Antibiotic susceptibility testing

*[Carl/Rick: Please add details on the experimental methods used for sequencing and AST]*

DNA was extracted using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA), according to manufacturer’s instructions. Genomic DNA was then subjected to indexed whole genome sequencing on an Illumina Hiseq 2500 platform to generate paired-end reads of 100–150 bp in length.

The Clinical and Laboratory Standards Institute (CLSI) protocol was followed for antimicrobial susceptibility testing using disk diffusion. Briefly, individual GBS strains were grown on Mueller-Hinton Agar (MHA) with 5% sheep blood (BBL; ThermoFisher) at 37°C supplemented with 5% CO2 following the placement of antimicrobial discs (BD Sensi-Disc; ThermoFisher) using a standard dispenser to ensure proper spacing between disks. Tested antibiotics included penicillin (10U), tetracycline (30μg), clindamycin (2μg), erythromycin (15μg), and levofloxacin (5μg). Zones of inhibition were measured using a digital caliper following 18 hours of growth and recorded as susceptible, intermediate, or resistant according to the CLSI breakpoints ref

## Genomic analysis

Trimmomatic (v.0.39) was used to remove sequencing adapters from raw reads, followed by FastQC (v0.12.1) and Kraken (v.1.0) on trimmed reads to verify low adapter content and identify con- tamination with the miniKraken database (minikraken\_20171019\_8GB), respectively. [[23](#_bookmark18)–[25](#_bookmark19)] We eliminated 4 isolates that had less than 50% of reads classified as ‘Streptococcus agalactiae’ (GBS). Srst2 (v.0.2.0) and the Sanger GBS-Typer (v.1.0.12) tool were used on trimmed reads to obtain the serotype and Sequence Type (ST) of each isolate, and to identify the presence of genes encoding surface proteins and antibiotic resistance. [[26](#_bookmark20), [27](#_bookmark21)] The ST of each isolates was then translated into its corresponding clonal complex (CC) based on known ST-CC matches available in the PubMLST Streptococcus agalactiae typing database. [[28](#_bookmark22)] Unicycler (v0.4.8) was used for genome assembly, followed by QUAST to check the quality of assemblies. [[29](#_bookmark23), [30](#_bookmark24)] All our isolates passed the QC criteria, since they had fewer than 250 contigs, N50 *≥* 30*,* 000, genome lengths between 1.7m and 2.4m, and a GC content between 30% and 40%. Abricate (v.1.0.1) was used on assembled genomes to detect virulence factor genes with the vfdb databases and the SIP gene with a custom database. [[31](#_bookmark25)] Snippy (v.4.6.0) was used to obtain a whole genome SNP alignment, which was input into RAxML (v8.2.12) to infer a Maximum likelihood (ML) phylogenetic tree. [[32](#_bookmark26), [33](#_bookmark27)] 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. [[34](#_bookmark28)] A Streptococcus pyogenes (Group A Streptococcus, GAS) strain (ENA accession: SRR1104967) was used as an outgroup to root the inferred phylogenetic tree of the GBS isolates. [[35](#_bookmark29)] The resulting tree was visualized and annotated using the R package ggtree (v3.8.2). [[36](#_bookmark30)] GBS was isolated from both the blood and CSF in 9 patients, yielding identical genome sequences. We therefore removed duplicate strains from the genomic analysis, leaving us with a total of 87 samples for analysis. For the pair of VLOD twins identified in our dataset, the SNP-distances between their isolates were calculated with *snp-dists*. [[37](#_bookmark31)] Finally, the BCH LOD strains (n=48) were contextualized at the national and global level using publicly available LOD genomes from the US CDC (n=267, project accession: PRJNA355303) and other global cohorts: Netherlands (n=175, project accession: PR- JEB14124), Malawi (n=47, project accession: PRJEB8986), Canada (n=34, project accession: PRJNA295774), Ireland (n=16, project accession: PRJEB26339). [[4](#_bookmark5), [21](#_bookmark16), [38](#_bookmark32), [39](#_bookmark33)] A global phylogeny was built with RAxML to compare the phylogeographic distribution of samples. [[40](#_bookmark34)] The code for bioinformatic steps is available in GitHub at: <https://github.com/Leacavalli/BCH-GBS>.

## Analysis of Clinical and Molecular Correlates of Severe Disease

We examined hematological data of infant infections across level of care (ICU vs. inpatient pediatric unit), and meningitis status. The hematological data analyzed included WBC count, hemoglobin level, ANC, and platelet counts, along with the occurrence of leukopenia, leukocytosis, and neutropenia. To classify leukopenia, leukocytosis, and neutropenia, we binarized the WBC and ANC values using infant-specific thresholds: neutropenia was defined as ANC *<* 10 *×* 103 cells/mm3, leukocytosis as WBC count *≥* 10*.*50 *×* 103 cells/mL, and leukopenia as WBC count *<* 3*.*50 *×* 103 cells/mL. Logistic regression was employed to calculate odds ratios and 95% confidence intervals for each hematological indicator. The Bonferroni correction was applied for multiple testing, setting a significance threshold of 0.007 (0.05/7) for the seven hematological indicators.

We conducted a genetic association study to identify molecular risk factors associated with clinical outcomes. The outcomes of interest included the age of disease onset and severity indicators

such as ICU admission, meningitis, and abnormal hematological parameters, including leukopenia, leukocytosis, and neutropenia. Due to limitations in statistical power arising from sample size, the study concentrated on known virulence genes instead of conducting a genome-wide SNP analysis or evaluating all genes in the pangenome. We used a linear mixed model in Pyseer to examine gene- phenotype associations incorporating patristic distances from the RAxML phylogeny to account for population structure. [[41](#_bookmark35)] A Bonferroni correction was applied to address multiple testing, establishing a significance threshold of 0.003 (0.05/17) to evaluate associations with 17 virulence factors. Data analysis and visualizations were performed using R (V.4.3.1), and the corresponding code is available on GitHub at <https://github.com/Leacavalli/BCH-GBS>. [[42](#_bookmark36)]

## Analysis of Clinical and Molecular Correlates of Age of Onset Among Infants Beyond EOD

Logistic regression with Bonferonni correction was similarly applied to compare the clinical presentation of LOD and VLOD, examining hematological data, ICU admission rates and the incidence of meningitis. We then examined potential differences in the strains of the GBS isolates present across age groups by comparing the distribution of clonal complexes and serotypes using Fisher’s test FDR-adjusted p-values. Finally, we used the linear mixed model in Pyseer to determine whether specific molecular risk factors associated with age of onset among infant beyond EOD. Age of onset was tested both as a binary outcome classified into LOD and VLOD, and as a continuous outcome, to account for the arbitrary nature of the 3 month LOD cutoff.

# Results

***3.0 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 children 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, 32 (46.3 %) 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 (29%) infants had meningitis.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1 : Clinical laboratory parameters stratified outcome and age | | | | | | | |
|  | **Hemoglobin  (mean (SD))** | **WBC (mean (SD))** | **Leukocytosis n (%)** | **Leukopenia n (%)** | **Platelet count (mean (SD))** | **ANC (mean (SD))** | **Neutropenia n (%)** |
| **< 1 year (n = 69)** | 10.86 (2.26) | 9.19 (6.41) | 27 (39.1) | 13 (18.8) | 342.20 (158.55) | 5.57 (4.87) | 13 (18.8) |
| **> 1 year (n = 17)** | 11.49 (2.50) | 10.23 (7.76) | 04 (41.2) | 03 (17.6) | 235.74 (201.76) | 7.85 (6.50) | 03 (17.6) |
| P value | 0.293 | 0.555 | 0.466 | 0.776 | **0.018** | 0.105 | 0.725 |
|  | | | | | | | |
| **No ICU  (n = 37)** | 10.90 (1.79) | 11.84 (6.03) | 21 (56.8) | 01 (0.36) | 394.05 (166.9) | 7.31 (4.59) | 03 (0.28) |
| **ICU (n=32)** | 10.80 (2.78) | 5.81 (5.26) | 06 (18.8) | 12 (37.5) | 276.03 (120.2) | 3.35 (4.33) | 10 (31) |
| P-value | 0.865 | **<0.001** | **<0.001** | **0.001** | **0.002** | **0.001** | **0.009** |
|  | | | | | | | |
| **No Meningitis (n = 43)** | 10.75 (2.17) | 10.04 (6.53) | 19 (44.2) | 07 (16.3) | 366.88 (167.88) | 6.38 (4.92) | 05 (11.6) |
| **Meningitis (n = 18)** | 11.79 (2.43) | 6.96 (5.58) | 05 (27.8) | 05 (27.8) | 298.00 (141.94) | 3.69 (3.70) | 06 (33.3) |
| P-value | 0.104 | 0.086 | 0.091 | 0.079 | 0.132 | **0.041** | **0.045** |

WBC – White Blood Cell Count. ANC – Absolute Neutrophil Count

## Molecular Surveillance of Invasive GBS

Our final dataset comprised 96 clinical isolates collected from 87 patients with invasive GBS infec- tions between 2007 and 2021. In 72 patients, GBS was isolated from the blood only; in 6 patients, it was found in the cerebrospinal fluid (CSF); and in 9 cases, it was isolated from both sources. In total, 70 infants, defined as children under one year old, were diagnosed with invasive GBS infec- tion. Two patients were diagnosed within 7 days of birth (EOD), 48 between 7 days and 3 months (LOD), and 20 from 3 to 12 months (VLOD) (Figure [1](#_bookmark0), Table S1). The remaining 17 isolates were collected from 10 older children and 7 adults. Throughout the study period, serotype III was the most prevalent in infants (n=45, 64*.*3%) followed by serotype Ia (n=15, 21*.*4%), which was also predominant among older children aged 1 to 18 years (n = 5, 50%) (Figure [1](#_bookmark0), Table S1). In this sample, no isolates with capsular serotypes VI –IX and no non-typeable isolates were identified.

We identified six distinct Clonal Complexes (CCs) among the invasive GBS isolates: CC1 (N=4, 4*.*6%), CC12 (N=8, 9*.*2%), CC17 (N=41, 47*.*1%), CC19 (N=9, 10*.*3%), CC23 (N=21,

24*.*1%) and C459 (N=4, 4*.*6%) (Figures [1](#_bookmark0), Table S2). These were further divided into 18 known Sequence Types (STs), with ST-17 being the most frequent (n=33, 37*.*9%), followed by ST-23 (n=17, 19*.*5%) and ST-19 (n=5, 5*.*7%) (Table S2). Additionally, a new MLST profile that had not been previously reported in the pubMLST database was discovered: adhP(4) pheS(1) atr(4) glnA(4) sdhA(3) glcK(3) tkt(2). This new ST differed from ST-8 by one housekeeping gene (glnA), and from ST-10 by two housekeeping genes (adhP and glnA), classifying it within CC12 (Table S3). With one exception, each ST was associated with a single serotype, and each CC represented by a dominant serotype (Figure [2](#_bookmark1)). Specifically, CC17 was exclusively serotype III and CC23 was exclusively serotype Ia. Finally, we found evidence of capsule switching in CC19 and CC12. In CC19, ST-19 and ST-335 were serotype III while ST-28 was serotype II. CC1/ST-1 was serotype V, and ST-459 was serotype IV. All CC12 isolates were serotype Ib, including the newly identified sequence type, with the exception of one ST-10/cps II isolate.

The distribution of clonal complexes (CCs) and serotypes varied by age at infection, with in- fants differing significantly from older patients. Specifically, CCs differed between LOD and older children, LOD and adults, and VLOD and adults, with Fisher’s test FDR-adjusted p-values of 0.00123 for each comparison (refer to Figure [1](#_bookmark0) and Table S1). There were also serotype differences between LOD and older children (p=0.002) and between LOD and adults (p=0.036. Although CC17/cps-III was more prevalent in LOD isolates (65%) compared to VLOD isolates (45%), and CC23/cps-Ia showed the opposite trend (LOD: 15%, VLOD: 30%), no significant difference was observed in the CC distribution between LOD and VLOD (p=0.095). Additionally, no clonal expansion over time was detected (X-squared = 73.483, p-value = 0.365, and the serotype distri- bution remained stable as well (X-squared = 75.915, p-value = 0.294) (Figure [1](#_bookmark0)). These results indicate a stable diversity within the invasive GBS population over the study period, potentially reflecting the absence of selective pressures from clinical interventions targeting specific clones or capsular types.

Finally, the genetic diversity of LOD isolates in our study was consistent with that observed in globally and nationally representative GBS populations (Figure **??**). The isolates exhibited polyphyletic structures similar to those seen across broader geographical scales, encompassing various clonal complexes representative of distinct genotypes. Furthermore, there was no evidence of genetic isolation by geographic location within each clonal complex, suggesting a history of recombination and frequent international migration rather than localized transmission clusters.

1.00

0.15

0.75

Proportion

0.50

0.25

0.10

0.05

Proportion

CC

1

459

 19

 12

 23

 17

0.00 0.00

EOD

LOD

VLOD

Older Children

Adults

2007

2008

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

2019

2020

2021

1.00

Age group

0.15

n=2 n=48 n=20 n=10 n=7

Year of Isolation

0.75

Proportion

0.50

0.25

0.10

0.05

Proportion

Serotype Ia

Ib  II  III  IV  V

n=2 n=48 n=20 n=10 n=7

0.00 0.00

EOD

LOD

VLOD

Adults

2007

2008

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

2019

2020

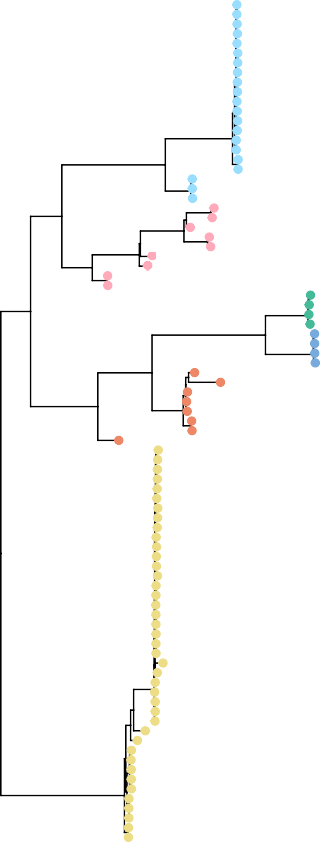
2021

Age group Year of Isolation

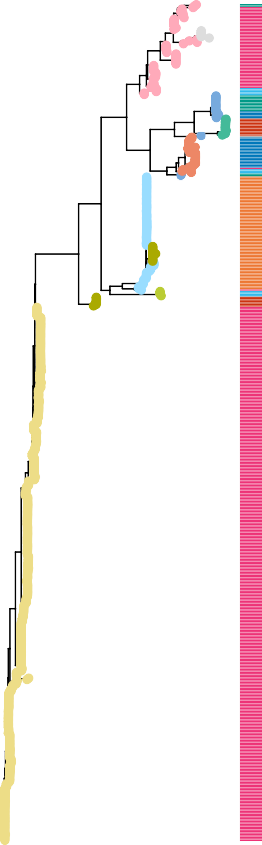
Older Children

**Figure 1. Distribution of group B Streptococcus (GBS) clonal complexes (CCs) and serotypes across age groups and over time.** The top panel displays the proportions of CCs across different age groups (left) and annually from 2007 to 2021 (right), while the bottom panel presents the corresponding distributions for serotypes. The age groups are defined as follows: early-onset disease (EOD) is diagnosed within 7 days of birth, late-onset disease (LOD) between 7 days and 3 months, very late-onset disease (VLOD) between 3 months and 1 year, older chil- dren range from 1 to 18 years, and adults are 18 years and older.



A



B

1 2 3 4 5 A B C D E F GH I J K L M N O P Q 1 6

Clonal Complex

1: Serotype 2: Collection date

3: Age group

4: ICU

6: Location of isolation

 1

 12

 17

 19

 22

 23

 327

 452

 459

Ia Ib II III IV V VI

2020

2016

2012

2008

EOD LOD

VLOD

Older Children Adults

No Yes

5: Meningitis

No Yes

BCH USA

other

**Figure 2. Phylogenies of Boston Children’s Hospital (BCH) and Global Group B Streptococcus (GBS) isolates.** Panel A presents the phylogenetic tree of Group B Strep- tococcus (GBS) isolates from Boston Children’s Hospital (BCH), categorized by age groups: early-onset disease (EOD, diagnosed within 7 days of birth), late-onset disease (LOD, 7 days to 3 months), very late-onset disease (VLOD, 3 months to 1 year), older children (1 to 18 years), and adults (18 years and older). Various genes for virulence factors and surface proteins are detailed, in- cluding Alpha-like proteins (ALP1, ALP23, Alpha, RIB), pilus islands (PI-1, PI-2a1, PI-2a2, PI-2b), and other factors such as the hypervirulence gene cluster A (HVGA), serine-rich repeat proteins (SRR1, SRR2), Sip, laminin-binding protein (lmb), C5a peptidase (scpB), hyaluronidase (hylB), and fibrinogen-binding protein (fbsB). Panel B extends the context by situating BCH LOD isolates (in blue on heatmap 6) within a broader phylogenetic framework, including na- tional and global LOD isolates. It features isolates from the USA (yellow on heatmap 6), gath- ered through the CDC’s ABCs program, and from other international sources such as Ireland, Malawi, Canada, and The Netherlands (orange on heatmap 6).

## Coverage of genes encoding vaccine targets

In assessing the potential coverage of the two main candidate GBS vaccines, the hexavalent vaccine (GBS6 by Pfizer) targeting capsular polysaccharides (cps) of serotypes Ia, Ib, II, III, IV and V would theoretically cover 100% of infant cases in our study (Figure [1](#_bookmark0), Table S4) [[11](#_bookmark12)]. The other leading vaccine candidate (GBS-NN2 by Minervax) contains fusion proteins from the alpha-like protein (Alp) family of GBS. [[12](#_bookmark13)] All infant isolates in our study contained genes encoding at least one Alp protein: *rib* (n=46, 66%), *alphaC* (n=7, 10%), *alp1* (n=15, 21%), and *alp23* (n=2, 3%) (Figure 2, Table S4). The *rib* gene was exclusively found in CC17 (n=40) and CC19 (n=6) isolates, alphaC in CC12 (n=7), *alp1* in CC23 (n=14) and CC459 (n=1), and *alp23* in CC23 (N=1) and CC1 (n=1) (Figure [2](#_bookmark1)). The capsular types and proteins targeted by these vaccines are present in all isolates from older children and adults in our dataset, indicating their potential as suitable candidates for adult vaccination in outbreaks or other high-risk scenarios. [[43](#_bookmark37)] For example, this includes the 2015 outbreak in Singapore caused by ST283/cpsIII in non-pregnant adults linked to the consumption of raw fish. [[44](#_bookmark38)]

We also examined the distribution of other immunogenic proteins encoded by isolates in our study cohort. All infant isolates encoded at least one of the three pilus islands: *PI-1* (n=49, 70%), PI-2a (n=27, 39%) and *PI-2b* (n=38, 54%), as well as the *sip* gene (Table S4). These immunogenic proteins have the potential to serve as universal vaccine targets. In contrast, not all infant isolates carried genes encoding *C5a peptidase* (n=68, 97%), *Lmb* (n=69, 99%), and *FbsB* (n=55, 79%) (Table S4). The gene encoding FbsB protein was exclusively present in hypervirulent clones CC17 (n=40, 100%) and CC23 (n=10, 100%) isolates, while absent from all other clonal complexes (Figure [2](#_bookmark1)).

## Linking Phenotypic Antibiotic Resistance to Genotypic Markers

Phenotypic antimicrobial resistance (AMR) testing showed all isolates were susceptible to peni- cillin, the preferred treatment for GBS infections, and to vancomycin, which is recommended for those with penicillin allergies and clindamycin resistance. Thirty-three (38%) isolates exhibited resistance to erythromycin, twenty-three (26%) were resistant to clindamycin, and three (3%) had intermediate resistance to clindamycin. Genes encoding macrolide resistance (ermA, ermB, ermT, and mefA-msrD) were identified in all erythromycin-resistant isolates and in 4 (7%) erythromycin- susceptible isolates. Genes linked to lincosamide resistance (ermA, ermB, and ermT) were found in all clindamycin-resistant isolates, one of the three clindamycin-intermediate isolates, and seven of the 61 that were clindamycin-susceptible.

We also assessed the presence of genetic resistance markers associated with resistance to other antibiotics (Table S5, Fig S1). Tetracycline resistance predicted in 78 isolates (90%) in our study from the presence of tet-M (*n* = 76, 87%), tet-O (*n* = 2, 2%), or both (*n* = 6, 7%). The highest rates of tetracycline resistance was observed in CC1 (*n* = 4*/*4, 100%) and CC17 (*n* = 39*/*40, 98%) compared to other CCs (CC12: *n* = 6*/*8, 75%; CC459: *n* = 2*/*4, 50%; CC19: *n* = 8*/*9, 89%; CC23: *n* = 18*/*21, 86%). Aminoglycoside resistance was predicted in 4 isolates (5%) due to the presence of the aph(3’)-III and ant(6) resistance genes, all belonging to CC17/ST-17. No SNP variants conferring reduced penicillin susceptibility were found in the rRNA 23-S1 and 23S-3 genes. However, gyrA or parC SNP variants, typically associated with fluoroquinolone resistance, were detected in three isolates: two CC12 isolates carried the S81L mutation in GyrA and S79F in ParC, and one CC17 isolate carried the S79F mutation in ParC alone.

In 2010, the GBS AIP guidelines for pregnant individuals were revised, notably removing erythromycin due to concerns about resistance. In light of this change during the study period, we evaluated the temporal changes in genetic elements that confer resistance (see Figure S1). We found that isolates carrying the aph(3’)III/ant(6) genes only appeared from 2014 onwards. In contrast, the mef(A)-msr(D) genes, which confer resistance to erythromycin, were absent after 2013. Meanwhile, the frequency of erm genes, associated with resistance to both clindamycin and erythromycin, increased from 29% to 41%.

## Clinical and Molecular Correlates of Severe Infant Disease

Among the infant cases, 31 (44*.*3%) required ICU admission due to risks of clinical decompensation, mental status changes or escalating respiratory support. Analysis indicated that leucocytosis, thrombocytosis , and high absolute neutrophil counts (ANC) were significantly correlated with reduced odds of ICU admission (WBC: OR=0.84, 95% CI=0.75-0.93, p-value=0.001, platelet: OR= 0.99, 95% CI=0.99-1, p-value=0.003; ANC: OR= 0.83, 95% CI= 0.73-0.94, p-value= 0.004; Table S6). In line with this, leukopenia was associated with significantly higher odds of ICU admission (OR= 6.5, 95% CI= 1.84-22.98, p-value= 0.004), while leukocytosis showed a protective effect (OR= 0.1, 95% CI= 0-0.5, p-value= 0.002). [*Carl/Rick: are these hemotological indices expected to be significantly associated with the odds of ICU admission ? Does it reflect any spe- cific guidelines ?* ] Meningitis was documented in thirteen (18*.*6%) infants, but no hematological indices were associated with meningitis after multiple testing corrections (Table S6).

Among infants, none of the virulence factors we evaluated were associated with ICU admission, meningitis, or abnormal hematological indices (leukopenia, leukocytosis, neutropenia) after accounting for population structure and correcting for multiple tests (Tables S8–S11). Furthermore, these clinical characteristics were not significantly correlated with any particular clonal complex: ICU admission (*χ*2 = 6*.*58*, p* = 0*.*25), meningitis (*χ*2 = 2*.*39*, p* = 0*.*79), neutropenia (*χ*2 = 6*.*09*, p* = 0*.*30), leukopenia (*χ*2 = 8*.*72*, p* = 0*.*12), and leukocytosis (*χ*2 = 3*.*75*, p* = 0*.*59).

**3.5. Bacteriological Similarities Between VLOD and LOD Underscore the Arbitrary 3-Month Cutoff**

Among the virulence factors we investigated, we found that the presence of pilus island PI-2A1 was associated with reducing the odds of LOD by 0.60 times compared to VLOD (95% CI: 0.42 - 0.85, *p* = 0*.*002). Although this effect remained important even after controlling for population structure (p=0.0058), it did not quite meet the stringent significance threshold of 0.003 required by multiple testing correction, despite being close (Table S13). We also evaluated the association between PI-2A1 and continuous age among infants to increase statistical power and account for the clinically arbitrary 3-month cut-off for LOD, but found the effect remained similar (Table S14). Nevertheless, considering the conservative nature of the Bonferroni correction and our small size, this finding may still indicate a genuine biological effect. None of the other virulence factors we investigated showed significant difference between LOD and VLOD isolates, possibly reflecting the similarities in their strains distributions (Table S).

The genomic similarity of GBS strains in LOD and VLOD suggests similar acquisition through similar routes from similar niches. In our data, a pair of VLOD twins, similar to previously reported LOD twins, provides insights into potential VLOD routes of acquisition. In 2020, this pair of twin infants presented with fever and seizures 107 days after birth at BCH and were diagnosed with meningitis following the isolation of GBS from their CSF. One twin was admitted to the ICU for treatment with ceftriaxone, vancomycin, and penicillin, while the other, treated with ceftriaxone and penicillin G, remained in the inpatient unit. Isolates from both twins were identified as the CC23/cpsIa strain. Phylogenetic analysis showed the isolates clustered together, differing by only 5 single nucleotide polymorphisms. Consistent with this, they encoded identical virulence factors (ALP1, PI2A1, SRR1) and antibiotic resistance markers (TET-M, TET-SM). Although both twins had negative neonatal GBS screening results, the genetic similarity of the isolates suggests a shared source of infection. While early intestinal colonization could explain the infections, simultaneous acquisition from an external source, such as community or enteral transmission, seems more plausible, similar to other cases reported in LOD twins. [[45](#_bookmark39)]In line with the known clinical similarities between LOD and VLOD, our study found no significant differences in the odds of ICU admission and meningitis odds between LOD and VLOD patients (ICU admission: OR= 3, 95% CI=0.9-9.6, p-value=0.063; meningitis: OR= 0.8 , 95% CI=0.2-3, p-value=0.743). Similarly, no significant differences were observed in any hematological indices (Hb, WBC, platelet, ANC counts, or leukocytosis, leukopenia, neutropenia) between LOD and VLOD patients (Table S6).

# Discussion

Our study characterizes cases of post-EOD GBS infant invasive disease over a 15-year period at Boston Children’s Hospital. We describe clinical characteristics of patients and the genomic determinants of bacterial isolates and compare them to broader contexts, alongside detailing their clinical presentations, to identify molecular risk factors associated with severe disease.

among patients

The diversity observed in clonal complexes and serotypes within our cohort reflects global trends, with hypervirulent clones CC17/cpsIII and CC23/cpsIa being predominant in cases of post-EOD infant invasive disease. All invasive disease isolates encoded targets of both maternal vaccine candidates currently in development: GBS6 and GBS-NN2. [[11](#_bookmark12), [50](#_bookmark44)] While these vaccines aim to prevent EOD and LOD, the observed potential coverage and the maternal antibodies they induce —reported to persist in infants for over three months — also suggest they could help prevent VLOD. [[50](#_bookmark44), [51](#_bookmark45)] Moreover, the broad coverage of these vaccines in older pediatric and adult patients highlights their value for adult vaccination during outbreaks or in high-risk settings, which is significant given the rise in invasive disease among non-pregnant adults since 1990. [[43](#_bookmark37), [52](#_bookmark46)] Despite this promise, the global diversity of CPS types and Alp sequences raises concerns about universal vaccine coverage. Gaps in coverage due to non-vaccine types and non-typeable isolates underscore the need for post-vaccine GBS population monitoring to guide and inform future clinical interventions. [[53](#_bookmark47)–[55](#_bookmark48)] Several other surface proteins have been identified as potential vaccine candidates due to their immunogenic nature observed in preclinical research. [[56](#_bookmark49)] The presence of Sip and at least one pilus island in all our isolates supports their potential as targets in a universal GBS vaccine. [14] In contrast, C5a, Lmb, FbsB and Srr1/2 proteins were absent from several of our isolates, which may limit their use as vaccine targets.

With regards to antimicrobial resistance, our findings also align with broader trends. AST revealed no phenotypic non-susceptibility to penicillin and vancomycin, which are the recommended first and last line antibiotics in AIP, respectively. [[57](#_bookmark50)] Additionally, we did not identify any genetic resistance markers for these antibiotics.Similarly, no genetic markers for resistance were detected for other beta- lactams such as ampicillin or cefazolin, also recommended for AIP in certain cases. [[57](#_bookmark50)] Conversely, approximately 29% and 40% of isolates were non-susceptible to clindamycin and erythromycin, respectively. The high rates of erythromycin resistance among invasive GBS isolates are well-documented and prompted its removal from U.S. AIP guidelines in 2010.[[57](#_bookmark50)] Although not statistically significant, the absence of mef(A)/msr(D) genetic markers and a rise in the prevalence of erm genes after 2013 might have biological and clinical significance. The absence of mef(A)/msr(D), which confers resistance to erythromycin but not clindamycin, may reflect reduced selective pressure for erythromycin resistance following the 2010 revision of AIP guidelines. [[57](#_bookmark50)] The ongoing presence of ermA, ermB, and ermT genes, conferring resistance to both erythromycin and clindamycin, aligns with clindamycin’s continued use in treating GBS infections.It has been suggested that the rise clindamycin resistance in GBS may result from its overuse in penicillin-allergic patients, highlighting the importance of conducting clindamycin susceptibility testing in maternal carriage isolates and considering vancomycin for treating resistant strains. [57, 58] However, a similar resistance trend is observed in other Gram-positive organisms, such as Staphylococcus aureus and Group A Streptococcus, such that this pattern might also be attributed to bystander selection from clindamycin use in various contexts. Finally, tetracycline exhibits the highest resistance rate in GBS (90% in our study) despite having never been used to treat this infection and rarely been prescribed in the last 20 years. This high resistance rate is an inherent characteristic of invasive GBS, likely due to the widespread use of tetracycline in the 1950s to prevent and treat various infections, which resulted in lasting changes in the bacterial population and reportedly contributed to the emergence of neonatal GBS in the 1960s.

Finally, with our cohort primarily consisting of post-EOD cases, including 20 VLOD cases, we took the opportunity to compare them to LOD cases. VLOD cases, making up one-third of our cohort, showed no significant clinical differences from LOD. This aligns with previous reports and underscores the widely acknowledged arbitrary nature of the 3-month cut-off for LOD. [5, 6] The strain diversity was also similar between LOD and VLOD cases, with CC17/cpsIII prevailing in LOD and CC23/cpsIa in VLOD. Interestingly, the presence of PI-2A variant 1 was associated with higher odds of VLOD, which persisted when population structure was taken into account. PI-2A stands out as particularly variable among the three GBS pilus islands. [[47](#_bookmark41)] The surface protein typing tool included PI-2A1 and PI-2A2 to maximize PI-2A coverage, rather than to signify distinct biological phenotypes. [[48](#_bookmark42)] Nonetheless, these subvariants are genomically distinct and may have significant biological differences. The conservation of PI-2A alleles indicates that genomic variations among subvariants likely do not affect functionality, suggesting PI-2A1 may not have unique virulence traits. Instead, the observed trends might result from antigenic variation due to immune selection pressures, such as prior exposure or waning maternal antibodies, which can increase susceptibility to different antigenic alleles as infants age—a pattern also seen in a related pilus locus in pneumococci. [[49](#_bookmark43)] Aside from these possible age-related changes in exposure and immune landscape, the similarities between LOD and VLOD cases suggest they are likely bacteriologically equivalent. This implies acquisition from the same niches, as supported by a case of VLOD twins pointing to simultaneous external acquisition—consistent with proposed LOD acquisition routes notably exemplified by LOD twins. [[9](#_bookmark10), [45](#_bookmark39), [46](#_bookmark40)] emphasizing the arbitrary nature of the 3month cut off and suggesting the VLOD cases might be included alongside LOD in studies of infant invasive disease.

Overall, our study highlights the complex host-bacterial interactions underlying invasive infant GBS disease, affecting outcomes likes disease severity. It calls for broader investigations of the molecular risk factors linked to these outcomes, such as genome-wide association studies that require larger datasets, which are now becoming increasingly available.

# References

1. S. Shabayek and B. Spellerberg, “Group b streptococcal colonization, molecular characteristics, and epidemiology,” *Frontiers in microbiology*, vol. 9, p. 437, 2018.
2. World Health Organization (WHO), “Group b streptococcus infection causes an estimated 150,000 preventable stillbirths and infant deaths every year,” 2017. Accessed: 04 December 2024.
3. Center for Disease Control and Prevention (CDC), “ABCs bact facts interactive data dash- board,” 2024. Accessed: 04 December 2024.
4. L. McGee, S. Chochua, Z. Li, S. Mathis, J. Rivers, B. Metcalf, A. Ryan, N. Alden, M. M. Farley, L. H. Harrison, *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,” *Clinical Infectious Diseases*, vol. 72, no. 6, pp. 1004–1013, 2021.
5. J. B. Cantey, C. Baldridge, R. Jamison, and L. A. Shanley, “Late and very late onset group b streptococcus sepsis: one and the same?,” *World Journal of Pediatrics*, vol. 10, pp. 24–28, 2014.
6. A. W. Bartlett, B. Smith, C. R. George, B. McMullan, A. Kesson, M. M. Lahra, and P. Palas- anthiran, “Epidemiology of late and very late onset group b streptococcal disease: fifteen-year experience from two australian tertiary pediatric facilities,” *The Pediatric infectious disease journal*, vol. 36, no. 1, pp. 20–24, 2017.
7. C. R. Phares, R. Lynfield, M. M. Farley, J. Mohle-Boetani, L. H. Harrison, S. Petit, A. S. Craig, W. Schaffner, S. M. Zansky, K. Gershman, *et al.*, “Epidemiology of invasive group b streptococcal disease in the united states, 1999-2005,” *Jama*, vol. 299, no. 17, pp. 2056–2065, 2008.
8. S. A. Nanduri, S. Petit, C. Smelser, M. Apostol, N. B. Alden, L. H. Harrison, R. Lynfield, P. S. Vagnone, K. Burzlaff, N. L. Spina, *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 pediatrics*, vol. 173, no. 3, pp. 224–233, 2019.
9. F. Miselli, I. Frabboni, M. Di Martino, I. Zinani, M. Buttera, A. Insalaco, F. Stefanelli,

L. Lugli, and A. Berardi, “Transmission of group b streptococcus in late-onset neonatal disease: A narrative review of current evidence,” *Therapeutic Advances in Infectious Disease*, vol. 9,

p. 20499361221142732, 2022.

1. K. M. Puopolo, R. Lynfield, J. J. Cummings, I. Hand, I. Adams-Chapman, B. Poindexter,

D. L. Stewart, S. W. Aucott, J. P. Goldsmith, M. Mowitz, *et al.*, “Management of infants at risk for group b streptococcal disease,” *Pediatrics*, vol. 144, no. 2, 2019.

1. J. Absalon, N. Segall, S. L. Block, K. J. Center, I. L. Scully, P. C. Giardina, J. Peterson, W. J. Watson, W. C. Gruber, K. U. Jansen, *et al.*, “Safety and immunogenicity of a novel hexavalent group b streptococcus conjugate vaccine in healthy, non-pregnant adults: a phase 1/2, ran- domised, placebo-controlled, observer-blinded, dose-escalation trial,” *The Lancet Infectious Diseases*, vol. 21, no. 2, pp. 263–274, 2021.
2. C. Banks, B. J. Lindbom, G. Kitson, M. Darsley, and P. B. Fischer, “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 Research*, vol. 115, no. 9, pp. 933–944, 2023.
3. M. N. van Kassel, G. de Boer, S. A. Teeri, D. Jamrozy, S. D. Bentley, M. C. Brouwer,

A. van der Ende, D. van de Beek, and M. W. Bijlsma, “Molecular epidemiology and mortality of group b streptococcal meningitis and infant sepsis in the netherlands: a 30-year nationwide surveillance study,” *The lancet microbe*, vol. 2, no. 1, pp. e32–e40, 2021.

1. E. Horváth-Puhó, M. N. van Kassel, B. P. Gonçalves, B. de Gier, S. R. Procter, P. Paul,

A. van der Ende, K. K. Søgaard, S. J. Hahné, J. Chandna, *et al.*, “Mortality, neurodevelop- mental impairments, and economic outcomes after invasive group b streptococcal disease in early infancy in denmark and the netherlands: a national matched cohort study,” *The Lancet Child & Adolescent Health*, vol. 5, no. 6, pp. 398–407, 2021.

1. J. Bramugy, H. Mucasse, S. Massora, P. Vitorino, C. Aerts, I. Mandomando, P. Paul,

J. Chandna, F. Seedat, J. E. Lawn, *et al.*, “Short-and long-term outcomes of group b strepto- coccus invasive disease in mozambican children: results of a matched cohort and retrospective observational study and implications for future vaccine introduction,” *Clinical Infectious Dis- eases*, vol. 74, no. Supplement\_1, pp. S14–S23, 2022.

1. J. Chandna, W.-H. Liu, Z. Dangor, S. Leahy, S. Sridhar, H. B. John, H. Mucasse, Q. Bassat,

A. Bardaji, A. Abubakar, *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,” *Clinical Infectious Diseases*, vol. 74, no. Supplement\_1, pp. S35–S43, 2022.

1. P. Paul, J. Chandna, S. R. Procter, Z. Dangor, S. Leahy, S. Santhanam, H. B. John, Q. Bassat,

J. Bramugy, A. Bardají, *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*, vol. 47, 2022.

1. L. M. Harden, S. Leahy, S. G. Lala, P. Paul, J. Chandna, S. Lowick, S. Mbatha, T. Jaye,

B. Laughton, A. Ghoor, *et al.*, “South african children: a matched cohort study of neurode- velopmental impairment in survivors of invasive group b streptococcus disease aged 5 to 8 years,” *Clinical Infectious Diseases*, vol. 74, no. Supplement\_1, pp. S5–S13, 2022.

1. H. B. John, A. Arumugam, M. Priya, N. Murugesan, N. Rajendraprasad, G. Rebekah, P. Paul,

J. Chandna, J. E. Lawn, and S. Santhanam, “South indian children’s neurodevelopmental outcomes after group b streptococcus invasive disease: a matched-cohort study,” *Clinical Infectious Diseases*, vol. 74, no. Supplement\_1, pp. S24–S34, 2022.

1. B. P. Gonçalves, S. R. Procter, P. Paul, J. Chandna, A. Lewin, F. Seedat, A. Koukounari,

Z. Dangor, S. Leahy, S. Santhanam, *et al.*, “Group b streptococcus infection during pregnancy and infancy: estimates of regional and global burden,” *The Lancet Global Health*, vol. 10, no. 6, pp. e807–e819, 2022.

1. C. Chaguza, D. Jamrozy, M. W. Bijlsma, T. W. Kuijpers, D. van de Beek, A. van der Ende, and

S. D. Bentley, “Population genomics of group b streptococcus reveals the genetics of neonatal disease onset and meningeal invasion,” *Nature communications*, vol. 13, no. 1, p. 4215, 2022.

1. W. M. Association *et al.*, “World medical association declaration of helsinki: ethical principles for medical research involving human subjects,” *Jama*, vol. 310, no. 20, pp. 2191–2194, 2013.
2. A. M. Bolger, M. Lohse, and B. Usadel, “Trimmomatic: a flexible trimmer for illumina se- quence data,” *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, 2014.
3. D. E. Wood and S. L. Salzberg, “Kraken: ultrafast metagenomic sequence classification using exact alignments,” *Genome biology*, vol. 15, pp. 1–12, 2014.
4. S. Andrews *et al.*, “Fastqc: a quality control tool for high throughput sequence data,” 2010.
5. M. Inouye, H. Dashnow, L.-A. Raven, M. B. Schultz, B. J. Pope, T. Tomita, J. Zobel, and

K. E. Holt, “Srst2: Rapid genomic surveillance for public health and hospital microbiology labs,” *Genome medicine*, vol. 6, pp. 1–16, 2014.

1. V. Dyster, H. Hung, and K. Pepper, “Gbs-typer-sanger-nf.” [https://github.com/](https://github.com/sanger-bentley-group/GBS-Typer-sanger-nf) [sanger-bentley-group/GBS-Typer-sanger-nf](https://github.com/sanger-bentley-group/GBS-Typer-sanger-nf), 2022.
2. K. A. Jolley, J. E. Bray, and M. C. Maiden, “Open-access bacterial population genomics: Bigsdb software, the pubmlst. org website and their applications,” *Wellcome open research*, vol. 3, 2018.
3. R. R. Wick, L. M. Judd, C. L. Gorrie, and K. E. Holt, “Unicycler: resolving bacterial genome assemblies from short and long sequencing reads,” *PLoS computational biology*, vol. 13, no. 6,

p. e1005595, 2017.

1. A. Gurevich, V. Saveliev, N. Vyahhi, and G. Tesler, “Quast: quality assessment tool for genome assemblies,” *Bioinformatics*, vol. 29, no. 8, pp. 1072–1075, 2013.
2. T. Seemann, “Abricate: Mass screening of contigs for antimicrobial and virulence genes..”

<https://github.com/tseemann/abricate>, 2016.

1. T. Seemann, “Snippy: Rapid haploid variant calling and core genome alignment..” [https:](https://github.com/tseemann/snippy)

[//github.com/tseemann/snippy](https://github.com/tseemann/snippy), 2014.

1. A. Stamatakis, “Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, 2014.
2. Z. Yang, “Maximum likelihood phylogenetic estimation from dna sequences with variable rates over sites: approximate methods,” *Journal of Molecular evolution*, vol. 39, pp. 306–314, 1994.
3. T. B. Athey, S. Teatero, A. Li, A. Marchand-Austin, B. W. Beall, and N. Fittipaldi, “Deriving group a streptococcus typing information from short-read whole-genome sequencing data,” *Journal of clinical microbiology*, vol. 52, no. 6, pp. 1871–1876, 2014.
4. G. Yu, D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam, “ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data,” *Methods in Ecology and Evolution*, vol. 8, no. 1, pp. 28–36, 2017.
5. T. Seemann, F. Klötzl, and J. Jonathan Terhorst, “snp-dists.” [https://github.com/](https://github.com/tseemann/snp-dists) [tseemann/snp-dists](https://github.com/tseemann/snp-dists).
6. S. Teatero, E. Ramoutar, A. McGeer, A. Li, R. G. Melano, J. Wasserscheid, K. Dewar, and

N. Fittipaldi, “Clonal complex 17 group b streptococcus strains causing invasive disease in neonates and adults originate from the same genetic pool,” *Scientific reports*, vol. 6, no. 1,

p. 20047, 2016.

1. M. Meehan, M. Eogan, N. McCallion, R. Cunney, J. E. Bray, K. A. Jolley, A. Unitt, M. C. Maiden, O. B. Harrison, and R. J. Drew, “Genomic epidemiology of group b streptococci spanning 10 years in an irish maternity hospital, 2008–2017,” *Journal of Infection*, vol. 83, no. 1, pp. 37–45, 2021.
2. L. S. Katz, T. Griswold, S. S. Morrison, J. A. Caravas, S. Zhang, H. C. den Bakker, X. Deng, and H. A. Carleton, “Mashtree: a rapid comparison of whole genome sequence files,” *Journal of Open Source Software*, vol. 4, no. 44, 2019.
3. J. A. Lees, M. Galardini, S. D. Bentley, J. N. Weiser, and J. Corander, “Pyseer: a comprehen- sive tool for microbial pangenome-wide association studies,” *Bioinformatics*, vol. 34, no. 24,

pp. 4310–4312, 2018.

1. R. C. Team *et al.*, “R: A language and environment for statistical computing,” *Foundation for Statistical Computing, Vienna, Austria*, 2013.
2. C. L. Trotter, M. Alderson, Z. Dangor, M. Ip, K. Le Doare, E. Nakabembe, S. R. Procter,

M. Sekikubo, and P. Lambach, “Vaccine value profile for group b streptococcus,” *Vaccine*, vol. 41, pp. S41–S52, 2023.

1. S. Tan, Y. Lin, K. Foo, H. F. Koh, C. Tow, Y. Zhang, L. W. Ang, L. Cui, H. Badaruddin, P. L. Ooi, *et al.*, “Group b streptococcus serotype iii sequence type 283 bacteremia associated with consumption of raw fish, singapore,” *Emerging infectious diseases*, vol. 22, no. 11, p. 1970, 2016.
2. R. Elling, M. Hufnagel, A. De Zoysa, F. Lander, K. Zumstein, M. Krueger, and P. Hen- neke, “Synchronous recurrence of group b streptococcal late-onset sepsis in twins,” *Pediatrics*, vol. 133, no. 5, pp. e1388–e1391, 2014.
3. A. Furuta, A. Brokaw, G. Manuel, M. Dacanay, L. Marcell, R. Seepersaud, L. Rajagopal, and

K. Adams Waldorf, “Bacterial and host determinants of group b streptococcal infection of the neonate and infant,” *Frontiers in Microbiology*, vol. 13, p. 820365, 2022.

1. I. Margarit, C. D. Rinaudo, C. L. Galeotti, D. Maione, C. Ghezzo, E. Buttazzoni, R. Rosini,

Y. Runci, M. Mora, S. Buccato, *et al.*, “Preventing bacterial infections with pilus-based vac- cines: the group b streptococcus paradigm,” *The Journal of infectious diseases*, vol. 199, no. 1,

pp. 108–115, 2009.

1. B. Metcalf, “GBS\_Scripts\_Reference.” [https://github.com/BenJamesMetcalf/GBS\_](https://github.com/BenJamesMetcalf/GBS_Scripts_Reference) [Scripts\_Reference](https://github.com/BenJamesMetcalf/GBS_Scripts_Reference).
2. G. Regev-Yochay, W. P. Hanage, K. Trzcinski, S. L. Rifas-Shiman, G. Lee, A. Bessolo, S. S. Huang, S. I. Pelton, A. J. McAdam, J. A. Finkelstein, *et al.*, “Re-emergence of the type 1 pilus among streptococcus pneumoniae isolates in massachusetts, usa,” *Vaccine*, vol. 28, no. 30,

pp. 4842–4846, 2010.

1. M. Gonzalez-Miro, A. Pawlowski, J. Lehtonen, D. Cao, S. Larsson, M. Darsley, G. Kitson, P. B. Fischer, and B. Johansson-Lindbom, “Safety and immunogenicity of the group b streptococcus vaccine alpn in a placebo-controlled double-blind phase 1 trial,” *Iscience*, vol. 26, no. 3, 2023.
2. S. A. Madhi, A. S. Anderson, J. Absalon, D. Radley, R. Simon, B. Jongihlati, R. Strehlau,

A. M. Van Niekerk, A. Izu, N. Naidoo, *et al.*, “Potential for maternally administered vaccine for infant group b streptococcus,” *New England Journal of Medicine*, vol. 389, no. 3, pp. 215–227, 2023.

1. L. K. F. Watkins, L. McGee, S. J. Schrag, B. Beall, J. H. Jain, T. Pondo, M. M. Farley,

L. H. Harrison, S. M. Zansky, J. Baumbach, *et al.*, “Epidemiology of invasive group b strepto- coccal infections among nonpregnant adults in the united states, 2008-2016,” *JAMA internal medicine*, vol. 179, no. 4, pp. 479–488, 2019.

1. F. Bianchi-Jassir, P. Paul, K.-N. To, C. Carreras-Abad, A. C. Seale, E. Jauneikaite, S. A. Madhi, N. J. Russell, J. Hall, L. Madrid, *et al.*, “Systematic review of group b streptococcal capsular types, sequence types and surface proteins as potential vaccine candidates,” *Vaccine*, vol. 38, no. 43, pp. 6682–6694, 2020.
2. S. Bellais, A. Six, A. Fouet, M. Longo, N. Dmytruk, P. Glaser, P. Trieu-Cuot, and C. Poyart, “Capsular switching in group b streptococcus cc17 hypervirulent clone: a future challenge for polysaccharide vaccine development,” *The Journal of infectious diseases*, vol. 206, no. 11,

pp. 1745–1752, 2012.

1. N. J. Croucher, J. A. Finkelstein, S. I. Pelton, P. K. Mitchell, G. M. Lee, J. Parkhill, S. D. Bentley, W. P. Hanage, and M. Lipsitch, “Population genomics of post-vaccine changes in pneumococcal epidemiology,” *Nature genetics*, vol. 45, no. 6, pp. 656–663, 2013.
2. A. Dobrut and M. Brzychczy-Włoch, “Immunogenic proteins of group b streptococ- cus—potential antigens in immunodiagnostic assay for gbs detection,” *Pathogens*, vol. 11, no. 1, p. 43, 2021.
3. J. R. Verani, L. McGee, S. J. Schrag, *et al.*, “Prevention of perinatal group b streptococcal disease: revised guidelines from cdc, 2010,” 2010.
4. B. L. Pineles, K. E. Goodman, L. Pineles, and A. D. Harris, “Appropriate antibiotic use for group b streptococcus prophylaxis among penicillin-allergic patients in academic and nonaca- demic hospitals,” in *Open Forum Infectious Diseases*, vol. 9, p. ofac514, Oxford University Press US, 2022.