

Genomic epidemiology of group B streptococci spanning 10 years in an Irish Maternity Hospital, 2008-2017

Mary Meehan^{a,*}, Maeve Eogan^b, Naomi McCallion^{c, d}, Robert Cunney^{a, g}, James E Bray^e, Keith A Jolley^e, Anastasia Unitt^e, Martin CJ Maiden^e, Odile B Harrison^e and Richard J Drew^{a, f, g}

^aIrish Meningitis and Sepsis Reference Laboratory, Children's Health Ireland at Temple Street, Dublin, Ireland.

^bDepartment of Obstetrics and Gynaecology, Rotunda Hospital, Dublin, Ireland.

^cDepartment of Neonatology, The Rotunda Hospital, Dublin, Ireland;

^dDepartment of Paediatrics, The Royal College of Surgeons in Ireland, Dublin, Ireland.

^eDepartment of Zoology, Peter Medawar Building, University of Oxford, Oxford, OX1 3SY, UK.

^fClinical Innovation Unit, Rotunda Hospital, Dublin, Ireland.

^gDepartment of Clinical Microbiology, Royal College of Surgeons in Ireland, Dublin, Ireland.

Corresponding author: Mary Meehan, Irish Meningitis and Sepsis Reference Laboratory, Children's Health Ireland at Temple Street, Dublin, Ireland. email: mary.meehan@cuh.ie. Tel: 00-353-1-8784854

Key words: Group b streptococcus, invasive, PubMLST, Core genome MLST.

Running title: Genomic epidemiology of group b streptococci

Declaration of competing interest: KAJ reports grants from the Wellcome Trust during the conduct of this work and personal fees from GlaxoSmithKline outside of the submitted work. All other authors report no conflict of interest.

Abstract

Objectives: The genomic epidemiology of group b streptococcus (GBS) isolates from the Rotunda maternity hospital, Dublin, 2008-2017, was investigated.

Methods: Whole genome sequences of isolates (invasive, n = 114; non-invasive, n = 76) from infants and women were analysed using the PubMLST database (<https://pubmlst.org/sagalactiae/>).

Results: Serotypes III (36%), Ia (18%), V (17%), II (11%) and Ib, (9%) and sequence types (ST) 17 (23%), ST-23 (14%), ST-1 (12%) and ST-19 (7%) were most common. Core genome MLST (cgMLST) differentiated isolates of the same ST, grouped STs into five lineages congruent with known clonal complexes and identified known mother-baby pairs and suspected linked infant cases. Clonal complex (CC) 17 accounted for 40% and 22% of infant and maternal invasive cases, respectively and 21% of non-invasive isolates. CC23 and CC19 were associated with maternal disease (30%) and carriage (24%), respectively. Erythromycin (26%) and clindamycin (18%) resistance increased over the study period and was associated with presence of the *erm(B)* gene (55%), CC1 (33%) and CC19 (24%). A multi-resistant integrative conjugative element incorporated in the PI-1 locus was detected in CC17, an ST-12 and ST-23 isolate confirming the global dissemination of this element. All isolates possessed one or more pilus islands. Genes encoding other potential protective proteins including Sip, C5a peptidase and Srr1 were present in 100%, 99.5% and 65.8% of isolates, respectively. The *srr2* gene was unique to CC17.

Conclusions: The PubMLST.org website provides a valuable framework for genomic GBS surveillance to inform on local and global GBS epidemiology, preventive and control measures.

Introduction

Streptococcus agalactiae (group B streptococcus; GBS) is a leading cause of invasive disease in infants (1-3). Most cases occur within the first week of life and are classified as early onset disease ([EOD]; < 7 days). Disease in infants aged from 7 days to 3 months is classified as late onset (LOD). The bacterium is carried asymptomatically in the urogenital and gastrointestinal tracts and, maternal rectovaginal carriage rates of up to 35% have been reported (4). Colonization of women at or close to delivery is the principal risk factor for maternal bacteraemia, EOD, stillbirth and premature birth (2). Late onset disease is associated with vertical or community transmission in addition to acquisition in a hospital setting (5, 6). The global incidence of infant GBS disease is estimated to be 0.49 (95%CI 0.43 – 0.56) per 1000 live births (3). Maternal disease is estimated to be 0.23 (95%CI 0.09 – 0.37) per 1000 maternities in high income countries (2). The incidence of EOD has declined in countries where maternal screening and intrapartum antibiotic prophylaxis (IAP) have been implemented (1, 7). Some countries, including the UK and Ireland, have noted an increase in infant disease despite the presence of national preventive risk-based guidelines (0.72 per 1000 live births, 2000-01 and 0.94 per 1000 live births, 2014-15) (7). Given the persistence of GBS disease in infants and concern over emerging antimicrobial resistance, maternal vaccination is considered to be the most promising preventive measure to protect infants and mothers (1, 8).

Group b streptococci express ten known capsular polysaccharide serotypes (cps Ia, Ib and II-IX). Serotype III is most common in infant disease and, cps Ia, III, and V are most frequent in adult carriage and disease (1). Maternal antibodies against type-specific capsular polysaccharide are protective (9) and trans-placental transfer of vaccine-specific antibodies was confirmed in phase II trials of a trivalent (Ia, Ib and III) conjugate vaccine (10). Focus has now shifted to pentavalent (Ia, Ib, II, III, V) and hexavalent

conjugate (additionally covering cps IV) vaccines covering additional serotypes (11). Several protein-based vaccines are also at various stages of development (12, 13).

Group b streptococcus can be further classified through multilocus sequence typing (MLST; <https://pubmlst.org/sagalactiae/>) (14). Sequence types (ST) of most human strains of GBS cluster into five main clonal complexes (CC1, CC10/12, CC17, CC19, CC23) (15). Clonal complex 17, is the leading cause of invasive disease and meningitis in infants (16) whereas CC1, CC12 and CC23 are common in adult disease (1). CC17, referred to worldwide as the GBS hypervirulent clone, is almost entirely cps III although capsular switching to cps IV has been described (17, 18). The other main clonal complexes (CC1, CC10/12, CC19, CC23) can present with multiple serotypes but are mainly associated with cps V, Ib/II, III and Ia, respectively (15).

Whole genome sequencing (WGS) has revolutionised the field of clinical microbiology, providing clinicians and public health officials with the means to identify outbreaks, enhance surveillance and understand the population biology of bacterial pathogens (19). Comparative WGS analysis is based on single nucleotide polymorphisms (SNP) and/or gene-by-gene comparison of the core genome (cg) or the whole genome (wg) (20). The SNP-based approach requires inter-lab standardization and the choice of reference genome is important. The gene-by-gene approach employs *de novo* assembly and the large number of loci in cg and wg MLST provides the high resolution required to distinguish between closely related strains (21). The PubMLST *Neisseria* database (<https://pubmlst.org/neisseria>) has for several years provided the *Neisseria* scientific community with the tools to store and analyse WGS data (22). This has allowed, for example, the epidemiology of a highly virulent serogroup W *Neisseria meningitidis* strain to be characterized leading to the recommendation for immunisation of the ACWY conjugate vaccine to teenagers in the UK (23). In contrast to *N. meningitidis*, fewer WGS studies have been undertaken with GBS, although they are increasing.

Here, we examined the epidemiology of invasive and carried GBS isolates obtained from infants and women of child-bearing years collected over a 10-year period (2008-2017) from the Rotunda Maternity Hospital Dublin. This study made use of the publicly available PubMLST *S. agalactiae* database where through a gene-by-gene approach, WGS data belonging to GBS were annotated, curated and compared in association with phenotypic and accompanying provenance metadata.

Methods

Bacterial strains

Whole genome sequencing was performed on 190 GBS isolates recovered from clinical specimens in the Rotunda Maternity Hospital in Ireland during the period January 2008 to December 2017 and submitted to the Irish Meningitis and Sepsis Reference Laboratory, which is based at Children's Health Ireland at Temple Street. The Rotunda Hospital delivers approximately 8,500 babies per year serving a catchment area of approximately 1.08 million people representing 22.7% of the national population. Isolates included 114 from invasive GBS disease (defined as GBS cultured from blood or another sterile site), comprising those from EOD (n = 34), LOD (n= 16), maternal bacteraemia (n = 60) cases and four perinatal cases (Table S1). The genomes of a random selection of non-invasive or colonizing isolates (herein referred to as non-invasive) collected over the 10-year period were also sequenced including 71 isolates from women of child-bearing age (information was not available on whether women were pregnant or were gynaecology patients) and five from infants (< 90 days). Thirty of the sequenced non-invasive isolates were from 2017 and between zero and 14 isolates in other years (Table S1). Isolates were cultured overnight at 37°C in 5% CO₂ on Columbia horse blood agar plates. Some isolates in this study were previously reported (18, 24).

Serotype and ST were routinely determined for all isolates. Capsular polysaccharide serotypes were determined by latex co-agglutination (Statens Serum Institute, Denmark) and by PCR as previously

described (18). Gene fragments of the seven housekeeping genes employed in MLST were amplified by PCR and sequenced (14). Sequence type and allele numbers were assigned using PubMLST. All GBS isolates were routinely tested for susceptibility to erythromycin, clindamycin, benzylpenicillin, cefotaxime and vancomycin using E-test® (BioMérieux, Marcy l'Etoile, France) and to tetracycline using disc diffusion. Methods and interpretative criteria were according to Clinical and Laboratory Standards Institute (CLSI) guidelines (<https://clsi.org/>).

Whole genome sequencing and analysis

Chromosomal DNA was extracted using the PurElute™ bacterial genomic kit (Edgebiosystems, San Jose, CA) from GBS cells pre-treated with mutanolysin. Genomic DNA was sequenced on an Illumina 2500 HiSeq sequencer (Wellcome Trust Center for Human Genetics, University of Oxford, UK) using 150-bp paired-end reads. The resulting short-read sequences were assembled *de novo* using Velvet (25) with the VelvetOptimiser algorithm (<https://github.com/tseemann/VelvetOptimiser>) as part of a pipeline developed in Department of Zoology, University of Oxford. Resulting assemblies (Table S1) were deposited in the *S. agalactiae* PubMLST database, which uses the Bacterial Isolate Genome Sequence Database (BIGSdb) platform (22). Short read sequences are deposited in the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>) under the accession number PRJEB26339.

A total of 2439 GBS loci are defined in the PubMLST *S. agalactiae* database (accessed 01-Feb-2021). These are characterised by the nomenclature 'SAG' followed by an arbitrary 4-digit number. Loci are then organised into schemes associated with function with 43 distinct schemes developed ranging from metabolism, capsular polysaccharide biosynthesis, pilus islands to haemolysin operon (Table S2). The antimicrobial resistance scheme was added during preparation of this manuscript. Assembled contigs deposited in the database are automatically curated and annotated for all defined loci, including MLST profiles, by the BIGSdb platform that runs the PubMLST databases (26). This uses the BLAST algorithm

to query sequences against a database containing all defined alleles. New alleles are automatically assigned when these share $\geq 98\%$ sequence identity to previously defined alleles (22). Alleles with less than 98% identity are manually curated. The core genome MLST scheme (h_*S.agalactiae* cgMLST v1.0) consists of 1405 loci variously associated with metabolism, capsule and b carbohydrate synthesis, oxidative phosphorylation, the pentose phosphate pathway, protein translation, the TCA cycle, pilus regulation in addition to loci encoding the C5s peptidase and haemolysin (27).

Comparative genome analysis was undertaken by means of the Genome Comparator tool implemented within PubMLST as previously described (22). The total sum of allele differences for each locus following pairwise comparison of isolate allele profiles yielded output distance matrices (Nexus format) which were used to generate neighbour net and neighbour joining graphs using SplitsTree (v4.14.6) (28). The BIGSdb Genome Comparator tool was also used to compare the profiles of the capsular-specific genes against serotypes pre-determined by PCR and serology.

Single nucleotide polymorphisms were called using Snippy (version 4.6.0, <https://github.com/tseemann/snippy>). GUBBINS (version 2.4.1) was used to remove regions of high SNP density due to suspected recombination (29). Reference genomes included *S. agalactiae* strains SS1 (NZ_CP010867.1), B507 (CP021771.1), COH1 (NZ_HG939456.1), 2603V (NC_004116.1), CCH210801006 (ERS337511) and NZ_CP007572.1 (GBS6) for isolates from CC1, CC10/12, CC17/cps III, CC19, CC23/cps Ia and ST-525, respectively.

Detection of genes encoding antimicrobial resistance and immunogenic proteins

Antimicrobial resistance genes were identified in contigs using ABRicate (<https://github.com/tseemann/abrigate>) and agreement between CARD (30), ARG-ANNOT (31), Resfinder (32) and NCBI databases (33). This information was used to develop an antimicrobial resistance scheme consisting of acquired resistance genes in <https://pubmlst.org/sagalactiae/>.

Sequences of genes encoding several potential protective proteins including the Alp protein family, the serine rich repeat proteins, the C5a and Lmb proteins and the Streptococcal immunogenic protein (Sip) were screened for using the Genome Comparator tool and, the PubMLST isolate BLAST analysis tool (plugin v1.5.1 running BLAST v2.6.0+) for genes that are not curated (22, 34).

Ethical Approval

Patient (age, disease category, year of isolate, mother/baby pairs, suspected linked infant cases) and isolate (MIC, ST, serotype) metadata were recorded in a database linked to sequence information. The data were then irrevocably anonymised from any patient identifiable information (name, hospital number, specimen number and data of birth). Most non-invasive isolates were randomly collected and had limited to no associated patient metadata. The work was approved by the Research Ethics Committee, The Rotunda Hospital, Dublin, Ireland (reference No. RAG-2018-016)

Results

Serotype and sequence type distribution

Whole genome sequences were obtained for 85.9% (114 of 128 cases) of all infant and maternal invasive cases that occurred in the Rotunda hospital, 2008-2017. Cases included 34 of 41 EOD cases, 16 of 22 LOD cases and 60 of 65 maternal bacteraemia cases (Table S3). The 14 cases not included in this study were either diagnosed by PCR only or isolates were not available for this study. The average incidence of infant disease (0.56 per 1000 live births) calculated from all cases (n=63) over the 10 year period was within the range of global estimates which are considered minimum rates. Total annual incidence rates of EOD and LOD cases ranged from 0.12 to 0.79 per 1000 live births and 0 – 0.34 per 1000 live births, respectively (Table S4). The average rate (0.6 per 1000 maternities) of all maternal cases (n=65) was higher than the estimated incidence (0.23 per 1000 maternities) for high income

countries (Table S4). Of isolates sequenced, the most commonly expressed capsular serotypes were cps III, Ia and V (36%, 18% and 17%, respectively) followed by cps II, Ib and IV (11%, 9%, and 3%, respectively). Serotypes VI –VII and IX were represented by one and two isolates, respectively (Table 1). Seven isolates were non-typeable (NT). Overall, serotype III was the most frequent (38% invasive and 34% non-invasive isolates). Serotype Ia was more prevalent in invasive (24%) than non-invasive isolates (9%). Isolates grouped into 50 STs. Sequence types 17, ST-23, ST-1, ST-19 and ST-12 were most common (23% [n=44], 14% [n=26], 12% [n=23], 7% [n=13] and 5% [n=10], respectively). Sequence type-17 was exclusively serotype III whereas the other main STs presented different serotypes (ST-23: Ia [n=23], III [n=2], V [n=1]); ST-1: V [n=16], Ib [n=2], II [n=1], IV [n=1], VII [n=1], NT [n=2]; ST-19: III [n=10], V [n=2], NT [n=1]; ST-12: II [n=5], Ib [n=3], V [n=1], NT [n=1]). Serotypes and STs pre-determined by serology/PCR and Sanger sequencing, respectively were in agreement isolate with gene/allelic profiles called on PubMLST.

Core genome MLST analysis and disease association

Of the 1405 loci in the cgMLST scheme, 1336 were present in 90-100% of isolates. Most isolates clustered (n=182 of 190; 96%) into the five main CCs of human-associated GBS (Fig. 1; CC1 n=34, CC10/12 n=24, CC17 n=49, CC19 n=33, CC23 n=42) and grouped within CCs according to serotype. The founder of each CC (ST-1, ST-12, ST-19, ST-17 and ST-23) was the most frequent (67% of CC1; 42% of CC10/12; 90% of CC17; 39% of CC19; 62% of CC23) (Fig. 1). Seven loci had identical alleles in all isolates and encoded ribosomal proteins (SAG0129, SAG1358, SAG1368, SAG1429, and SAG2109), a hypothetical protein (SAG1214) and the PTS system mannose-specific transporter subunit IIA (SAG1652). Clonal complex 1, CC10/CC12, and CC23 isolates had 303, 375, and 380 identical allelic loci, respectively and CC17 and CC19 each had 458 identical loci. Analysis of pairwise locus distances showed that CC1 and CC17 had the highest (0-713 [mean \pm SD; 270 \pm 202]) and lowest pairwise locus differences (0-286; mean

\pm SD, 115 \pm 44; Fig. S1). The other CCs had pairwise locus differences ranging of 4–612 [301 \pm 142], 0–455 [237 \pm 105], 0–671 [221 \pm 172], for CC10/12, CC19 and CC23, respectively (Fig. S1). These differences are consistent with recombination rates based on SNP distances previously reported by Da Cunha *et al.* demonstrating a very low rate of recombination for CC17 involving 3% of the genome compared to a high recombination rate for CC1 of 47% of the genome (15). In agreement with other studies, however, most CC1/cps V isolates (20 of 21) formed a tight cluster (pairwise differences 0 – 98; mean \pm SD, 71 \pm 14) (35).

Invasive and non-invasive isolates distributed among clonal complexes (Fig. 2, Fig. S2 – S6). Clonal complex 1, CC17 and CC23 accounted for 80% of infant disease (24%, 40% and 16%, respectively; Table 1). Clonal complex 17 was most frequent in infant disease (EOD, 14 of 34 [41%]; LOD, 6 of 16 [37.5%]) and was more frequent in infants compared to maternal bacteraemia/perinatal cases (20%, p = 0.023 OR 2.62 [95%CI 1.14 – 6.00]) or non-invasive infections (21%) therefore supporting the over representation of this clonal complex in this patient cohort (1, 36, 37). Comparison of Irish isolates with CC17 from other global locations revealed two main clades (Clade 1 and 2; Fig S7). Most isolates, including those from Ireland (78%; n = 38 of 49)) clustered into Clade 1.

Clonal complex 23 and CC17 were the two most frequent clonal complexes in maternal disease (30% and 20%, respectively) as reported in other studies (1) (Table 1, Fig. 2). From 2012, isolates (n =9) of the ST-24 sublineage of CC23 were detected and represented by two serotype clusters (cps Ia [n =5] and cps V [n = 4]). The ST-24/cps Ia sublineage has been associated with disease in the Mediterranean region (36). Seventy nine and 85 percent of CC10/CC12 and CC19 isolates were associated with invasive and non-invasive maternal cases (CC12: p = 0.25, OR 1.83 [95%CI 0.65–5.17]; CC19: p = 0.036, OR 2.94 [95%CI 1.07–8.04]). Overall, CC10/CC12 and CC19 accounted for 13% and 17% of maternal bacteraemia

and 16% and 25% of non-invasive maternal isolates, respectively and were less frequent in infant disease (CC10/CC12: 6% [3 of 50]; CC19: 8%[4 of 50]).

The CovRS (control of virulence regulator and sensor) two-component system plays a key role in modulating virulence-associated genes and GBS pathogenicity (38-40). There were seven and 14 alleles of *covR* (SAG1625) and *covS* (SAG1624), respectively associated with CC1 (n=32), CC12 (n=5), CC17 (n=2) and singletons (n=3) resulting in two and nine protein variants compared to sequences in *S. agalactiae* 2603V, respectively (Table S5). The CovR variants (A70V and E123K) were present in two invasive isolates (ST-516 and ST-17). There was no significant difference in the distribution of CovS variants between invasive and non-invasive isolates (Table S5). To the authors knowledge the CovRS variants reported in this present study have not been previously documented and their role in GBS invasive disease is unclear at present (38-40).

Detection of infection clusters

The isolate collection contained a number of known or suspected linked cases. The resolution of cgMLST to identify these cases was investigated. There were five pairs of isolates collected from maternal sources and EOD and, one maternal/stillbirth isolate pair (Table 2). Core genome MLST confirmed their linkage and showed that these six isolate pairs differed by 0 to 3 cgMLST loci (Table 2). There were two clusters (cluster 7 and 8) involving infant pairs where there had been suspicion of cross infection. Cluster 7 consisted of two non-typeable GBS of ST-1163 with identical cgMLST loci collected from two infants with non-invasive respiratory tract infections. Cluster 8 consisted of an ST-8 isolate collected from a colonized index case (skin) which differed by four loci from an invasive LOD case strongly supporting cross infection. Two pairs of maternal isolates from blood and vaginal swabs in the same year had identical cgMLST loci (cluster 9 and 10). No patient demographic data were provided for the two non-invasive isolates and it was not possible to confirm relatedness with the respective invasive

cases. Analysis of the lowest pairwise loci differences between all other unrelated (to the authors' knowledge) isolates revealed differences of 37, 40, 19, 12 and 17 loci for CC1, CC10/CC12, CC17, CC19 and CC23 isolates, respectively. Single nucleotide polymorphism analysis confirmed the results of cgMLST and showed that eight clusters had no SNP differences and two clusters (clusters 4 and 8) differed by one SNP. The next closest isolate pairs differed by 21, 38 and 20 SNPs for CC12, CC17/ cps III and CC23/cps Ia and, 14 SNPs for CC1 and CC19 (data not shown). These latter two isolate pairs were collected from different patients several years apart.

Characterisation of non-typeable isolates

In addition to frequent capsular exchange, loss of capsular expression occurs in 5-20% of GBS isolates (41, 42). In this study, NT isolates were at a low frequency (3.7%, n = 7; Table 1). The majority of NT isolates were from non-invasive infections (n = 6 of 7) as reported elsewhere (43). One NT isolate was collected from a stillbirth. Serotypes were determined by PCR and WGS for five of these isolates (Table 3). Two isolates from an infant case cluster could not be typed by either serology or PCR. To identify sequence alterations predicted to have resulted in loss of capsule expression, the *cps* (SAG1162 – 2176; SAG2247 – 2248) and *neu* (SAG1158 – SAG1161, SAG2185) genes of the NT isolates were compared to encapsulated isolates from the same CC using the Genome Comparator tool on PubMLST and by alignment of the *cps* operons and 5' promoter regions (Table 3). Four isolates (ST-1 [n=2], ST-12 [n=1], ST-19 [n=1]) possessed single nucleotide deletions/mutations in the *cpsE* (SAG1171), *cpsG* (SAG1169) or *cpsN* (SAG1166) genes encoding the first two (essential) enzymes in polysaccharide synthesis (44) and the *cpsN* glycosyl transferase gene, respectively (Table 3). No detectable sequence alterations in the capsular operon gene were detected in one ST-27 isolate. Two ST-1163 isolates collected from a non-invasive infection cluster of two infants lacked complete *cps* genes and likely arose due to a

recombination event between direct repeats in *neuB* (SAG1161) and *cpsA* (SAG1175) and the excision of an 11.1 kb genome fragment (Fig. S8).

Distribution of antimicrobial resistance genes

Susceptibility testing revealed that all isolates were susceptible to penicillin, cefotaxime and vancomycin (Table S6). Twelve isolates possessed a penicillin MIC (0.12 µg/ml) on the threshold for non-susceptibility (45). Mutations in the *pbp2x* gene encoding penicillin binding protein 2X are associated with reduced susceptibility to beta-lactams although other *pbp* genes may additionally be involved (45, 46). Analysis of the distributions of *pbp* gene polymorphisms for all isolates revealed 19, 20, 16, 20 and 14 allelic variants for *pbp1a* (SAG0298), *pbp1b* (SAG0159), *pbp2x* (SAG0287), *pbp2a* (SAG2066), and *pbp2b* (SAG0765), respectively resulting in six to twelve protein variants which exhibited CC-specific distributions (Table S7). Comparison of PBP2X substitutions with the reference *S. agalactiae* 2603V sequence revealed four (G329V, I377V, G627V P396L) were previously described and associated with penicillin susceptibility. The remaining substitution (T718S) has not been previously documented and is distant from the active domain (₄₀₂SSN₄₀₄ and ₅₅₂KSG₅₅₄) (45, 46).

Forty-nine and thirty four isolates (26% and 17.9%) were resistant to erythromycin and clindamycin, respectively. There was somewhat higher frequency of macrolide resistant isolates among non-invasive isolates (30%, [23/76]) compared to invasive isolates (23% [26/114], Table 1, Fig. S2 – S6). Comparison of the macrolide resistance rates showed that macrolide rates increased in the last five years of the study compared to the first five years (18% [6/33], 18% [6/33] and 16% [3/19], 2008-2012; 29% [5/17], 35% [11/31] and 35% [20/57], 2013-2017) for infant disease, maternal/perinatal disease and non-invasive infections, respectively). The distributions of constitutive MLS_B (cMLS_B), inducible MLS_B (iMLS_B) and M phenotypes were 65% (n=32), 18% (n=9) and 16% (n=8) encoded by the methylase genes *erm*(B) (SAG2253; n=28, 88%), *erm*(TR) (SAG2235; n=9, 100%) and, the efflux *mef* (SAG2254) and *msr*(D) genes

(SAG2258; n=8, 100%), respectively (Table 1). Two CC19 isolates (ST19 and ST861) exhibited the L phenotype (erythromycin sensitive, clindamycin resistant) and harboured *Inu*(B) (SAG2255) and *Isa*(E) (CP021773.1:c1449340-1447856) and, *Isa*(C) (SAG2256) genes, respectively. No resistance gene was detected in one macrolide resistant CC19 isolate. BLAST analysis against all *erm* alleles known to be present in group b streptococcus did not reveal any novel homologues (<http://faculty.washington.edu/marilynr/ermweb1.pdf>). CC1 and CC19 were most associated with erythromycin and clindamycin resistance with 50% ($p < 0.002$ OR 3.29 [95%CI 1.91–6.94]) and 33% ($p = 0.17$ OR 1.73 [95%CI 0.78–3.83]) of isolates exhibiting resistance, respectively. Clonal complex 10/CC12, CC17 and CC23 each possessed 17–18% resistance. There was 88% (n=168) tetracycline resistance, associated with *tet*(M) (SAG0923; n=148, 88%) and/or *tet*(O) (SAG2251; n=25, 15%), confirming the stability of this resistance phenotype (15). CC10/CC12 had the lowest tetracycline resistance rate (67%, 16 of 24) compared to the other CCs (91%–93%).

Aminoglycoside resistance, predicted from the detection of resistance genes only, was detected in eleven isolates (*aph*(3')-III [SAG2252] and *ant*(6) [SAG2029], n= 9; *aph*(3')-III, n=1; *ant*(6), n=1). Eight of these isolates (ST-17 n=6 [two from an infection cluster], ST=12 [n=1], ST-23 [n=1]) notably lacked pilus island-1 (SAG0645 – SAG0650) and additionally harboured *erm*(B) and *tet*(O). Sequence analysis revealed the insertion of a large 74kb integrative conjugative element (ICE) from the ICESa2603 family (<http://db-mmml.sjtu.edu.cn/ICEberg/>) at the same location on the genome as the pilus-1 island (47). All ICE had sequences essentially similar (99% identity) to ICESag37 (CP019978.1; Fig. S9) previously reported in an ST12 isolate in China and detected in ST-17 strains lacking PI-1 (48-50).

Potential vaccine coverage

Vaccine coverage for invasive isolates was predicted to be 70% (72% Infants, 69% maternal bacteraemia/perinatal cases) for a trivalent (Ia, Ib and III) conjugate vaccine, 95% (96% Infants, 94%

maternal bacteraemia/perinatal cases) for a pentavalent vaccine (cps Ia, Ib, II, III, and V) and 98% (98% Infants, 98% maternal bacteraemia/perinatal cases) for a hexavalent conjugate vaccine (cps Ia, Ib, II-V) (12). Reduced vaccine coverage was predicted for non-invasive isolates (53%, 87% and 88%, respectively).

The distributions of genes encoding a number of potential protein vaccine candidates were determined (12, 13). All isolates possessed at least one pilus island (Fig. 3 and Fig. S2 – S6). The most common pilus island combinations were: i) PI-1 and PI-2a (SAG0645 – SAG0650 and SAG1404 – SAG1408; n = 81), ii) PI-2a (n = 57) and, iii) PI-1 and PI-2b (SAG2190–SAG2194; n =45). The *srr1* (SAG1462) gene was prevalent in CC1, CC12 and CC23 (95% – 100%) but less frequently found in CC19 (61%, n = 20 of 33) with *srr1* absence mainly associated with ST-28 and non-invasive isolates (absent in 56% and 20% of non-invasive and invasive, respectively). The *srr2* gene (AY669067.1) was present only in CC17 isolates (100%). Genes encoding Sip (SAG0032) and, C5a peptidase (SAG1236) and the associated laminin-binding protein (Lmb; [SAG1235]), data not shown) were present in all isolates with the exception of one (singleton) ST-485 isolate that lacked the C5a peptidase and Lmb protein genes. All isolates possessed genes encoding one of the Alp protein family with Rib (SAG0748), Alp1 (CP051114.1), Alp3 (SAG2198) and, the alpha (SAG2196) and/or beta (SAG2195) components of C-antigen most frequent in CC17 and CC19 (98%), CC23 (64%), CC1 (72%) and CC12 (100%), respectively.

Discussion

GBS is a significant cause of infant morbidity and mortality. Invasive disease in mothers and older adults with co-morbidities can also have devastating outcomes (1-3). Despite advances in prevention and treatment, GBS remains a global public health concern. To date, there have been limited GBS genomic studies in the Republic of Ireland (5, 7). Here, the population genomics of GBS collected from infants and mothers in an Irish maternity hospital was analysed using a gene-by-gene approach facilitated by the S.

agalactiae database on PubMLST. This web-based resource, provides open access easy-to-use genomic analysis tools which do not require extensive command line expertise (22). BIGSdb implemented in PubMLST links isolate provenance/phenotypic and sequence data to an expandable catalogue of loci. Newly identified loci can be added to the existing locus catalogue. At the time of writing (07-Feb-21), 10160 GBS genomes originating from 35 countries and spanning nearly seven decades (1953 to 2020) have been deposited in the database. It is acknowledged that genomic comparative studies using the gene-by-gene approach can have limitations including the fact that there may be issues with duplicated genes and repetitive regions (both of which cannot be easily reconstructed using current *de novo* assemblers). Gene-by-gene approaches will not include information about regions outside of coding sequences or loci not yet added to a typing scheme. This is similar to reference-based mapping which will only include sequences present in the reference genome. Analyses of paralogous genes and repeat regions will also be problematic using SNP-mapping approaches with additional challenges encountered when applied to more diverse bacterial species undergoing extensive horizontal gene transfer (20, 22).

In the absence of a vaccine, antimicrobials are vital in the treatment and prevention of GBS. There has, however, been a worrying decline in coverage for potential therapeutics. During the present study, an acquired antimicrobial resistance scheme was added to PubMLST enhancing surveillance of antimicrobial resistance. All of the isolates examined here were susceptible to penicillin, the antibiotic of choice for GBS and no isolates possessed PBP2X mutations associated with reduced beta-lactam susceptibility which have been detected in several global locations at low frequencies (45, 46, 51). As reported in this and other worldwide studies, there has been increased (rates of 16-80%) resistance to lincosamides and macrolides (used in IAP for those allergic to penicillin and prophylaxis for premature rupture of membranes, respectively) (36, 52, 53) and resulting in revised guidelines recommending cephalosporin or vancomycin (for severe penicillin allergy) for IAP (54). Resistance to vancomycin (the last line antibiotic) associated with *vanG* has been reported in two isolates in the United States (55). In

the present study, erythromycin and clindamycin resistance was more prevalent in CC1 and CC19 isolates but was detected in all CCs as reported elsewhere (36). In 2016 and 2017, we detected a CC17 clone in our isolate collection which harboured a mobile genetic element conferring multiple resistance to tetracycline, macrolides, lincosamides and aminoglycosides, the latter which is used in severe infections in combination with penicillin. This clone has been reported in many global locations including Canada, Portugal and France from 2010-2011 and China from 2015 (36, 50, 52, 56, 57). Interrogation of WGS data available on PubMLST (6-Oct-2020) revealed a low frequency of this element in ST-12 (3 of 313, 1%) and ST-23 genomes (1 of 1447 ST23, 0.07%) compared to ST-17 (76 of 1416 [5.5%]) indicating these resistant genotypes have not expanded in a manner similar to multi-resistant CC17. However, as submission to PubMLST is voluntary these calculated frequencies are an estimate and may not be a true representation of the overall global GBS population.

Core genome MLST analysis revealed a population structure congruent with that based on the seven-gene MLST (14). Most isolates grouped into the five main CCs (CC1, CC12, CC17, CC19 and CC23) associated with human isolates (15). Each CC was characterised by a small number of closely related sub clones consisting of a dominant serotype or minor serotypes indicating frequent capsular exchange as reported elsewhere (15). The limited number of tetracycline-resistant CCs and clusters within each CC is consistent with studies showing an evolutionary bottleneck in the GBS population due to tetracycline usage from the mid-twentieth century and the comparatively recent expansion of tetracycline resistant clones associated with human disease (15). In agreement with numerous studies, CCs differed in disease potential with CC1, CC17 and CC23 associated with invasive disease and, CC12 and CC19 associated with non-invasive infections (1). A number of CC-specific genes have been recently identified which may play a role in invasion or colonisation (27). CC17 in particular possesses a number of unique genes which may contribute to its virulence in infants (58). The detection of a dominant CC17 clade in this study concurs with other recent studies demonstrating the expansion of a major CC17 clade from the 1990s (59).

A GBS vaccine is now a WHO priority with the recent ratification of the defeating meningitis by 2030 roadmap (60). Vaccination would not only protect infants and mothers but also have potential to protect older non-pregnant adults, a patient group in which GBS has been increasing. A hexavalent conjugate vaccine, designed to cover 98% of worldwide GBS, is at the most advanced stages of vaccine development with phase 1/2 trials completed and would cover 94% of our infant and mother isolate collection (11). The lack of coverage against some serotypes and NT strains and, possible capsular exchange events indicates potential for serotype replacement or capsular switching as reported in the post pneumococcal vaccine era (61). The detection of invasive NT strains in this and other studies suggest that capsule expression may not be required for all infection types (41, 42). Monitoring the GBS population in a post conjugate vaccine era will be critical.

In the present study, capsular biosynthesis loci defined on PubMLST were concordant with serotypes determined by serology and PCR and, assisted in the investigation of NT isolates. A universal GBS vaccine covering all possible genotypes would be optimal and several protein antigens have been investigated to date (12, 13). The ever-increasing availability of WGS data will help to identify new protein vaccine targets. Using Genome Comparator and BLAST tools implemented on PubMLST, we examined coverage of several potential protein vaccine candidates. A vaccine based on the N-terminal domain of the Alpha C protein and Rib (GBS-NN) proteins reached phase 1 clinical trials and, although Alp proteins were prevalent, the heterogeneity of Alp sequences may restrict their use as a universal vaccine candidate (12, 13). The presence of Sip or C5a proteins in all but one C5a peptidase-negative isolate lends further support as potential vaccine targets (12, 13). The potential for pilus island loss by genome replacement events and absence of Srr1/2 proteins in some strains may abrogate their vaccine usage (13).

GBS is a recognised causative agent of nosocomial infections particularly among infants (5, 6). Recent studies showed that there was long intervals between cases and, without the use of (SNP-based) WGS

analysis and active surveillance, over half of LOD case clusters in neonatal intensive care units would have gone undetected with resultant missed opportunities for vital infection control and prevention strategies (5, 6). Gene-by-gene and SNP-based genome analysis approaches have been shown to exhibit congruence in invasive bacterial outbreak investigations (62). In the present study, we demonstrated cgMLST and SNP-based analysis were concordant in the identification of GBS infection clusters. Our results are consistent with the recent study of Collin *et al* which demonstrated eight of nine case clusters differed by ≤ 1 SNPs and one cluster differed by three SNPs with an overall cut-off of ≤ 5 SNPs (5, 6). Of the two infant clusters in our study, cross infection was partially confirmed in one cluster without the need for WGS because the isolates were NT and possessed a previously unassigned sequence type (ST-1163). The second cluster involved an index infant colonized with ST-8 and a subsequent LOD case. Notwithstanding the low frequency of ST-8 (6 of 190), WGS was required to confirm cross infection. All invasive GBS isolates will be typed by WGS in the Irish meningitis and Sepsis Reference laboratory going forward which will aid in the proactive surveillance of future GBS case clusters.

A global meningitis genome partnership, with contributions from high, middle and low income countries, has been proposed to aid the WHO defeating meningitis roadmap (63). For GBS, WGS has been central in population surveillance and elucidating the population structure and evolution (15, 37, 49, 64). The publicly available genome sequences and metadata presented here will provide a baseline for monitoring the genomic epidemiology of infant and maternal GBS in Ireland. Global databases such as those on PubMLST.org will support on-going surveillance of GBS to inform future vaccine strategies, monitor the impact of preventive/control measures and, emergence of new virulent and antimicrobial resistant clones.

Acknowledgements

The authors gratefully acknowledge the assistance of the staff of Microbiology laboratory, Rotunda Hospital with collection of group b streptococcus isolates and the Rotunda Maternity Hospital, for funding whole genome sequencing of the GBS isolates utilised in this study. Development of PubMLST and BIGSdb is supported by a Wellcome Trust Biomedical Resource Grant (218205).

References

1. Bianchi-Jassir F, Paul P, To KN, Carreras-Abad C, Seale AC, Jauneikaite E, et al. Systematic review of Group B Streptococcal capsular types, sequence types and surface proteins as potential vaccine candidates. *Vaccine*. 2020 Sep 1.
2. Hall J, Adams NH, Bartlett L, Seale AC, Lamagni T, Bianchi-Jassir F, et al. Maternal Disease With Group B Streptococcus and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. *Clin Infect Dis*. 2017 Nov 6;65(suppl_2):S112-S24.
3. Madrid L, Seale AC, Kohli-Lynch M, Edmond KM, Lawn JE, Heath PT, et al. Infant Group B Streptococcal Disease Incidence and Serotypes Worldwide: Systematic Review and Meta-analyses. *Clin Infect Dis*. 2017 Nov 6;65(suppl_2):S160-S72.
4. Russell NJ, Seale AC, O'Driscoll M, O'Sullivan C, Bianchi-Jassir F, Gonzalez-Guarin J, et al. Maternal Colonization With Group B Streptococcus and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. *Clin Infect Dis*. 2017 Nov 6;65(suppl_2):S100-S11.
5. Collin SM, Groves N, O'Sullivan C, Jauneikaite E, Patel D, Cunney R, et al. Uncovering infant group B streptococcal (GBS) disease clusters in the UK and Ireland through genomic analysis: a population-based epidemiological study. *Clin Infect Dis*. 2020 Aug 7.
6. Jauneikaite E, Kapatai G, Davies F, Gozar I, Coelho J, Bamford KB, et al. Serial Clustering of Late-Onset Group B Streptococcal Infections in the Neonatal Unit: A Genomic Re-evaluation of Causality. *Clin Infect Dis*. 2018 Aug 31;67(6):854-60.
7. O'Sullivan CP, Lamagni T, Patel D, Efstratiou A, Cunney R, Meehan M, et al. Group B streptococcal disease in UK and Irish infants younger than 90 days, 2014-15: a prospective surveillance study. *Lancet Infect Dis*. 2019 Jan;19(1):83-90.
8. Melin P, Efstratiou A. Group B streptococcal epidemiology and vaccine needs in developed countries. *Vaccine*. 2013 Aug 28;31 Suppl 4:D31-42.
9. Baker CJ, Kasper DL. Correlation of maternal antibody deficiency with susceptibility to neonatal group B streptococcal infection. *N Engl J Med*. 1976 Apr 1;294(14):753-6.
10. Leroux-Roels G, Bebia Z, Maes C, Aerssens A, De Boever F, Grassano L, et al. Safety and immunogenicity of a second dose of an investigational maternal trivalent Group B streptococcus vaccine in non-pregnant women 4-6 years after a first dose: results from a phase 2 trial. *Clin Infect Dis*. 2019 Aug 8.
11. 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*. 2020 Sep 3.

12. Carreras-Abad C, Ramkhalawon L, Heath PT, Le Doare K. A Vaccine Against Group B Streptococcus: Recent Advances. *Infect Drug Resist.* 2020;13:1263-72.
13. Nuccitelli A, Rinaudo CD, Maione D. Group B Streptococcus vaccine: state of the art. *Ther Adv Vaccines.* 2015 May;3(3):76-90.
14. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, Kunst F, et al. Multilocus sequence typing system for group B streptococcus. *J Clin Microbiol.* 2003 Jun;41(6):2530-6.
15. Da Cunha V, Davies MR, Douarre PE, Rosinski-Chupin I, Margarit I, Spinali S, et al. Streptococcus agalactiae clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nat Commun.* 2014 Aug 4;5:4544.
16. 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.* 2012 Oct 25;207(11):2313-22.
17. Bellais S, Six A, Fouet A, Longo M, Dmytruk N, Glaser P, et al. Capsular switching in group B Streptococcus CC17 hypervirulent clone: a future challenge for polysaccharide vaccine development. *J Infect Dis.* 2012 Dec 1;206(11):1745-52.
18. Meehan M, Cunney R, Cafferkey M. Molecular epidemiology of group B streptococci in Ireland reveals a diverse population with evidence of capsular switching. *Eur J Clin Microbiol Infect Dis.* 2014 Jul;33(7):1155-62.
19. Van Goethem N, Descamps T, Devleeschauwer B, Roosens NHC, Boon NAM, Van Oyen H, et al. Status and potential of bacterial genomics for public health practice: a scoping review. *Implement Sci.* 2019 Aug 13;14(1):79.
20. Schurch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect.* 2018 Apr;24(4):350-4.
21. Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol.* 2013 Oct;11(10):728-36.
22. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 2018;3:124.
23. Lucidarme J, Hill DM, Bratcher HB, Gray SJ, du Plessis M, Tsang RS, et al. Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage. *J Infect.* 2015 Nov;71(5):544-52.
24. O'Connor C, Meehan M, Cunney R, Eogan M, Drew RJ. Unexpected increase in invasive maternal Group B Streptococcus bacteraemia in a maternity hospital in Dublin, Ireland (May to September 2017) associated with subtle signs of sepsis and unreliable sepsis biomarkers. *J Infect.* 2018 Apr;76(4):426-8.
25. Zerbino DR. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics.* 2010 Sep;Chapter 11:Unit 11 5.
26. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics.* 2010 Dec 10;11:595.
27. Gori A, Harrison OB, Mlia E, Nishihara Y, Chan JM, Msefula J, et al. Pan-GWAS of Streptococcus agalactiae Highlights Lineage-Specific Genes Associated with Virulence and Niche Adaptation. *mBio.* 2020 Jun 9;11(3).
28. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics.* 1998;14(1):68-73.
29. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 2014 Feb 18;43(3):e15.

30. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D566-D73.
31. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother.* 2014;58(1):212-20.
32. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* 2012 Nov;67(11):2640-4.
33. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob Agents Chemother.* 2019 Nov;63(11):e00483-19.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990 Oct 5;215(3):403-10.
35. Flores AR, Galloway-Pena J, Sahasrabhojane P, Saldana M, Yao H, Su X, et al. Sequence type 1 group B *Streptococcus*, an emerging cause of invasive disease in adults, evolves by small genetic changes. *Proc Natl Acad Sci U S A.* 2015 May 19;112(20):6431-6.
36. Martins ER, Pedroso-Roussado C, Melo-Cristino J, Ramirez M. *Streptococcus agalactiae* Causing Neonatal Infections in Portugal (2005-2015): Diversification and Emergence of a CC17/PI-2b Multidrug Resistant Sublineage. *Front Microbiol.* 2017;8:499.
37. 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 US: 2015-2017. *Clin Infect Dis.* 2020 Feb 15.
38. Almeida A, Rosinski-Chupin I, Plainvert C, Douarre PE, Borrego MJ, Poyart C, et al. Parallel Evolution of Group B *Streptococcus* Hypervirulent Clonal Complex 17 Unveils New Pathoadaptive Mutations. *mSystems.* 2017 Sep-Oct;2(5).
39. Lupo A, Ruppen C, Hemphill A, Spellerberg B, Sendi P. Phenotypic and molecular characterization of hyperpigmented group B *Streptococci*. *Int J Med Microbiol.* 2014 Jul;304(5-6):717-24.
40. Lembo A, Gurney MA, Burnside K, Banerjee A, de los Reyes M, Connelly JE, et al. Regulation of CovR expression in Group B *Streptococcus* impacts blood-brain barrier penetration. *Mol Microbiol.* 2010 Jul;77(2):431-43.
41. Ippolito DL, James WA, Tinnemore D, Huang RR, Dehart MJ, Williams J, et al. Group B streptococcus serotype prevalence in reproductive-age women at a tertiary care military medical center relative to global serotype distribution. *BMC Infect Dis.* 2010 Nov 24;10:336.
42. Rosini R, Campisi E, De Chiara M, Tettelin H, Rinaudo D, Toniolo C, et al. Genomic analysis reveals the molecular basis for capsule loss in the group B *Streptococcus* population. *PLoS One.* 2015;10(5):e0125985.
43. Teatero S, Ferrieri P, Martin I, Demczuk W, McGeer A, Fittipaldi N. Serotype Distribution, Population Structure, and Antimicrobial Resistance of Group B *Streptococcus* Strains Recovered from Colonized Pregnant Women. *J Clin Microbiol.* 2017 Feb;55(2):412-22.
44. Cieslewicz MJ, Kasper DL, Wang Y, Wessels MR. Functional analysis in type Ia group B *Streptococcus* of a cluster of genes involved in extracellular polysaccharide production by diverse species of streptococci. *J Biol Chem.* 2001 Jan 5;276(1):139-46.
45. Kobayashi M, McGee L, Chochua S, Apostol M, Alden NB, Farley MM, et al. Low but Increasing Prevalence of Reduced Beta-lactam Susceptibility Among Invasive Group B *Streptococcal* Isolates, US Population-Based Surveillance, 1998-2018. *Open Forum Infect Dis.* 2021 Feb;8(2):ofaa634.
46. van der Linden M, Mamede R, Levina N, Helwig P, Vila-Cerqueira P, Carrico JA, et al. Heterogeneity of penicillin-non-susceptible group B streptococci isolated from a single patient in Germany. *J Antimicrob Chemother.* 2019 Nov 18.

47. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D660-D5.
48. Campisi E, Rosini R, Ji W, Guidotti S, Rojas-Lopez M, Geng G, et al. Genomic Analysis Reveals Multi-Drug Resistance Clusters in Group B Streptococcus CC17 Hypervirulent Isolates Causing Neonatal Invasive Disease in Southern Mainland China. *Front Microbiol.* 2016;7:1265.
49. Teatero S, Ramoutar E, McGeer A, Li A, Melano RG, Wasserscheid J, et al. Clonal Complex 17 Group B Streptococcus strains causing invasive disease in neonates and adults originate from the same genetic pool. *Sci Rep.* 2016 Feb 4;6:20047.
50. Zhou K, Xie L, Han L, Guo X, Wang Y, Sun J. ICESag37, a Novel Integrative and Conjugative Element Carrying Antimicrobial Resistance Genes and Potential Virulence Factors in *Streptococcus agalactiae*. *Front Microbiol.* 2017;8:1921.
51. Kimura K, Suzuki S, Wachino J, Kurokawa H, Yamane K, Shibata N, et al. First molecular characterization of group B streptococci with reduced penicillin susceptibility. *Antimicrob Agents Chemother.* 2008 Aug;52(8):2890-7.
52. Plainvert C, Anselem O, Joubrel C, Marcou V, Falloukh A, Frigo A, et al. Persistence of group B Streptococcus vaginal colonization and prevalence of hypervirulent CC-17 clone correlate with the country of birth: a prospective 3-month follow-up cohort study. *Eur J Clin Microbiol Infect Dis.* 2020 Aug 18.
53. Gao K, Guan X, Zeng L, Qian J, Zhu S, Deng Q, et al. An increasing trend of neonatal invasive multidrug-resistant group B streptococcus infections in southern China, 2011-2017. *Infect Drug Resist.* 2018;11:2561-9.
54. Hughes RG, Brocklehurst P, Steer PJ, Heath P, Gynaecologists. SBobotRCoOa. Prevention of early-onset neonatal group B streptococcal disease. Green-top Guideline No. 36. *British Journal of Gynaecology.* 2017;124:e280–e305.
55. Park C, Nichols M, Schrag SJ. Two cases of invasive vancomycin-resistant group B streptococcus infection. *N Engl J Med.* 2014 Feb 27;370(9):885-6.
56. Zhou K, Zhu D, Tao Y, Xie L, Han L, Zhang Y, et al. New genetic context of lnu(B) composed of two multi-resistance gene clusters in clinical *Streptococcus agalactiae* ST-19 strains. *Antimicrob Resist Infect Control.* 2019;8:117.
57. Campisi E, Rosini RJW, Guidotti S, Rojas-López M, Geng G, Deng Q, et al. Genomic Analysis Reveals Multi-Drug Resistance Clusters in Group B Streptococcus CC17 Hypervirulent Isolates Causing Neonatal Invasive Disease in Southern Mainland China. *Front Microbiol.* 2016;7:1265.
58. Tazi A, Bellais S, Tardieux I, Dramsi S, Trieu-Cuot P, Poyart C. Group B Streptococcus surface proteins as major determinants for meningeal tropism. *Curr Opin Microbiol.* 2011 Feb;15(1):44-9.
59. Jamroz D, de Goffau MC, Bijlsma MW, van de Beek D, Kuijpers TW, Parkhill J, et al. Temporal population structure of invasive Group B Streptococcus during a period of rising disease incidence shows expansion of a CC17 clone. *Biorxiv.* 2018.
60. WHO. Defeating meningitis by 2030: a global roadmap. 2020:1-36.
61. Feikin DR, Kagucia EW, Loo JD, Link-Gelles R, Puhon MA, Cherian T, et al. Serotype-specific changes in invasive pneumococcal disease after pneumococcal conjugate vaccine introduction: a pooled analysis of multiple surveillance sites. *PLoS Med.* 2013;10(9):e1001517.
62. Pearce ME, Alikhan NF, Dallman TJ, Zhou Z, Grant K, Maiden MCJ. Comparative analysis of core genome MLST and SNP typing within a European *Salmonella* serovar Enteritidis outbreak. *Int J Food Microbiol.* 2018 Jun 2;274:1-11.
63. Rodgers E, Bentley SD, Borrow R, Bratcher HB, Brisse S, Brueggemann AB, et al. The global meningitis genome partnership. *J Infect.* 2020 Jun 29.
64. Furfaro LL, Chang BJ, Kahler CM, Payne MS. Genomic characterisation of perinatal Western Australian *Streptococcus agalactiae* isolates. *PLoS One.* 2019;14(10):e0223256.

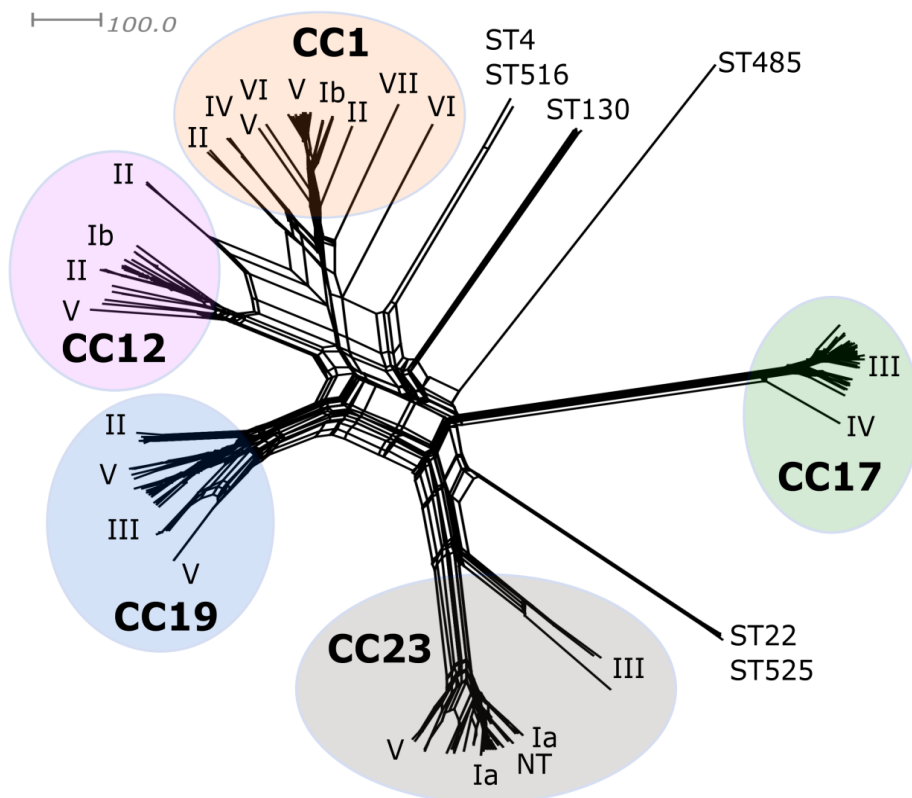


Figure 1. Neighbour-net analysis based on allelic profiles of core genome MLST loci from group b streptococcal isolates collected from the Rotunda hospital, 2008-2017. The five main clonal complexes and capsular polysaccharide genotypes (cps Ia, Ib, II – VII and non-typeable [NT]) of sub-clusters are indicated. There were eight isolates in six singleton groups that included ST-4 (cps Ia), ST-22 (cps II), ST-130 (IX), ST485 (cps Ia), ST-516 (cps Ia) and ST-525 (cps II).

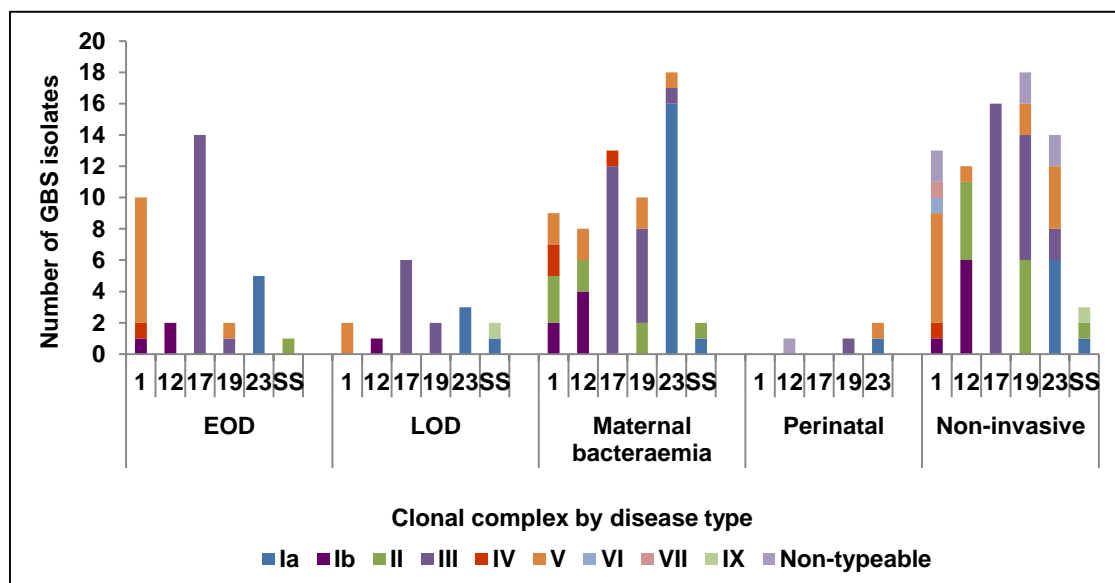


Figure 2. Serotype and clonal complex distribution of group B streptococcal isolates by disease type collected from the Rotunda Hospital Dublin, 2008-2017. Isolates grouped in clonal complex (CC) 1, 12, 17, 19 and 23 or singletons not associated with any clonal complex. Seven isolates were non-typeable by serology. Capsular polysaccharide serotypes were confirmed by PCR and WGS for five of these isolates (CC1/cps V (n=2), CC12/cps II (n=1), CC19/cps III (n=1) and CC19/cps V (n=1)).

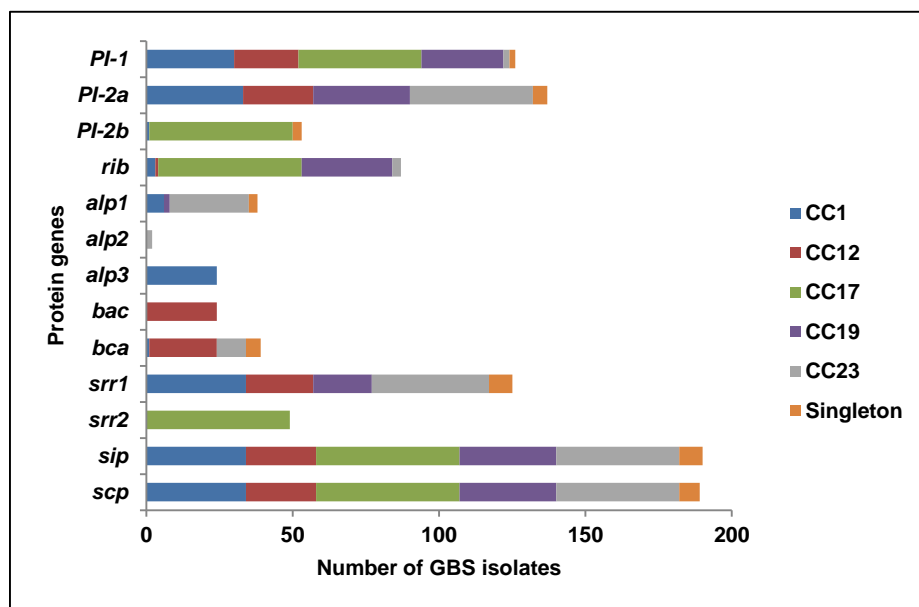


Figure 3. Distribution of genes encoding pilus islands (PI-1, PI-2a and PI-2b), the alpha-like protein (Alp) family (*rib*, *alp1* – 3, the alpha [*bca*] and beta [*bac*] components of the C protein), serine rich proteins (*srr1* and *srr2*), streptococcal surface protein (*sip*) and C5a peptidase (*scp*).

Table 1. Distribution of GBS serotypes and, antimicrobial resistance phenotypes and associated genes by clonal complex and disease

Clonal complex	Disease	Serotype	Tetracycline		Macrolide and lincosamides		Other resistance genes
			Phenotype	Resistance genes	Phenotype	Resistance genes	
CC1 (34)	EOD (10), LOD (2), Maternal bacteraemia (9), Non-invasive (13)	Ib (4), II (3), IV (4), V (19), VI (1), VII (1), NT (2)	R (31), S (3)	<i>tetM</i> (30), <i>tetO+tetW</i> (1)	cMLSB (13), iMLSB (4), S (17)	<i>ermB</i> (12), <i>ermTR</i> (5)	<i>ant(6)+aph(3')-III</i> (1)
CC12 (24)	EOD (2), LOD (1), Maternal bacteraemia (8), Perinatal (1), Non-invasive (12)	Ib (13), II (7), V (3), NT (1)	R (16), S (8)	<i>tetM</i> (10), <i>tetM+tetO</i> (1), <i>tetO</i> (5)	cMLSB (2), iMLSB (2), S (20)	<i>ermB</i> (2), <i>ermTR</i> (2)	<i>ant(6)+aph(3')-III</i> (1)
CC17 (49)	EOD (14), LOD (6), Maternal bacteraemia (13), Non-invasive (16)	III (48), IV (1)	R (46), S (3)	<i>tetM</i> (37), <i>tetM+tetL</i> (2), <i>tetM+tetO</i> (2), <i>tetO</i> (5)	cMLSB (7), iMLSB (2), S (40)	<i>ermB</i> (5), <i>ermB+mef+msrD</i> (2), <i>ermTR</i> (2)	<i>ant(6)+aph(3')-III</i> (6)
CC19 (33)	EOD (2), LOD (2), Maternal bacteraemia (10), Perinatal (1), Non-invasive (18)	II (8), III (18), V (5), NT (2)	R (32), S (1)	<i>tetM</i> (21), <i>tetM+tetO</i> (2), <i>tetO</i> (9)	cMLSB (7), iMLSB (1), M (2), L (2), S (21)	<i>ermB</i> (5), <i>ermB+isaC</i> (1), <i>IsaC</i> (1), <i>mef+msrD</i> (2), <i>ermTR</i> (1), <i>lnuB+IsaE</i> (1), no gene (1) ^a	<i>aph(3')-III</i> (1), <i>cat</i> (1)
CC23(42)	EOD (5), LOD (3), Maternal bacteraemia (18), Perinatal (2), Non-invasive (14)	Ia (31), III (3), V (6), NT (2)	R (38), S (4)	<i>tetM</i> (37), <i>tetM+tetO</i> (1)	cMLSB (1), M (6), S (35)	<i>ermB+lnuB+IsaE</i> (1), <i>mef msrD</i> (6)	<i>ant(6)+ aph(3')-III</i> (1)
Singleton (8)	EOD (1), LOD (2), Maternal bacteraemia (2), Non-invasive (3)	Ia (3), II (3), IX (2)	R (5), S (3)	<i>tetM</i> (5)	cMLSB (1) iMLSB (1), S(6)	<i>lnuB+IsaE+mef+msrD</i> (1), <i>ermTR</i> (1)	<i>ant(6)</i> (1)

EOD early onset disease; LOD late onset disease, NT nontypeable; R resistant; S sensitive

^aNo gene was detected in one isolate exhibiting the cMLSB phenotype

Table 2

Genetic differences in cgMLST loci detected in known and potentially linked cases

Cluster	Isolate	Strain	Disease category	Sequence type	Type	Locus	Name	Classification
Cluster 1	4910	Infant 1	EOD	ST-17	SNP	SAG0067	30S Ribosomal protein	Missense (L33P)
					SNP	SAG1061	Oxidoreductase, FMN-binding	Missense (G164R)
					SNP	SAG1587	Class I aminotransferase	Missense (N384S)
	4911	Mother1	Non-invasive	ST-17				
Cluster 2	4887	Infant2	stillbirth	ST-23	None			
	4882	Mother2	Invasive	ST-23				
Cluster 3	4788	Infant3	EOD	ST-110	None			
	4790	Mother3	Non-invasive	ST-110				
Cluster 4	4756	Infant4	EOD	ST-1	None			
	4757	Mother4	invasive	ST-1				
Cluster 5	4736	Infant5	EOD	ST-525	None			
	4735	Mother5	invasive	ST-525				
Cluster 6	4765	Infant 6	EOD	ST-600	None			
	4766	Mother6	invasive	ST-600				
Cluster 7	4851	Infant7	Non-invasive	ST-1163	None			
	4852	Infant8	Non-invasive	ST-1163	None			
Cluster 8	4739	Infant9	LOD	ST-8	Indel	SAG0098	Molecular chaperone DNAJ	G84_F85insGFDGGGFGG
					SNP	SAG0142	Hypothetical protein	No change
					SNP	SAG0278	Transketolase	No change
					SNP	SAG1647	DhaKLM operon coactivator DhaQ	Missense (A325S)
	4740	Infant10	Non-invasive	ST-8				
Cluster9	4874	Mother7	Invasive	ST-17	None			
	4880**	Mother7b	Non-invasive	ST-17				
Cluster 10	4890	Mother8	Invasive	ST-1169	None			
	4900**	Mother8b	Non-invasive	ST-1169				

**No metadata was available for the non-invasive isolate and the relationship to the invasive isolate could not be confirmed

Table 3**Genetic alterations associated with non-typeable group b streptococci**

Isolate	Infection type	Source	Genotype	Sequence type	Genetic alterations in <i>cps</i> and <i>neu</i> genes
4745	Non-invasive	Placenta	V	ST-19	SNP in <i>cpsE</i> resulting in CpsE missence mutation (I283R)
4801	Non-invasive	Placenta	V	ST-1	Single nucleotide deletion in <i>cpsN</i> resulting in frameshift and premature stop codon.
4806	Non-invasive	Breast milk	V	ST-1	SNP in <i>cpsD</i> resulting in premature stop codon Single nucleotide deletion in <i>cpsG</i> resulting in frameshift and premature stop codon
4812	Perinatal	Swab	II	ST-12	Single nucleotide deletion in <i>cpsE</i> resulting in frameshift and premature stop codon
4836	Non-invasive	High vaginal swab	III	ST-27	No alterations detected in <i>cps</i> locus
4850	Non-invasive	Endotracheal secretion	NT	ST-1163	Loss of <i>cps</i> genes likely due to recombination between direct repeats in <i>neuB</i> and <i>cpsA</i>
4851	Non-invasive	Endotracheal secretion	NT	ST-1163	Loss of <i>cps</i> genes likely due to recombination between direct repeats in <i>neuB</i> and <i>cpsA</i>

Corresponding loci assigned in PubMLST *cpsE* , SAG1171; *cpsN* (V), SAG1166; *cpsD*, SAG1172; *neuB*, SAG1161; *cpsA*, SAG1175