Perinatal *Streptococcus agalactiae* epidemiology and surveillance targets

Lucy L. Furfaro*, Barbara J. Chang, Matthew S. Payne

1 School of Medicine, Division of Obstetrics and Gynaecology, The University of Western Australia, Australia.

2 School of Biomedical Sciences, The Marshall Centre for Infectious Diseases Research and Training, The University of Western Australia, Australia.

*Corresponding author

Email: lucy.furfaro@uwa.edu.au

Keywords: Group B Streptococcus, *Streptococcus agalactiae*, epidemiology, serotype, sequence type, clonal complex, surveillance, vaccine.
Table of Contents

Title page
Summary

Streptococcus agalactiae
Identification and classification
Serotyping

Surface Proteins

Pulsed Field Gel Electrophoresis

Multi-locus Sequence Typing
Random Amplified Polymorphic DNA and Repetitive-sequence-based Polymerase Chain Reaction

Multi-locus Variant-repeat Assay
Clustered Regularly Interspaced Short Palindromic Repeats Locus 1

From Targets to Treatment
Conclusion

Acknowledgments
References

Author biographies
**SUMMARY**

*Streptococcus agalactiae* or Group B Streptococcus (GBS) is a major neonatal pathogen. Recent data has elucidated the global prevalence of maternal and neonatal colonisation, however, gaps still remain when considering the epidemiology of this species. A number of phenotypic and genotypic classifications can be used to identify the diversity of GBS and some are more discriminatory than others. This review explores the main schemes used for GBS epidemiology and further details the targets for epidemiological surveillance. Current screening practices across the world provide a unique opportunity to gain detailed information on maternal colonising strains and neonatal disease-causing strains, information which is vital for monitoring and therapeutics, if sufficient detail can be extracted. Deciphering which isolates are circulating within specific populations and recording targets within invasive strains are crucial steps for monitoring the implementation of therapeutics such as vaccines, as well as developing novel therapies against prevalent GBS strains. Having a detailed understanding of global GBS epidemiology will prove invaluable for understanding the pathogenesis of this organism and equipping future prevention strategies for success.

**STREPTOCOCCUS AGALACTIAE**

*Streptococcus agalactiae* is the sole member of the Lancefield group, Group B Streptococcus and therefore is commonly referred to as GBS (1). This Gram positive coccus resides as a commensal of the gastrointestinal and genitourinary tract of humans; however, it has the capacity to cause serious infections. These are often opportunistic in nature, affecting the
elderly, immunocompromised, and particularly neonates, in which GBS is a major cause of morbidity and mortality (2, 3). Recently, infections in healthy adults have also been reported (4, 5). In these populations the spectrum of disease ranges from sepsis, pneumonia and meningitis to endocarditis. GBS disease can be classified based on the time of onset: early-onset disease (EOD) occurs in the first week of life (6), while any disease presentation after this up to 3 months of life is classified as late-onset disease (LOD) (7). In neonates, the majority of early-onset GBS disease occurs as a result of transmission from a colonised mother either through ascending infection or during the process of vaginal birth (8). In contrast, late-onset sepsis is believed to be contracted via nosocomial or community-acquired means (9). Due to the potential for transmission from the mother, universal screening of pregnant women by either risk-based or culture based-methods has been implemented in most countries. In such countries, intrapartum antibiotic prophylaxis (IAP) is administered before delivery to women at risk of GBS disease (delivery at <37 weeks’ gestation, intrapartum temperature ≥38.0°C, or rupture of membranes for ≥18 hours) or with GBS colonisation (10). This widespread treatment strategy has been successful in reducing EOD significantly, however, has had no effect on LOD (10). Screening for GBS by culture currently involves vaginal and rectal swabs collected at 35-37 weeks gestation and cultured in selective enrichment media (11). Additionally, diagnostic molecular techniques to detect GBS have also been described (12). The outcome of this screening is either GBS colonisation or absence. However, there is a plethora of other information about these organisms that could be obtained during routine screening that may prove vital in furthering our understanding of their epidemiology; as such we are currently under-utilising this potentially valuable surveillance source. Expanding our knowledge of current targets
through the use of phenotypic and molecular assays to identify and differentiate isolates within the population will likely prove essential for future GBS interventions.

IDENTIFICATION AND CLASSIFICATION

Several methods have been described to identify and characterise GBS for epidemiological and diagnostic purposes (13). These include: serotyping (based on the capsular polysaccharide); surface proteins; multi-locus sequence typing; multiple-locus variable repeat assays; and more recently, clustered, regularly-interspaced, short palindromic repeat locus assays. All have advantages and disadvantages, and all convey different levels of epidemiologic data in the context of surveillance.

Serotyping

The capsular polysaccharide (CPS) that encapsulates GBS is a major virulence factor, but additionally determines the serotype. A total of ten serotypes have been identified, with CPS IX the most recently proposed (14). Serotypes can be identified through latex agglutination and molecular assays such as multiplex PCR (13, 15-17). As one of the first discriminating factors of GBS, the serotypes have been described in the literature as a way of grouping isolates in an attempt to understand prevalence in maternal colonisation and neonatal disease. Serotyping provides valuable information that allows classification into serologically similar groups, and patterns have emerged based on global prevalence. The most common global serotypes include CPS Ia, Ib, II, III and V. When considering a perinatal context, however, many studies describe limited serotype testing (18-20) and others were
completed prior to inclusion of CPS IX (21-23). These are major limitations in such studies, however, there are still numerous studies that have included all known serotypes and these generally support the overarching predominance of the five main CPS types in a perinatal context (24, 25).

Global serotype distribution

A recent meta-analysis of maternal colonisation described a global dataset of serotype prevalence (24). Russell and colleagues highlighted the regional variation for the less common serotypes such as CPS IV, VI, VII, VIII and IX. Prior to this analysis the general consensus was that approximately 10 – 30 % of pregnant women were colonised with GBS; this analysis supports previous estimates, and suggests that the global colonisation rate in pregnant women is 18%, with regional variation estimates between 11 – 35%. South and East Asian countries had the lowest reported prevalence (11-12.5%), while the Caribbean was observed to have the greatest prevalence (34.7%) (24). Subsequent neonatal disease rates of GBS have been estimated at 0.49 per 1000 live births globally, with highest incidence in Africa (1.12 per 1000 live births) and lowest in Asia (0.30 per 1000 live births) (25). Five serotypes (Ia, Ib, II, III and V) have been estimated to encompass approximately 98% and 97% of the serotypes identified during maternal colonisation and neonatal disease, respectively. This leaves a small proportion of colonisation/neonatal disease attributable to the remaining five serotypes, however, it is important to highlight these serotypes as the dynamic has the potential to shift. Several studies highlight the variation in serotype prevalence globally and are summarised with general trends below for each serotype (Table 1).
The frequency of different serotypes as colonisers makes determination of disease development an extremely difficult task. The asymptomatic colonisation in some, yet invasive disease progression in others emphasises the complexity of the host-GBS interaction. Maternal colonisation by this serotype is seen globally (24) and it is the most prevalent serotype in maternal disease according to a limited meta-analysis (including only USA, UK and France) (26). Regarding neonatal disease, previous meta-analysis found Ia as the most frequent serotype contributing to EOD, however, this global estimate was lacking in data from Asia (27). The latest analysis includes Asia and while overall trends of prevalence remain similar, South America is the only region in which CPS Ia predominates over CPS III in cases of EOD (25). In contrast, in Eastern Asia unlike other regions, prevalence of Ib surpassed Ia (25).

Eastern Asia has the highest proportion of neonatal disease attributable to CPS Ib, and within the Asian region, Ib is second only to CPS III in predominance (25). In Taiwan, Lin and colleagues found CPS Ib to account for 21.6% of the isolates collected from a mixed population of pregnant women, neonates and non-pregnant adults, representing the most common serotype in the study (28). This frequency, however, was largely accounted for by non-pregnant adult invasive disease isolates which represented 26/29 CPS Ib isolates. Similarly, in South Korea, 22% of GBS serotypes causing infection in adults were identified as
Differences in distribution have been observed within different cohorts, such as pregnant, neonatal and adult populations. Maternal colonisation with CPS Ib in Mexico and South America has been reported as considerably higher compared to other countries (24). On the opposite end of the spectrum, CPS Ib was the least prevalent, after non-typeable (NT) isolates, to account for maternal disease (USA, UK and France) (26).

Serotype II is the fourth most prevalent serotype in maternal disease according to a study including USA, UK and France (26) and the fifth most prevalent in neonatal disease worldwide (25). Predominance of this serotype was observed in pregnant women at the Thai-Myanmar border, where 24% of isolates recovered were CPS II as reported by Turner and others (30). A study in Catalonia, Spain assessed GBS isolates from pregnant women and pathogenic strains isolated from neonates with sepsis and found serotypes Ia and II were significantly more frequent as colonising strains (31). Colonisation of Nigerian pregnant women also showed predominance by serotype Ia, while serotype II accounted for 71.4% of the neonatal colonising isolates (32). Likewise, GBS isolated from the nasopharynx of Gambian neonatal carriers found highest prevalence of CPS V and II (33).

Well-known for its association with disease, in particular EOD, CPS III has been reported thoroughly worldwide and remains the leading cause of neonatal disease (25). Maternal prevalence differs by continent, as Africa, Australia and New Zealand have much higher prevalence, while the Middle East, Eastern Europe, South Asia (including India and
Bangladesh) and Central America have lower rates (24). In coastal Kenya, CPS III accounted for the majority of maternal colonising isolates and 70% of the neonatal disease strains (34).

In France, serotype III accounted for 83.9% of infant meningitis caused by GBS infection (35).

Temporal dynamics of this serotype were reported for adult invasive infections in Iceland, in which the only significant decline in serotype prevalence over the 39 year period was CPS III (36). This observation contrasts with other populations, such as neonates, in which it is attributed to the majority of invasive disease (25).


IV

Serotype IV is less prevalent globally compared to the other main serotypes (Ia, Ib, II, III and V). Recent studies, however, have noticed the emergence of this serotype, particularly in Canada and USA (37-39). These studies also reported evidence of capsular switching relevant to this serotype, a concern for targeted therapies. Serotype IV presence within a predominantly CPS III cluster of isolates from France was suggested to have occurred through capsular switching (40) and capsular switching to serotype IV within clonal complex 17 was also reported in Kenya (34) which further supports this as a cause for concern.

V

Early studies reporting the emergence of CPS V noted a close genetic relationship among the strains of this serotype and its recovery from all populations, however, more so from non-pregnant adults (41). It is recognised as the third most prevalent serotype associated with maternal disease in USA, UK and France (26). Africa, more specifically Western Africa,
has reported higher frequency of CPS V maternal colonisation than other regions such as
North America, Australia and New Zealand, which have a much lower prevalence (24). In
Gabon, approximately one third of serotypes isolated from pregnant women were CPS V
(42). This supports previous observations of CPS V predominance by other studies in
Gambian mothers and their infants, and pregnant women in Egypt (43, 44).

**VI**

Considered one of the rarer serotypes (amongst VII, VIII and IX), CPS VI is often reported in
Asian countries, in particular Japan (45) and Malaysia (18, 46). While global colonisation
estimates place CPS VI in the minority, this serotype is thought to account for as much as
20% of serotypes observed in South-East Asian pregnant women (24).

**VII**

Similarly, CPS VII is less represented across the globe (24). Interestingly, it is found to be the
most common serotype in Southern Ghana, the first study to report such serotype
prevalence in pregnant women from West Africa (47). A number of Malaysian studies
observed serotype VII in their cohorts ranging from 1.7% to 21.4% of serotypes (18, 46, 48).
This supports the observation of intra-country variation that was similarly reported
(although not for CPS VII) for serotypes in different regions of Brazil (19). A study of
maternal colonisation in India observed CPS VII accounting for 6.7% of the serotypes and
greater than Ib prevalence, which is indicative of Asian countries, yet unique in the
distribution (49). More recently, Islam et al. observed 37.1% of the GBS specimens isolated
from newborn umbilical cord areas (colonising isolates) belonged to CPS VII in Bangladesh (50).

Serotype VIII is strongly associated with Japan, with an early study reporting high frequency in pregnant women (45). They found 35.6% prevalence of CPS VIII among serotypes, followed by CPS VI (also considered rare globally) among colonising isolates of Japanese pregnant women (45). Thereafter, incidences of its presence began to be described outside of Japan. In Denmark, seven instances of invasive disease among non-pregnant adults belonging to CPS VIII were reported from 1999 – 2002 (51). Also, detection of one isolate colonising a pregnant woman was recognised as CPS VIII in Maryland, USA (52). Lately, frequency of CPS VIII is generally low outside of Japan and often studies describe minimal isolates (53, 54).

The newest serotype described in 2007 has had fewer chances for description in the literature, however in the last decade studies have adopted new techniques and now most test for all serotypes. As mentioned above, a study in Southern Ghana observed CPS IX with the second highest prevalence behind VII (47). A Swedish study assessed changes in serotype distribution over time for invasive disease (neonatal and adult populations) and suggested circulating serotypes remained stable from 1990 onwards, however, noted the increase in CPS VII – IX (55). Additionally, a case study in Japan described an incidence of
ultra-late onset sepsis caused by CPS IX, and this demonstrates its capacity to play a role in
disease (56).

Non-typeable

Numerous serotyping studies result in isolates that do not represent any of the defined
types and are thus considered non-typeable (NT). It is not known whether this is a result of
technique specificity or whether such isolates may represent novel serotypes; either way NT
isolates could potentially skew the available data. Capsule loss or lack of expression are both
potential contributors to this NT category (57, 58). One study assessed a collection of GBS
isolates classified as NT by serological methods and went on to characterise them by a
number of different techniques (immunodiffusion, pulsed-field gel electrophoresis and
multi-locus sequence typing) (59). This resulted in all isolates being serotyped and found to
represent all types excluding IV and VII–IX, with instances of multiple serotypes per isolate,
which has been described previously (52). This may represent an incidence of capsular
switching, a phenomenon that involves the genetic exchange of capsular genes which
enables expression of alternative capsule types and it is proposed that such horizontal
transfer may be responsible for the capsular locus diversity (60-62). In the former study by
Ramaswamy and colleagues, 80% of these isolates containing multiple serotypes contained
a type V CPS-specific gene and the authors proposed that this serotype may have an
association with increased competency and ability to acquire DNA (59). This is an interesting
notion, but given that GBS is not thought to be naturally transformable, the idea that
capsular switching may result from increased competence of CPS V seems unlikely,
however, capsular switching has been reported by others (37, 40, 63, 64). This is a concern
for vaccine development targeting the capsule as this has been an issue for the naturally transformable *Streptococcus pneumoniae* due to capsular switching (65, 66).

One of the major limiting factors when discussing GBS serotype distribution is that a direct comparison between studies can be difficult due to the number of different study parameters assessed such as specimen type, study population and the added complexity of various serotyping methodologies. For optimal surveillance outcomes, all types of patient cohorts are vital to completing the broader picture of GBS epidemiology. Increasing our understanding of prevalent isolates associated with a specific body site and their potential to cause disease by monitoring the serotypes is the first step towards narrowing down invasive strain targets and surveying the serotype distribution dynamics.

Surface Proteins

In addition to the CPS, proteins are also present on the surface of GBS cells and studies have shown that antibodies against both CPS and surface proteins are protective against GBS infection (67-69). The alpha-like protein (Alp) family are among the well-characterised GBS surface proteins and are significant virulence factors. These include alpha-C protein, Rib, Alp2, Alp3, Alp4 and epsilon (Alp1), and are encoded by the *bca*, *rib*, *alp2*, *alp3*, *alp4* and *epsilon/alp1* genes respectively (70). Detection of the expressed surface proteins by serology or gene presence by molecular techniques such as PCR or sequencing can be used as another grouping tool. The C antigen was the first identified surface protein antigen and has been found in numerous GBS strains except for serotype III (71). Alps have been present in the majority of GBS isolates, however, non-Alp isolates do exist. A total of 1.1% of isolates
in one study were non-Alp and all were associated with invasive bloodstream infections (72). Smith and others examined associations between the common serotypes (Ia, Ib, II, III and V) and specific genes. They found that most genes tested (sbp1, bca, bac, rib, brp, pag, and psp) were common in only one or two serotypes, except for bca which was present in >50% of all Ib, II and V serotypes (73). A review of several surface proteins reported correlations between alpha protein presence in CPS Ia, Ib and II, yet almost never in III, and only rarely in V; Rib is expressed by the large majority of CPS III, many II and by a few V; Alp3 is expressed by CPS V and VIII and Alp2, more rarely, by CPS Ia, III and V (70).

Pulsed Field Gel Electrophoresis

Pulsed-field gel electrophoresis (PFGE) is another molecular method to distinguish isolates with greater discrimination than serotyping for evaluation of genetic relatedness. It involves nucleic acid digestion with one or more restriction enzymes (often Smal in the instance of GBS) to produce a fragment profile on an agarose gel, after which the profiles are then compared for similarity and grouped accordingly (74). Fasola and colleagues described this method for use in GBS in 1993 and later Benson et al. reported an updated and more rapid PFGE technique for GBS, reducing workflow from 6-8 days down to 3 days total (75, 76). PFGE profile correlations with serotype have shown variation; Skjaervold et al. observed homogeneity for Ib, III and V strains while IV, Ia and II were heterogenous (77). This suggests that serotype grouping does not permit an understanding of close genetic relationships. While greater depth of genetic relation could be observed using PFGE compared to serological typing, there are limitations with this method. Due to the lack of a uniform typing system with respect to naming, PFGE analyses can be difficult to compare between
studies: what may be PFGE clone A in one study could be 37 in another (78, 79). Therefore pattern and band sizes must be used for comparison between studies. In this respect, it is mainly useful for comparison within a study to see diversity among isolates and modes of transmission. For example, mother and infant paired GBS specimen fragment profiles were observed with common patterns, highlighting vertical transmission (80). In the past, identifying fragment banding on gels was subjective and lacked accuracy in reporting band sizes, making comparison across studies difficult. The idea of using restriction enzymes with minimal cutting sites aims to reduce this by producing fewer fragments for visualisation on the gel (74). Regardless of any limitations, PFGE has identified variation within and across serotypes and has led to the progression of the molecular epidemiology of GBS.

Multi-locus Sequence Typing

Jones and colleagues developed the multi-locus sequence typing (MLST) method for GBS using seven defined housekeeping genes \((adhP, atr, glcK, glnA, pheS, sdhA\text{ and } tkt)\) (81). The MLST database was produced by Jolley and colleagues and currently includes over 4,000 GBS isolates (82). Allelic variations amongst housekeeping genes allow the determination of different sequence types (STs). These STs can then be clustered into what are known as clonal complexes (CCs) which are conservatively defined by members having no more than one allelic difference from other members of the cluster.

The MLST database is a valuable resource that enables sharing of data; the publicly available information for GBS includes a range of criteria to assess, including the country of origin, serotype and specimen type. Although there remains a reporting bias, in that only
submitted information is available, the data represent a mix of global information which is very useful. An audit of the GBS MLST database (http://sagalactiae.mlst.net) found a total of 4,131 isolates from 38 different countries. Of these, 3,949 isolates had been assigned an ST (n = 635). Twelve STs account for 73% of the GBS isolates with available MLST data (Figure 1). The remaining STs (27%) each account for less than one percent of all isolates and 525 STs have a single isolate submitted to the database.

Having access to the location of each isolate’s origin enabled categorisation of STs by continent (Figure 2). This highlighted gaps in the global data, with no isolates from Australia and New Zealand available, unless not stated (categorised as “Unknown”). Examining the main contributors within each continent, Netherlands (n = 1230/1610), Kenya (n = 1133/1172), Japan (n = 490/550), USA (n = 325/491) and Brazil (n = 3/5) have submitted the most isolates within Europe, Africa, Asia, North America and South America, respectively.

Associations between ST and serotype have been reported in the literature with some showing strong correlation and others very little; this is observed in the MLST database also (Figure 3). Ramaswamy et al. observed a correlation between CPS III and ST-17, Ib and ST-12, and found CPS V was represented in all STs except for ST-17 (59). Similar trends are observed within the MLST database where each of the most prevalent STs appears to be dominated by a particular serotype (Figure 3). The serotype variability in some instances, for example ST-1, may be suggestive of discordance with serotype profiling. Trends observed by various studies reveal variable serotype presence within ST-1, variation between CPS Ib and II within ST-10, predominance of CPS III within ST-17 and ST-19 and CPS Ia within ST-23 (83, 84). Collectively, within the clonal complexes, homogeneity of serotypes is evident in the CC-17 group with all representing CPS III except for a small proportion which are
unspecified. Similarly, within CC-19, ST-182 includes CPS III only, however, other STs in this complex show variation, especially ST-28 which is predominantly CPS V (Figure 3). From an evolutionary perspective, caution must be taken when inferring phylogenetic relationships as suggested by Sorensen et al., due to the indication that different segments of the GBS genome evolved independently (85). With this in mind, MLST is perhaps not the best method for describing evolutionary relationships, however, for the purpose of epidemiology at the current scale, it should be acceptable now that we are aware of these factors.

Reoccurring themes in the literature include the numerous studies that have described the association between ST-17 and invasive neonatal disease (83, 86-90), and further, the relationship between this lineage and meningitis (91). Neonatal invasive isolates collected over a ten year period in Sweden showed variation in genotype prevalence over time and in different regions, in particular, the emergence of CC-1 and subsequent disappearance of CC-23 (88). These temporal changes emphasise the importance of understanding GBS epidemiology and tracking changes in prevalence worldwide; ST-17 may expand and/or other STs may become hypervirulent over time.

The MLST profiling scheme is publicly available and curated, making it an invaluable epidemiology resource. If the database was treated like a repository, the epidemiological surveillance would be greatly enhanced which would be desirable for targeted therapeutics assessment.

**Random Amplified Polymorphic DNA and Repetitive-sequence-based Polymerase Chain Reaction**
Random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) and repetitive-sequence-based PCR methods (rep-PCR) are techniques that rely on the amplification of DNA via random primers (RAPD) (92) or repetitive sequence target primers (rep-PCR) and result in a genomic fingerprinting profile as the generated amplicons lead to strain-specific band patterns on an electrophoretic gel.

The RAPD assay has been used in a number of genotyping contexts such as; evolution and genetic relatedness (93), epidemiology (46, 94), transmission (95) and isolate variation within individual patients (96-98). One study examined RAPD profiles of GBS isolated from a number of sites such as blood, cerebral spinal fluid and maternal breastmilk, in cases of LOD in neonates (n = 4) (96). They found identical banding profiles for all sample types from individual patients and differing types between patients, however, this was representative of a small sample size. Likewise, Brzychczy-Wolch et al. used RAPD to suggest that GBS isolates from colonised neonates originated from the colonised mother regardless of mode of delivery as a result of identical genotypic profiles (99). While detection of identical clones via RAPD profiles can suggest vertical transmission, other studies have used this to demonstrate horizontal transmission in neonates born to non-colonised mothers that likely acquired GBS from the hospital environment with identical profiles observed in other hospital wards (100).

Toresani et al. found 16 profiles from 21 GBS isolates tested by RAPD assay, furthermore, successive samples of two patients showed different genotype profiles, suggesting reinfection (97). Contrary to this, El Aila et al. compared culture methods and genotyped the multiple isolates collected from individual patients. RAPD testing of 118 isolates from 32 pregnant women resulted in 66 genotype profiles, including patients with multiple different
genotypes at each site (vagina and rectum), demonstrating assumed reinfection may be due
to lack of initial identification (98). This highlights the use of RAPD for single centre studies
and integration with high resolution capillary electrophoresis, which enhances accuracy and
provides digital fingerprints (98).

Numerous studies have observed correlation between RAPD profile and clustering of
resistant phenotypes (101, 102). The results from a study of macrolide resistant GBS isolates
suggested that PFGE was a more precise tool for analysing molecular epidemiology than
RAPD PCR (95). In contrast, Zhang and colleagues (103) found that the clustering of the
isolates by RAPD-PCR was in concordance with previous PFGE findings that suggested
clusters could be serotype specific (80).

The great advantage of RAPD is, however, a relatively short procedure time and simple
methodology making it a suitable tool for suggesting clonal relatedness in clinical settings
and monitoring of transmission. Comparison between studies can be difficult, however,
particularly when different primers are used. The random nature of this profiling technique
provides variation and enables discrimination between clones, however, similar to the PFGE
technique, due to the lack of specific targets the information this provides beyond strain
similarity is somewhat limited unless further genetic analysis is undertaken.

Literature assessing rep-PCR in GBS are limited, with the one study comparing rep-PCR
Diversilab® system with MLST and PFGE using clinical adult and neonatal GBS isolates (n =
179) (104). Compared to PFGE, the clustering concordance was similarly low between the
two techniques, however, PFGE had greater discriminatory power than rep-PCR (104). One
study compared both RAPD and rep-PCR techniques to genotype GBS isolates from fish and
observed RAPD to have a greater level of variability than rep-PCR between strains, with 13
and 9 profiles, respectively (105). This is not supportive of the use of rep-PCR over other methods tested including MLST and PFGE. The rep-PCR typing is notorious for its susceptibility to minor variations in experimental conditions and reagents, resulting in poor reproducibility (106, 107).  

**Multi-locus Variant-repeat Assay**  

Various bacterial species have been typed through the detection of variable numbers of tandem repeats (VNTR), which is based on the repeated sequences that make up DNA at different loci and that are found throughout bacterial genomes (108). As the name suggests, these VNTR loci vary in repeat number, allowing for an assay of a combination of several VNTR loci to generate different strain-specific profiles. This is known as a multi-locus variant-repeat assay (MLVA) and has a similar principle to MLST, however, rather than housekeeping genes, MLVA uses these VNTR loci (109). This typing scheme was developed for GBS by Radtke and colleagues and includes five of the most diverse loci (SATR1 (110), SATR2, SATR3 (111), SATR4 (112) and SATR5 (113)) (109). The authors found clustering of CPS/surface protein typing or MLST to generally correspond with the MLVA profiles, however, the latter was more discriminatory. A total of 126 GBS isolates tested were profiled into 70 different MLVA types as compared to 36 STs by MLST and 19 CPS/protein types. A specific number of 15 repeats within SATR2 was observed to correspond with serotype IX. Additionally, for one of the three described SATR3 alleles, all isolates of that allele (except one) belonged to the MLST ST-17, known to account for a high proportion of neonatal infections (81, 88). Other MLST groups were more dispersed, including CPS V/ST-1 for which SATR5 resolved into a number of profiles. This is interesting due to the often
homogeneous reports of this sub-group in other studies (28, 77). The SATR1 locus is considered to be composed of clustered regularly interspaced short palindromic repeats (CRISPR) and this worked efficiently as part of the MLVA scheme.

Additional MLVA schemes have been reported, one using six VNTR loci including three described by Radtke et al. (109, 114). Similar results were observed in this study, with greater isolate discrimination reported with the MLVA (98 types) technique compared to MLST (51 types). The clustering patterns observed for MLVA types showed similarity with MLST CCs, in particular, all human strains of MLST CC-17 appeared in MLVA cluster 9 (114). This scheme was able to categorise MLST CC-23 into two clusters, one accounting for CPS III and the other CPS Ia and highlights this scheme’s ability to distinguish strain variability within lineages (114).

A study of commensal GBS colonisation of non-pregnant women of reproductive age utilised MLVA for genotyping and observed 15 types among the 86 GBS isolates (115). In pregnant women, 30 MLVA types from a total of 41 GBS isolates were observed which suggested a highly diverse population structure (116). The MLVA types (MTs) are difficult to compare between studies though, as there is no mention of a database where naming conventions are uniformly enforced; however, a database has been created for this purpose (www.mlva.net) but unfortunately it does not include GBS.

**Clustered Regularly Interspaced Short Palindromic Repeats Locus 1**

Lastly, the molecular target which was mentioned in the MLVA scheme as one of the loci are the clustered regularly interspaced short palindromic repeats (CRISPR) and Cas (CRISPR-
associated proteins). These are known for their role in bacterial host defence and adaptive immunity against foreign invading DNA, such as bacteriophages and mobile genetic elements (117). The genetic arrangement that this system follows results in a polarised orientation, which allows tracking of ancestral invasions over time (118). In GBS, two CRISPR-Cas systems have been identified and these include a type II-A system (ubiquitous and functional) and a type I-C system (rare and often incomplete), associated with the CRISPR1 and CRISPR2 loci, respectively (119, 120). The ubiquitous nature of the CRISPR1 locus makes it an attractive molecular marker to track changes in the genome. The highly conserved direct repeats form the spacers between introduced genetic materials and can therefore be monitored for changes. Using this method, Beaureille and others were able to follow the vaginal GBS colonisation of 100 women over an 11-year period (121). This extensive study highlights the influence of genetic variables that alter these organisms over time, for example bacteriophage infection. This is an example of new epidemiological markers that may assist in understanding these bacteria, their colonisation and their pathogenic tendencies.

**FROM TARGETS TO TREATMENT**

The purpose of the targets and molecular assays presented above is to understand the population of GBS with respect to colonisation and disease; by distinguishing differences between strains we can begin to associate specific features with clinical outcomes. Equally, by identifying similarities that are represented across the majority or the entire population of GBS, we can target these bacteria more broadly. Within antenatal settings the current preventative measures for GBS include antenatal GBS screening (risk- or culture-based) and
subsequent intrapartum antibiotic prophylaxis prior to delivery. This has resulted in reductions in EOD rates as reported by the Centre for Disease Control and Prevention and is recommended by other countries also (10, 11, 122, 123). Unfortunately, no change in LOD has been observed with such preventative strategies, however, this is likely due to the nature of transmission (nosocomial or community acquired) (124).

Due to the recommended intrapartum antibiotic prophylaxis, antibiotic exposure has become widespread: the potential for resulting adverse events is unclear as highlighted in a recent systematic review (125). While consistency is lacking for development of antimicrobial resistance and long-term adverse events as a result of IAP for GBS, infant microbiome disruption has been observed by numerous studies (126-132). However, due to the lack of high-level (species) taxonomic data and the myriad of other methodological problems that are still rampant in microbiome studies (lack of negative controls and proper selection of primer sets) (133), with the current data it remains unclear as to what impact this may have on neonates. With potential implications in mind, however, it is important to examine alternative preventative and therapeutic strategies.

Maternal GBS vaccination has been proposed as an alternative preventative strategy due to the ability of maternal IgG to cross the placenta and provide immunity for the fetus (134). Baker identified the association between deficient maternal antibody levels and occurrence of neonatal GBS infection. This study also detected GBS antibody titres in the umbilical cord serum from three healthy neonates born to mothers with demonstrable antibody levels (60 – 80% of the maternal levels detected), suggesting transplacental transfer and illustrating the potential for a maternal vaccine (135). The benefit of maternal vaccination has since been described and a number of GBS vaccines have been developed (134, 136-138).
While no vaccines against GBS are currently licensed and available, several polysaccharide conjugate and surface protein vaccines are being trialled (137, 138). Through comparison of these vaccine targets with the global GBS epidemiology, we can identify areas of vaccine coverage. Numerous studies that have identified the less commonly recognised serotypes within their populations have raised concern for vaccinations targeting specific serotypes such as Ia, Ib and III — all of which are included in the trivalent vaccine currently furthest along in progress with completion of a phase Ib/2 clinical trial (NCT01193920) (139, 140). The review by Lin et al. describes such challenges faced by capsule-based conjugate vaccines against GBS, in addition to dealing with the potential for serotype replacement and capsular switching (138). More recently, a pentavalent vaccine targeting serotypes Ia, Ib, II, III and V has commenced a phase I trial on healthy volunteers (NCT03170609); if successful, this could be an effective GBS vaccine outside of Asia as it covers the five major circulating serotypes in the US, Europe and Australia (24). Alternative vaccination strategies have focused on proteins. Surface proteins have been included as vaccine targets in addition to capsule types, and given the variety of circulating serotypes described, these may provide a broader, more universal target to exploit which would be ideal for vaccine development (137). Although minimal surface proteins are conserved across all strains, Maione and colleagues provided evidence of protein vaccination in female mice with the ability to elicit protection in 43 – 80% of cases of GBS challenge to pups born to the vaccinated mother (141). Four proteins, including a conserved Sip protein (SAG0032) encoded in the core genome of strains were included in this study in addition to three other antigens (SAG1404, SAG0645 and SAG0649) which had variable presence in the genome (141). Studies in humans have been described by Heath et al. and include multiple studies of monovalent, one bivalent and several trivalent CPS-based vaccines, all in the proof-of-concept stage involving trials in pregnant women.
women. A bivalent vaccine including the N-terminal domains of surface proteins alpha-C and Rib is currently in development by MinervaX and has completed phase I clinical trials (NCT02459262). Larsson et al. reported the association between low protein alpha and Rib antibodies and occurrence of neonatal invasive infection (142). While the type of vaccine is important in terms of coverage, the implementation is equally as important when considering access, especially in low-income settings and developing countries where screening and widespread prophylaxis is often not practical (143). Based on the World Health Organisation guidelines, GBS vaccination, particularly in South Africa, would be very cost-effective according to cost-effectiveness analysis (144). In developed countries, such as the United States, Kim and others evaluated the cost-effectiveness of maternal vaccine compared to screening/IAP (145). They concluded that ≥70% vaccine efficacy in addition to IAP for unvaccinated women would prevent more disease than current strategies with similar economic burden (145). Epidemiological data about GBS is not only invaluable for vaccination development and success, but also for other future targeted therapeutic agents, such as bacteriophage therapy (146).

CONCLUSION

This review describes different commonly used and emerging techniques to define GBS epidemiology and phylogeny. Some traditional targets such as serotype are widely reported but lack the level of definition that other molecular targets can provide with respect to genetic relatedness and pathogenic features. The difficulty lies in recognising these targets and widely reporting them to provide a comparable basis for global surveillance. It is likely that future epidemiological efforts for monitoring GBS will include the reporting of a
combination of these targets, and in turn, refine the association between these different components. The epidemiology of GBS provides insight into the adaptation and evolution of these organisms through examination of their genetic relatedness and population structure. Understanding GBS as an opportunistic pathogen is one aspect to arise from the proposed surveillance efforts, another is for the guidance of future therapeutics. Vaccines based on serotypes commonly found in one area may not include prevalent serotypes of another due to the geographic variation observed, even at a regional level. To vet these coverage issues, surveillance of different targets within GBS are required to monitor prevalent isolates in all populations. Additionally, following vaccine implementation, this will provide a platform to monitor shifts in GBS dynamics and identify emerging pathogen types before outbreaks occur. Broader coverage may result from protein vaccine development and this further emphasises the importance of characterisation of a range of targets. The targets outlined here each have individual relevance, although uniformity in reporting profiles would be ideal for global comparisons. Currently MLST, while less discriminatory than MLVA, comprises the only established database to enable this comparison and provides the added confidence of curation.

ACKNOWLEDGMENTS

LLF is supported by the Australian Government Research Training Program Scholarship and the Professor Gordon King Postgraduate Scholarship provided by the Women’s and Infants Research Foundation. MSP is supported by an NHMRC Project Grant [1077931]. The authors state no conflicts of interest.
REFERENCES


agalactiae in Brazil: serotype distribution, virulence determinants and antimicrobial susceptibility. BMC Infect Dis 14:323.


46. Suhaimi MES, Desa MNM, Eskandarian N, Pillay SG, Ismail Z, Neela VK, Masri SN, Nordin SA. 2017. Characterization of a Group B Streptococcus infection based on the demographics, serotypes, antimicrobial susceptibility and genotypes of selected isolates from sterile and


Otaguiri ES, Morguette AE, Tavares ER, dos Santos PM, Morey AT, Cardoso JD, Perugini MR, Yamauchi LM, Yamada-Ogatta SF. 2013. Commensal Streptococcus agalactiae isolated
from patients seen at University Hospital of Londrina, Parana, Brazil: capsular types, genotyping, antimicrobial susceptibility and virulence determinants. BMC Microbiol 13:297.


Lucy L. Furfaro is currently completing her PhD at the University of Western Australia and previously completed her Bachelor of Science and Honours at the University of Western Australia also. Her research interests include perinatal microbiology with a focus on Group B Streptococcus and alternative treatment strategies to antibiotics such as bacteriophages. This area of research has been pursued over the last 4 years as part of Honours and postgraduate studies.

Professor Barbara J. Chang received her PhD from Monash University, Australia, followed by postdoctoral positions at the University of Calgary, Canada. Currently, she is the Director of Postgraduate programmes in Infectious Diseases in the School of Biomedical Sciences at the University of Western Australia. Her research interests include biology and genetics of bacteriophages, anti-quorum-sensing agents, and bacterial pathogens including Clostridium difficile, Moraxella catarrhalis, and Group B Streptococcus.

Dr Matthew S. Payne was awarded his PhD in molecular microbiology from the University of Queensland, Australia and has since conducted post-doctoral research at Kings College London, United Kingdom and La Trobe University, Australia. Dr Payne is currently at The University of Western Australia, where he is focused on preventing infection-mediated preterm birth. His research interests include the microbiology of preterm birth, microbial metagenomics, Ureaplasma and Mycoplasma spp., and antibiotic and non-antibiotic treatment regimens for prevention of preterm birth and neonatal sepsis.
**Figure 1.** Frequencies of the 12 most prevalent GBS sequence types available from the MLST database which includes a total of 635 sequence types (remaining STs account for <1% each).
Figure 2. Contributions by continent of GBS isolates available in the MLST database, as a percentage of all isolates (South America represents 0.1% and is therefore not visible).
Figure 3. Serotype distribution within each sequence type, using the public MLST database for GBS, for the 12 most represented sequence types within the database, as well as their corresponding clonal complexes (CCs), where appropriate.
Table 1. Summary of various serotyping distribution studies that highlight serotype variation showing the variety of patients, specimens, clinical presentation and methods used.

<table>
<thead>
<tr>
<th>Study</th>
<th>Location</th>
<th>Patients</th>
<th>Specimens</th>
<th>Clinical presentation</th>
<th>Serotyping method</th>
<th>Serotypes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin et al. 2016</td>
<td>Central Taiwan</td>
<td>Neonates - adults</td>
<td>Blood</td>
<td>Invasive</td>
<td>PCR</td>
<td>All 10</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pregnant women</td>
<td>Vaginal and rectal</td>
<td>Carriage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uh et al. 2005</td>
<td>South Korea</td>
<td>Adults</td>
<td>NS</td>
<td>Infection</td>
<td>Latex agglutination</td>
<td>Only Ia, Ib, II - VIII</td>
<td>25</td>
</tr>
<tr>
<td>Turner et al. 2012</td>
<td>Thai-Myanmar border</td>
<td>Pregnant women</td>
<td>Vaginal and rectal</td>
<td>Carriage</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lopez et al. 2017</td>
<td>Catalonia, Spain</td>
<td>Pregnant women</td>
<td>NS</td>
<td>Carriage</td>
<td>NS</td>
<td>NS</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elikwu et al. 2016</td>
<td>Nigeria</td>
<td>Pregnant women</td>
<td>Vaginal and rectal</td>
<td>Carriage</td>
<td>PCR</td>
<td>All 10</td>
<td>28</td>
</tr>
<tr>
<td>Foster-Nyarko et al. 2016</td>
<td>Gambia</td>
<td>Infants</td>
<td>Nasopharyngeal</td>
<td>Carriage</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romain et al. 2017</td>
<td>France</td>
<td>Neonate - infant</td>
<td>Blood and cerebrospinal fluids</td>
<td>Invasive</td>
<td>French National Reference Centre for</td>
<td>Only Ia - V</td>
<td>31</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Subjects</td>
<td>Specimens</td>
<td>Method</td>
<td>Other Notes</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------</td>
<td>------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------------</td>
<td>-------------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Bjorsdottir et al. 2016</td>
<td>Iceland</td>
<td>Adults</td>
<td>Blood and synovial fluid</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Teatero et al. 2017</td>
<td>Toronto, Canada</td>
<td>Pregnant women</td>
<td>Vaginal and rectal carriage</td>
<td>PCR and latex agglutination</td>
<td>All 10</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Teatero et al. 2015</td>
<td>Toronto, Canada</td>
<td>Neonates - adults</td>
<td>Blood, cerebrospinal fluid and tissue</td>
<td>PCR and latex agglutination</td>
<td>All 10</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Ferrieri et al. 2013</td>
<td>Minnesota, USA</td>
<td>Infant</td>
<td>Blood, cerebrospinal fluid, bone, joint and post-mortem from liver and lung</td>
<td>Immunoprecipitation</td>
<td>All 10</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Blumberg et al. 1996</td>
<td>Atlanta, USA</td>
<td>Infant - adults</td>
<td>Blood, peritoneal fluid and cerebrospinal fluid</td>
<td>Capillary precipitin</td>
<td>Only Ia, Ib, II - V</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Belard et al. 2015</td>
<td>Gabon</td>
<td>Pregnant women</td>
<td>Vaginal and rectal carriage</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Suara et al. 1994</td>
<td>Gambia</td>
<td>Pregnant women</td>
<td>Vaginal and rectal carriage</td>
<td>Capillary precipitin</td>
<td>Only I - VI</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neonates</td>
<td>Throat, rectal and umbilical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Group(s)</td>
<td>Carriage Sites</td>
<td>Detection Methods</td>
<td>Results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shabayek et al. 2014</td>
<td>Egypt</td>
<td>Pregnant and non-pregnant women</td>
<td>Vaginal Carriage</td>
<td>PCR and latex agglutination</td>
<td>All 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lachenauer et al. 1999</td>
<td>Japan</td>
<td>Pregnant women</td>
<td>Vaginal Carriage</td>
<td>Inhibition ELISA</td>
<td>Only Ia, Ib, II - VIII</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eskandrian et al. 2015</td>
<td>Malaysia</td>
<td>Neonates and adults</td>
<td>Vaginal and urine Carriage</td>
<td>PCR and latex agglutination</td>
<td>Only Ia, Ib II - VII</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Blood, wound and skin invasive/infection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ulcers, tracheal and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gastric aspirates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suhaimi et al. 2017</td>
<td>Malaysia</td>
<td>Neonates – adults</td>
<td>Vaginal, urine and placental Carriage</td>
<td>PCR and latex agglutination</td>
<td>All 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(80+ years)</td>
<td></td>
<td>Blood and tissue invasive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutra et al. 2014</td>
<td>Brazil</td>
<td>Pregnant women</td>
<td>Vaginal and perianal Carriage</td>
<td>Capillary precipitin</td>
<td>Only Ia, Ib, II - VIII</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adults</td>
<td>Blood, urine and male Carriage</td>
<td>invasive/infection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>genitale tract discharge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chaudhary et al. 2017</td>
<td>India</td>
<td>Pregnant women</td>
<td>Vaginal and rectal Carriage</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Islam et al. 2016</td>
<td>Bangladesh</td>
<td>Neonates</td>
<td>Umbilical cord Carriage</td>
<td>PCR</td>
<td>All 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Population</td>
<td>Sample Site(s)</td>
<td>Method</td>
<td>Serovars Detected</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------</td>
<td>--------------------------</td>
<td>----------------------------------------------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Ekelund et al. 2003</td>
<td>Denmark</td>
<td>Neonates – adults (80+ years)</td>
<td>Blood</td>
<td>Invasive</td>
<td>Precipitin test</td>
<td>Only Ia, Ib, II – VIII</td>
<td>46</td>
</tr>
<tr>
<td>Paoletti et al. 1999</td>
<td>Boston, USA</td>
<td>Neonates and adults</td>
<td>Genitourinary tract, others NS</td>
<td>NS</td>
<td>Precipitin test</td>
<td>Only Ia, Ib, III, V, VI and VIII</td>
<td>47</td>
</tr>
<tr>
<td>Yan et al. 2016</td>
<td>China</td>
<td>Pregnant women</td>
<td>NS</td>
<td>Carriage</td>
<td>PCR</td>
<td>All 10</td>
<td>48</td>
</tr>
<tr>
<td>Toyofuku et al. 2017</td>
<td>Japan</td>
<td>Neonates</td>
<td>Nasopharyngeal and rectal</td>
<td>Carriage</td>
<td>PCR</td>
<td>Only Ia, Ib, II – VIII</td>
<td>49</td>
</tr>
<tr>
<td>Slotved et al. 2017</td>
<td>Ghana</td>
<td>Pregnant women</td>
<td>Vaginal and rectal</td>
<td>Carriage</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td>42</td>
</tr>
<tr>
<td>Gudjonsdottir et al. 2015</td>
<td>Sweden</td>
<td>Neonates and adults</td>
<td>Blood, cerebrospinal fluid and synovial fluid</td>
<td>Invasive</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td>50</td>
</tr>
<tr>
<td>Takahara et al. 2015</td>
<td>Japan</td>
<td>Neonate (case study)</td>
<td>Blood and cerebral spinal fluid</td>
<td>Invasive</td>
<td>Latex agglutination</td>
<td>All 10</td>
<td>51</td>
</tr>
</tbody>
</table>