ORIGINAL ARTICLE

Phylogeny, recombination, and invasiveness of group B *Streptococcus* revealed by genomic comparisons of its global strains

Enze Lin [1,2,3] · Shengmei Zou [1,2,3] · Yue Wang [4] · Chien-Chung Lee [5,6] · Cheng-Hsun Chiu [5,6] · Ye Feng [1,2,3]

Abstract
Capsular polysaccharide (CPS) genes and pilus islands encode important virulence factors for group B *Streptococcus* (GBS) genomes. This study aims to detect phylogenetic inconsistency in CPS genes and pilus islands in GBSs and to explore its relationship with invasiveness. A total of 1016 GBS genomes were downloaded from the NCBI public database. The multilocus sequence typing (MLST) and Bayesian analysis of Population Structure (BAPS) analyses were both conducted for phylogeny construction. Serotyping and pilus typing were determined in silico using the genomic sequences. The CPS and pilus typing results were generally consistent with MLST and BAPS clustering. GBS isolates of serotype II and of the PI-1 + PI-2b and PI-2a types were more prone to phylogenetic inconsistency than the others. Isolates of serotype Ib and of PI-1 + PI-2a were more likely to appear as colonizing strains, whereas PI-2b was more likely to appear in invasive strains. For serotype V, phylogenetic inconsistency occurred more commonly in colonizing isolates, while for serotype III, the opposite occurred. The present study profiles for the first time the phylogenetic inconsistency of CPS genes and pilus islands in global GBS isolates, which is helpful for infection control and the development of new vaccines for the prevention of GBS occurrence.

Keywords Group B *Streptococcus* · Bayesian clustering · Multi-locus sequence typing · Pilus island · CPS type · Recombination · Invasiveness

Introduction

Group B *Streptococcus* (GBS, *Streptococcus agalactiae*) was initially described as the cause of bovine mastitis [1]. In humans, GBS usually occurs in the genitourinary or gastrointestinal tract of adult males and females and is not pathogenic.

In the birthing process, however, GBS from the genital tract of pregnant women can infect the lower gastrointestinal and upper respiratory tracts of newborns [2], resulting in neonatal sepsis and meningitis. Even with the appropriate antibiotic treatment, the mortality rates of late-onset GBS infection (age 7–90 days [3]) are high [4].

Enze Lin, Shengmei Zou and Yue Wang contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10096-020-04067-4) contains supplementary material, which is available to authorized users.

✉ Cheng-Hsun Chiu
chchiu@adm.cgmh.org.tw

✉ Ye Feng
pandafengye@zju.edu.cn

1 Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, People's Republic of China

2 Institute for Translational Medicine, Zhejiang University School of Medicine, Hangzhou, People's Republic of China

3 Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Hangzhou, People's Republic of China

4 Women's Hospital, Zhejiang University School of Medicine, Hangzhou Zhejiang, People's Republic of China

5 Molecular Infectious Disease Research Center, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Republic of China

6 Division of Pediatric Infectious Diseases, Department of Pediatrics, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Republic of China

Additionally, during the past decade, GBS has been increasingly associated with skin and soft tissue infections and bacteremia in nonpregnant adults [5].

Capsular polysaccharides (CPS) are considered to play a vital role as virulence factors [6]. The sialic acid residue on the CPS accelerates the dissociation of the C3 convertase, thereby inhibiting the complement cascade and escaping detection by the host's immune system. By identifying specific sequences in the CPS gene cluster, GBS can be subdivided into ten types [7–9]. The virulence of different CPS types varies: five major CPS types (Ia, Ib, II, III, and V) account for 96% and 88%, respectively, of the cases of invasive GBS infections in neonates and adults [10, 11]. Accordingly, similar to that found in *Streptococcus pneumoniae*, CPS is an essential target for the development of the GBS vaccine [12]. Currently, there is no vaccine on the market yet for GBS. The vaccines under development mainly target CPS Ia, Ib, II, III, and V [13–15].

It is well known that, in many bacterial pathogens, the use of vaccines can exert intense selective pressure upon bacterial evolution and change the epidemiological pattern. In *S. pneumoniae*, for example, the introduction of a 7-valent pneumococcal conjugate vaccine (PCV7) showed a decline of vaccine type (VT) and an increase of nonvaccine type (NVT) pneumococci in disease and nasopharyngeal carriage [16]. This alternation is achieved by capsule switching, which results actually from the recombination of essential CPS determinant genes [11, 17, 18]. GBS is naturally not transformable, and the rate of its capsular transformation shown by in vivo experiment is low [17]. Paradoxically, capsular switching has been frequently reported in GBS by a number of epidemiological studies, in particular in the clonal complex (CC) 1 and 17 [11, 18–22]. While most of these studies were conducted within the research subjects limited within a city or a country, the level of capsule switching on a global scale is still unclear. This information is of particular importance for the development of GBS vaccines because the propensity for capsule switching, if it really exists, would easily lead to "vaccine escape" once the monovalent vaccines start to use in the future.

The pilus-like structure in bacteria is another critical virulence factor and candidate target of vaccine development in GBS [23]. The structure consists of three components: a skeleton pilus protein and two accessory proteins that are involved in bacterial binding to host cells [24]. The GBS pilus is encoded by two loci in different regions of the genome, namely Pilus islands 1 and 2 (PI-1 and PI-2). The latter can be further divided into two distinct variants, PI-2a and PI-2b [25]. As a horizontal transfer element, typing based on pilus islands can be inconsistent with the true phylogeny of GBS [26]. Similar to CPS, knowledge on this on the global scale will benefit the development of GBS vaccines. Furthermore, when a certain GBS lineage possesses a phylogenetic inconsistent pilus type, it is unknown whether the lineage is more prone or more reluctant to a phylogenetic inconsistent CPS type, i.e., whether the recombination at the two loci will affect each other. It is also unknown whether these phylogenetic inconsistencies are associated with pathogenicity such as invasiveness.

The rapid development of WGS has made it cheaper for sequencing entire genomes. To date there have accumulated over 1000 GBS genomes in the public database, therefore providing good materials for addressing the above issues on a global scale. In the present study, we performed in silico prediction of the phylogeny, CPS, and pilus types for the GBS genomes in the NCBI Genbank database, and more importantly, investigated the relationship between the phylogenetic inconsistency and invasiveness of these isolates. It is hoped that this study helps better elucidate the relevance of recombination to GBS virulence and guide the development of GBS vaccines.

Methods

Bacterial isolates studied

All of the GBS genomes were downloaded from the NCBI Genbank database (downloaded on March 2, 2019; see Supplementary Table 1). The geographic and host information was extracted from the corresponding Biosample files. The pathogenicity of the human isolates was further divided as being invasive or colonizing based on the source information: those isolated from blood, cerebrospinal fluid, and placenta were considered invasive; those isolated from the genital tract or oral cavity were considered colonizing.

Phylogeny construction

The 7-gene multi-locus sequence typing (MLST) analysis (based on seven housekeeping genes) was performed using the online service, BacWGSTdb [27, 28]. MLST alleles and sequence types (STs) were assigned through the comparison of whole-genome data to the GBS MLST database (http://pubmlst.org/sagalactiae) [29]. The STs that differed by 1 or 2 alleles were further grouped to the same CC, following the nomenclature adopted by E. S. Björnsdóttir et al. [30].

Tree-independent hierarchical Bayesian clustering was performed to determine the population structure using hierBAPS (http://www.helsinki.fi/bsg/software/BAPS/). Briefly, the core genome multi-locus sequence typing (cgMLST) was performed using the BacWGSTdb service; the obtained allele matrix was used as input for the hierBAPS program; three levels of clustering were performed within the hierarchy, and a prior upper boundary of 20 clusters was established. A

neighbor-joining tree based on the concatenation of whole-genome single-nucleotide polymorphisms (SNPs) were built by using the BacWGSTdb service; the strain GBS-M002 (accession CP013908; serotype VI; pilus type 1, 2a; MLST ST1; collection region: Taiwan) was used as the reference genome.

CPS typing and pilus typing

For CPS typing, an in-house BLAST database was built based on information in the literature [31, 32], which included unique CPS genes for CPS identification. A sequence homology of > 95% with an alignment length of > 95% for the target gene was used as the threshold for predicting gene presence/absence. The CPS genes of CPS type IV and IX were too similar to be distinguished at the genomic level; thus, the two CPS types were split according to the SNPs at positions 327, 551, 1018, 1123, 1140, 1368, 1627, and 1832 [9, 33].

For pilus typing, an in-house BLAST database was established based on the literature [25, 34], which included all unique genes for pilus identification. A sequence homology of > 95% with an alignment length > 95% for the target gene was used as the threshold to predict gene presence/absence.

The inconsistency of phylogeny with CPS or pilus typing was defined as follows. Assume that CPS/pilus type a was most frequent within the BAPS cluster A whereas type b was the secondary one; meanwhile, majority of isolates of CPS/pilus type b belonged to another BAPS cluster B. Then we defined that the isolates of the BAPS cluster A carrying the CPS/pilus type b showed the inconsistency.

Statistical analysis

The data were analyzed for potential associations between such variables as CPS types and invasiveness using the SPSS program (Version 21.0 for Windows), utilizing the chi-square test of independence. The P value was further adjusted by FDR method when multiple comparisons were made. The significant level was set at $P < 0.05$.

Results

Phylogeny of GBS worldwide

All of the 1016 GBS genomes were downloaded from the NCBI Genbank database (Supplementary Table 1). These analyzed isolates were collected from 28 countries in Asia, Europe, South America, Africa, Australia, and North America (Supplementary Table 2). The seven gene-based MLST was performed to study the species' phylogeny. The predominant CCs of the isolates were identified as CC1, CC61, CC23, CC17, and CC19, accounting for 20.2% (n = 205), 16.7% (n

= 170), 15.3% (n = 156), 11.8% (n = 120), and 10.8% (n = 110) of the total isolates, respectively. Since housekeeping genes used for MLST may also undergo recombination that distorts the real phylogenetic relationships, we extracted the sequences of the core genomes and used a Bayesian clustering method (BAPS) for phylogeny reconstruction (Fig. 1). Thirteen BAPS clusters were identified based on the allele matrix of core genome MLST (cgMLST), including the most common cluster 3 (142 isolates, 14.0%), cluster 4 (142, 14.0%), cluster 2 (132, 13.0%), cluster 12 (120, 11.8%), cluster 5 (110, 10.8%), and cluster 8 (106, 10.4%).

Part of the BAPS clusters matched CCs perfectly, such as the BAPS cluster 3 corresponding to CC12, cluster 12 to CC17, and cluster 5 to CC19, whereas part of CCs seemed to exhibit a greater diversity as they comprised more than one BAPS cluster (Table 1; Fig. 1). For example, CC1 comprised BAPS clusters 4 and 11, CC23 comprised clusters 6, 8, and 10, CC61 comprised clusters 1 and 2. Overall, the two results were not contradictory to each other, showing the great compatibility of the two typing systems.

CPS typing of global GBS isolates

The CPS typing results demonstrated that CPS type II (250 isolates, 24.6%) was the most prevalent, followed by III (238, 23.4%), Ia (151, 14.9%), V (144, 14.2%), Ib (119, 11.7%), and IV (86, 8.5%). For a few BAPS clusters, their isolates possessed homogeneous CPS forms (Table 1; Fig. 1). For example, all isolates of BAPS cluster 6 possessed CPS type IV; cluster 7 possessed CPS type II; and cluster 13 possessed CPS type Ib. For the other BAPS clusters, however, their isolates showed miscellaneous CPS forms, with one CPS type appearing to be dominant and the secondary types likely to be phylogenetically inconsistent (Table 1). Statistically, the BAPS clusters 3, 4, and 5 were prone to such inconsistency, whereas BAPS clusters 2, 8, 9, 12, and 13 were reluctant to such inconsistency (Table 1).

From the view of CPS types, many CPS types had their infrequent BAPS clusters. While CPS type II was mostly found in BAPS clusters 1, 2, and 7, 27.6% (69/250) of its isolates appeared in BAPS clusters 3, 4, 5, 10, and 11. Similarly, 18.6% (16/86) of CPS type IV, 18.1% (26/144) of CPS type V, 14.3% (34/238) of CPS type III, and 2.0% (3/151) of CPS type Ia appeared in their infrequent BAPS clusters. Consequently, CPS types Ib and Ia were seldom involved in phylogenetical inconsistency, whereas CPS type II was most prone to phylogenetical inconsistency (Table 2).

Identification of Pilus Islands in GBS

Of the 1016 GBS isolates, 463 (45.8%) were found to carry one pilus island and 535 (52.7%) to carry two. PI-1, PI-2a, and
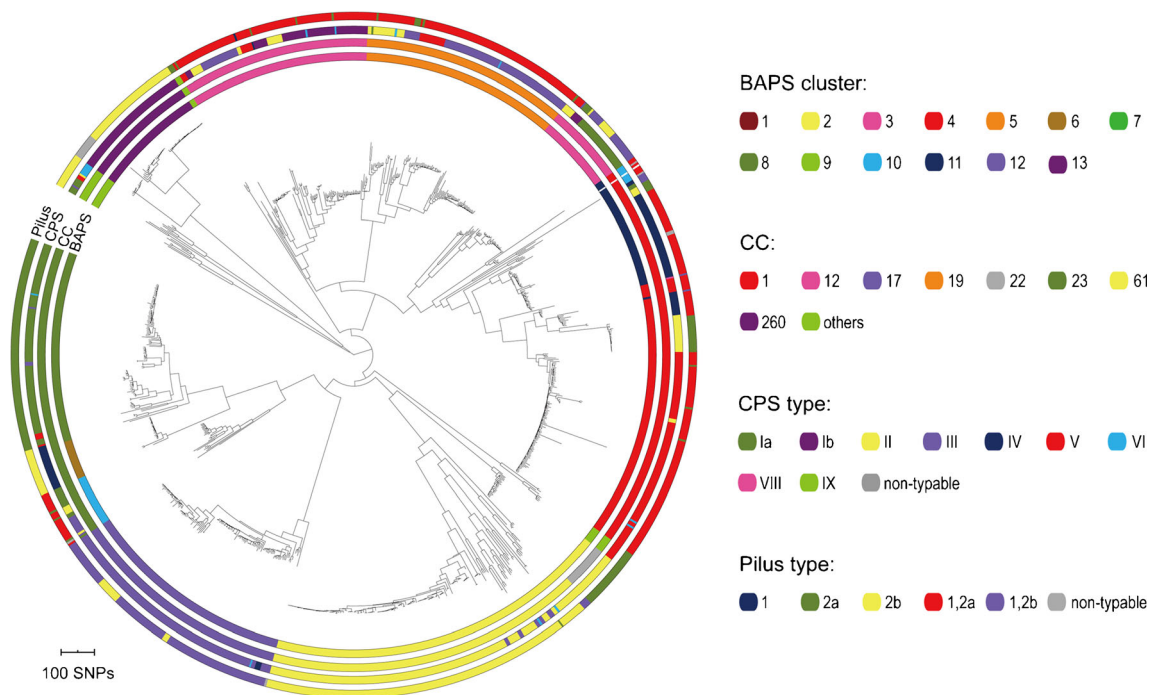
Fig. 1 Phylogeny and its relationship with BAPS, MLST, CPS type, and pilus type. The inner neighbor-joining tree was built based on the whole-genome SNPs. The rings from inner to outer represent BAPS cluster, clonal complex (CC) by 7-gene MLST, CPS type, and pilus type, respectively

PI-2b appeared in 540 (53.1%), 581 (57.2%), and 416 (40.9%) isolates. The most common form of pilus island in these isolates involved the simultaneous carriage of PI-1 + PI-2a (398, 39.2%). To further elaborate, all isolates of the BAPS cluster 2 carried PI-2b only, while most isolates of cluster 5 carried PI-1 + PI-2a (105/110, 95.5%), and most of cluster 8 carried PI-2a (105/106, 99.1%) (Fig. 1).

Since both the CPS typing and pilus typing results were overall consistent with BAPS clustering, majority of CPS type corresponded to a specific pilus type. The isolates carrying PI-1 + PI-2b were mainly of CPS type III (100/137, 73.0%), and the isolates of the type PI-2a were mainly of CPS types Ia and II (147/183, 80.3%). Conversely, the isolates of CPS type Ib mainly carried the PI-1 + PI-2a or PI-2b pilus (99/119, 83.2%), the isolates of CPS type III mainly carried the PI-1 + PI-2a or PI-1 + PI-2b pilus (203/238, 85.3%), and the isolates of CPS type V mainly carried the PI-1 + PI-2a pilus (125/144, 86.8%).

However, inconsistencies still existed between the pilus typing and the BAPS clustering. Taking the BAPS cluster as the unit, clusters 3 and 11 were prone to such inconsistency, whereas clusters 2, 8, 9, and 13 were reluctant to such inconsistency (Table 1). Taking the pilus type as the unit, isolates of the PI-1 + PI-2b and PI-2a types were prone to the phylogenetic inconsistency than the others, whereas isolates of the PI-1 + PI-2a type were reluctant to the inconsistency (Table 2).

Invasiveness of GBS

Of the total of 1016 isolates, 489 (48.1%) were of human sources. In detail, 187 isolates were derived from human blood, cerebrospinal fluid, and placenta and thus considered invasive. The remaining 302 isolates, mostly from the genital tract (125/302, 41.4%), were considered colonizing isolates. The BAPS clusters 4, 6, 7, and 12 appeared to be more virulent than the others, with > 40% of their isolates being invasive (Table 1).

Regarding the relationship with the CPS typing results, the invasive isolates were mainly distributed in CPS types III (69/187, 36.9%) and V (40/187, 21.4%) (Table 3). Although the numbers of invasiveness in these two CPS types was relatively high, the P value had not yet reached a significant level ($P > 0.05$). In contrast, CPS type Ib was more likely to appear in colonizing isolates than the other types ($P < 0.05$).

Regarding the distribution of the pilus types, the PI-2b type was more likely to appear in invasive isolates than other pilus types ($P < 0.05$; Table 3). In colonizing isolates, the PI-1 + PI-2a type was more common than the other types ($P < 0.05$; Table 3).

Relationships between recombination and invasiveness

Next, we analyzed whether the phylogenetically inconsistent isolates were concentrated in invasive isolates. For CPS type III isolates, 10 out of the 69 invasive isolates (14.5%) showed phylogenetic inconsistency; this proportion decreased to 2.1%

Table 1 The clonal complex (CC), CPS, and pilus composition in each BAPS cluster

BAPS	CC composition	CPS composition	Inconsistent with CPS typing	Pilus composition	Inconsistent with pilus typing	Invasiveness
1 (n = 38)	CC61 (n = 38)	II (n = 27, 71.1%) III (n = 9, 23.7%) Others (n = 2, 5.3%)	9 (23.7%)	2b (n = 34, 89.5%) 1, 2b (n = 4, 10.5%)	4 (10.5%)	0/2 (0.0%)
2 (n = 132)	CC61 (n = 132)	II (n = 130, 98.5%) III (n = 2, 1.5%)	2 (1.5%) *	2b (n = 132, 100%)	0 (0.0%) *	0/2 (0.0%)
3 (n = 142)	CC12 (n = 142)	Ib (n = 57, 40.1%) Ia (n = 31, 21.8%) Others (n = 54, 38.0%)	52 (36.6%) ‡	1, 2a (n = 100, 70.4%) 1, 2b (n = 23, 16.2%) Others (n = 19, 13.4%)	42 (29.6%) ‡	19/60 (31.7%)
4 (n = 142)	CC1 (n = 142)	V (n = 108, 76.1%) Others (n = 34, 24.0%)	32 (22.5%) ‡	1, 2a (n = 121, 85.21%) Others (n = 21, 14.79%)	21 (14.8%)	38/86 (44.2%)
5 (n = 110)	CC19 (n = 110)	III (n = 77, 70.0%) II (n = 17, 15.5%) Others (n = 16, 14.6%)	31 (28.2%) ‡	1, 2a (n = 105, 95.5%) 2a (n = 5, 4.5%)	5 (4.5%)	11/68 (16.2%)
6 (n = 22)	CC23 (n = 22)	IV (n = 22, 100%)	0 (0.0%)	2b (n = 22, 100%)	0 (0.0%)	15/20 (75.0%)
7 (n = 23)	CC22 (n = 23)	II (n = 23, 100%)	0 (0.0%)	2a (n = 23, 100%)	0 (0.0%)	6/14 (42.9%)
8 (n = 106)	CC23 (n = 106)	Ia (n = 97, 91.5%) Others (n = 9, 8.5%)	8 (7.5%) *	2a (n = 105, 99.1%) 2b (n = 1, 0.9%)	0 (0.0%) *	22/67 (32.8%)
9 (n = 28)	others (n = 28)	V (n = 10, 35.7%) Ia (n = 7, 25.00%) Others (n = 11, 39.3%)	0 (0.00%) *	2b (n = 17, 60.7%) 2a (n = 11, 39.3%)	0 (0.0%) *	1/8 (12.5%)
10 (n = 28)	CC23 (n = 28)	Ia (n = 13, 46.5%) III (n = 10, 35.7%) II (n = 5, 17.9%)	5 (17.9%)	1, 2a (n = 24, 85.7%) Others (n = 4, 14.3%)	4 (14.3%)	9/16 (56.3%)
11 (n = 63)	CC1 (n = 63)	IV (n = 48, 76.2%) Others (n = 15, 23.8%)	6 (9.5%)	1, 2a (n = 48, 76.2%) Others (n = 15, 23.8%)	11 (17.5%) ‡	19/49 (38.8%)
12 (n = 120)	CC17 (n = 120)	III (n = 116, 96.7%) Others (n = 4, 3.3%)	3 (2.5%) *	1, 2b (n = 104, 86.7%) Others (n = 16, 13.3%)	15 (12.5%)	47/99 (47.3%)
13 (n = 62)	CC260 (n = 62)	Ib (n = 62, 100%)	0 (0.0%) *	2b (n = 50, 80.6%) Others (n = 12, 19.4%)	0 (0.0%) *	na

* Significantly higher percentage of isolates with inconsistency with CPS or pilus typing

‡ Significantly lower percentage of isolates with inconsistency with CPS or pilus typing

Table 2 Percentage of isolates with phylogenetic inconsistency in each CPS and pilus type

	Phylogenetically inconsistent	Phylogenetically consistent	P value
Capsular genotype			
Ia	3 (2.0%)	148 (98.0%)	0.000
Ib	0 (0.0%)	119 (100.0%)	0.000
II	69 (27.6%)	181 (72.4%)	0.000
III	34 (14.3%)	204 (85.7%)	0.731
V	26 (18.1%)	118 (81.9%)	0.263
IV	16 (18.6%)	70 (81.4%)	0.324
Total	148 (15.0%)	840 (85.0%)	
Pilus type			
1 + 2a	0 (0.0%)	398 (100.0%)	0.000
1 + 2b	33 (24.1%)	104 (75.9%)	0.000
2a	43 (23.5%)	140 (76.5%)	0.000
2b	24 (8.6%)	255 (91.4%)	0.349
Total	100 (10.0%)	897 (90.0%)	

Note: minor CPS type or pilus type (< 1% of total isolates) are excluded from this analysis

(2/94) in the colonizing isolates (Fig. 2a). The situation in CPS type V was the converse: phylogenetic inconsistency in CPS type V isolates was more common in colonizing isolates (14/57, 24.6%) than invasive ones (2/40, 5.0%, P < 0.05; Fig. 2a).

For each pilus type, we did not find an uneven recombination distribution between invasive and colonizing isolates.

Discussion

This study determines the distribution of CPS types and pilus types among global GBS isolates and, in particular,

Table 3 The association between invasiveness and CPS and pilus typing

	Invasive isolates	Colonizing isolates	P value
CPS type			
Ia	22 (29.7%)	52 (70.3%)	0.102
Ib	2 (8.7%)	21 (91.3%)	0.003
II	16 (33.3%)	32 (66.7%)	0.461
III	69 (42.3%)	94 (57.7%)	0.188
IV	35 (46.1%)	41 (53.9%)	0.127
V	40 (41.2%)	57 (58.8%)	0.498
Pilus type			
1 + 2a	87 (33.6%)	172 (66.4%)	0.025
1 + 2b	36 (40.4%)	53 (59.6%)	0.636
2a	33 (34.0%)	64 (66.0%)	0.339
2b	30 (75.0%)	10 (25.0%)	0.000

Note: minor CPS type or pilus type (< 1% of total isolates) are excluded from this analysis

investigates whether the recombination events occur within CPSs and pilus islands. The motivation for this study is that capsular switching through recombination has been observed during the last decade in *S. pneumoniae* as the primary approach by which the pathogen eludes vaccines. As CPS and pilus typing has been frequently reported to show phylogenetic inconsistency in GBS, this pathogen is very likely to adopt the same strategy to *S. pneumoniae* to escape the host's immune system when vaccine starts to use in the future. It is therefore imperative to profile the global recombination pattern in these loci before GBS vaccines go into commercial use.

As a genetic event for laterally exchanging DNA, recombination is universally present in bacteria and results in an incongruent phylogeny between local recombinant fragments and adjacent regions. In the past decades, MLST has been considered a gold standard typing tool for characterizing bacterial isolates from the sequences of internal fragments of (usually) seven housekeeping genes. Nevertheless, recombination in the housekeeping genes has been reported in a number of bacterial pathogens, including *Escherichia coli*, *Salmonella enterica*, and *Staphylococcus aureus* [35–37]. The MLST-based phylogeny is not entirely free from the noise created by recombination, and scientists have turned to WGS to minimize this interference. By comparing the results of the genome-wide BAPS and the 7-gene MLST, we found that the two results matched well with each other, suggesting that the seven genes used for MLST in GBS had undergone little recombination and were good candidates for within-species typing. A combination of both BAPS clustering and MLST is expected to better accurately assess the bacterial phylogeny, based on which we further predicted the strains' phylogenetic consistency with the CPS and pilus typing results.
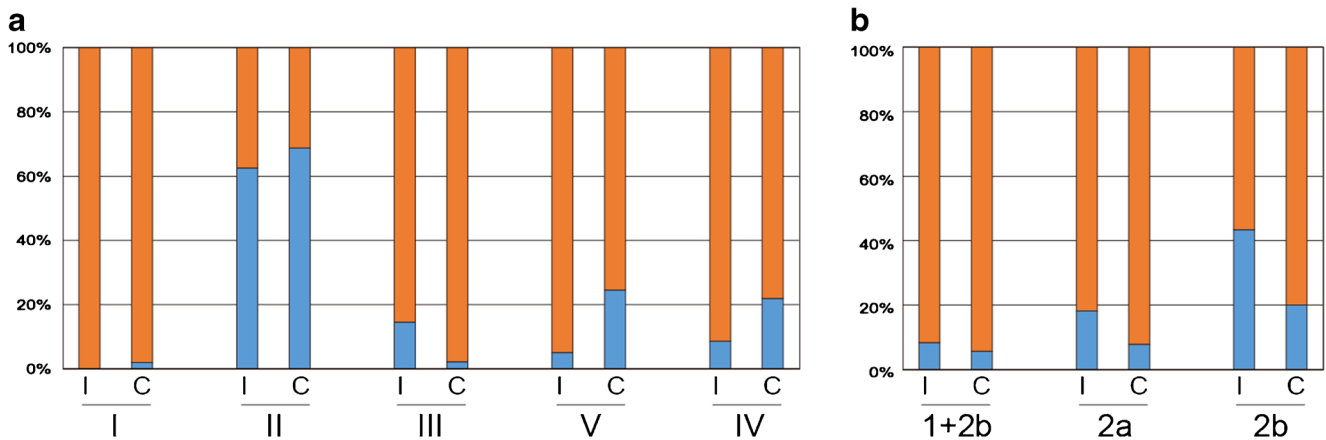
Fig. 2 The relationship between phylogenetic inconsistency and invasiveness. For each CPS type (in panel **a**) and pilus type (in panel **b**), the isolates were further divided into invasive isolates and colonizing isolates. The *y*-axis indicates the proportion of phylogenetically inconsistent isolates (in blue) and phylogenetically consistent isolates (in orange). I, invasive isolates; C, colonizing isolates

Epidemiologically, CPS types have been shown to vary between years and nations. Even more importantly, CPS types may be linked with human diseases such as early-onset (age 0–6 days) sepsis and meningitis [38, 39]. The varying pathogenic property is likely to be attributed to a specific CPS, but may also likely be due to a combination of virulence factors specific to the particular genetic lineage. Here, our analysis of global isolates revealed that the human invasive isolates were mainly concentrated in CPS types III and V, whereas CPS type Ib is more likely to appear in colonizing isolates than other types. Several studies based on other techniques than WGS have also found that CPS type V exhibits high invasiveness; this includes research performed in Shanghai, China [5]; England and Wales [40]; Portugal [41]; and Alberta, Canada [42]. Fortunately, the multivalent vaccines employed in the present trial currently have already covered CPS type V as well as Ia, Ib, II, III, and IV [13–15].

Despite the overall pronounced correlation between CPS type and BAPS clustering, a nonnegligible proportion of GBS isolates did show the signal of recombination in their CPS clusters. Among the main CPS types, CPS type II was more prone to phylogenetic inconsistency, whereas CPS type Ib and Ia had the reverse trend. Concerning the relevance with invasiveness, the colonizing isolates of CPS type V are more prone to phylogenetic inconsistency than the invasive ones. These findings are consistent with previous reports that the colonizing isolates have a greater tendency for capsular switching [43]. A plausible explanation is that persistent colonization facilitates the lateral genetic transfer of GBS isolates, thereby promoting capsular switching. Paradoxically, for CPS type III, phylogenetic inconsistency is less frequent in colonizing than in invasive isolates. We attribute this difference to an opposite driving force arising from the immunogenic pressure induced by host immunity [44]: bacteria resort to CPS switching to evade host immunity following the host's generation of antibodies against the CPS. The eventual extent of phylogenetic inconsistency may depend on the interaction of the above two opposing forces and the fitness of the recombinant strain.

The pilus-like structure plays a role in GBS colonization and pathogenicity, making it another potential target for the vaccine development. We found that PI-1 + PI-2a accounts for the highest proportion of pili. This combination of pilus types is more likely to occur in colonizing isolates but is less prone to phylogenetic inconsistency. This situation is in contrast with that for the CPS, possibly because this pilus combination may have adapted to a higher degree to the colonization environment. In contrast, the isolates carrying PI-2a or PI-1 + PI-2b have been frequently involved in phylogenetic inconsistency. According to previous reports, PI-2b is associated with invasive infections in neonates, and PI-2a with invasive diseases in adults [26, 31, 32, 45]. It is unclear whether this phenomenon results from the different immune statuses of children and adults. Here, we have only observed the association between PI-2b and invasiveness, probably due to the sampling bias of the public database towards child patients. Meanwhile, we also found that the PI-2b isolates did not enhance invasiveness by acquiring PI-2b in a recombinant manner; instead, PI-2b was inherently present in the isolates. It is possible that PI-2b itself benefits the evasion of the host immune system, or that the genetic background of these lineages carries other genes beneficial for invasion.

When the phylogenetic inconsistencies of the CPS and pilus typing results were taken into consideration together, we found that the BPAS clusters 4 and 5 had the CPS locus prone to the inconsistency whereas the BAPS cluster 11 had the pilus locus prone to the inconsistency. This suggests that for these lineages, recombination may occur independently in the two loci. However, the BAPS cluster 3 had its both loci prone to phylogenetic inconsistency, indicating that the double switch at the loci may better contribute to the cluster's fitness instead of a single switch does.

The strength of this study was the global collection of GBS genomes used for analysis, which meanwhile brings a few limitations. First, many isolates lack their detailed clinical information in the public database, therefore making many analyses unavailable such as the categorization between early-onset and late-onset diseases caused by GBS. Second, although the public database collects bacterial genomes worldwide, many genomes were released from a few countries, majority of which were from Asia, Europe, and North America. Consequently, it is unknown whether the conclusions drawn in this study applies to the countries not represented in the public database.

Conclusions

This study characterizes the global distribution of GBS's CPS type and pilus type as well as its relevance to invasive disease. While we found little recombination occurring in MLST genes, the recombination signals were detected in both CPS and pilus genes. This suggests that GBS may rely on recombination upon specific virulence genes to better defend the host's immune system. The present findings are beneficial for the current development of GBS vaccines. The CPS types and pilus types that are frequently involved in recombination need to be covered as the primary vaccine targets. Furthermore, continuing surveillance is required to reflect whether capsule switching would occur when the GBS vaccines start to use in the future, which can be revealed by comparison of current and future CPS distributions.

Compliance with ethical standards

References

1. Keefe GP (1997) Streptococcus agalactiae mastitis: a review. Can Vet J 38(7):429–437

2. Skoff TH, Farley MM, Petit S, Craig AS, Schaffner W, Gershman K, Harrison LH, Lynfield R, Mohle-Boetani J, Zansky S, Albanese BA, Stefonek K, Zell ER, Jackson D, Thompson T, Schrag SJ (2009) Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990-2007. Clin Infect Dis 49(1):85–92. https://doi.org/10.1086/599369

3. Heath PT, Balfour G, Weisner AM, Efstratiou A, Lamagni TL, Tighe H, O'Connell LA, Cafferkey M, Verlander NQ, Nicoll A, McCartney AC, Group PGBSW (2004) Group B streptococcal disease in UK and Irish infants younger than 90 days. Lancet 363(9405):292–294. https://doi.org/10.1016/S0140-6736(03)15389-5

4. Centers for Disease C, Prevention (2007) Perinatal group B streptococcal disease after universal screening recommendations–United States, 2003-2005. MMWR Morb Mortal Wkly Rep 56(28):701–705

5. Jiang H, Chen M, Li T, Liu H, Gong Y, Li M (2016) Molecular characterization of streptococcus agalactiae causing community- and hospital-acquired infections in Shanghai, China. Front Microbiol 7:1308. https://doi.org/10.3389/fmicb.2016.01308

6. Wessels MR, Rubens CE, Benedi VJ, Kasper DL (1989) Definition of a bacterial virulence factor: sialylation of the group B streptococcal capsule. Proc Natl Acad Sci U S A 86(22):8983–8987. https://doi.org/10.1073/pnas.86.22.8983

7. Ryc M, Jelinkova J, Motlova J, Wagner M (1988) Immuno-electronmicroscopic demonstration of capsules on group-B streptococci of new serotypes and type candidates. J Med Microbiol 25(2):147–149. https://doi.org/10.1099/00222615-25-2-147

8. Ferrieri P, Flores AE (1997) Surface protein expression in group B streptococcal invasive isolates. Adv Exp Med Biol 418:635–637. https://doi.org/10.1007/978-1-4899-1825-3_148

9. Slotved HC, Kong F, Lambertsen L, Sauer S, Gilbert GL (2007) Serotype IX, a proposed new streptococcus agalactiae serotype. J Clin Microbiol 45(9):2929–2936. https://doi.org/10.1128/JCM.00117-07

10. Dutra VG, Alves VM, Olendzki AN, Dias CA, de Bastos AF, Santos GO, de Amorin EL, Sousa MA, Santos R, Ribeiro PC, Fontes CF, Andrey M, Magalhaes K, Araujo AA, Paffadore LF, Marconi C, Murta EF, Fernandes PC Jr, Raddi MS, Marinho PS, Bornia RB, Palmeiro JK, Dalla-Costa LM, Pinto TC, Botelho AC, Teixeira LM, Fracalanzza SE (2014) Streptococcus agalactiae in Brazil: serotype distribution, virulence determinants and antimicrobial susceptibility. BMC Infect Dis 14:323. https://doi.org/10.1186/1471-2334-14-323

11. Bellais S, Six A, Fouet A, Longo M, Dmytruk N, Glaser P, Trieu-Cuot P, Poyart C (2012) Capsular switching in group B Streptococcus CC17 hypervirulent clone: a future challenge for polysaccharide vaccine development. J Infect Dis 206(11):1745–1752. https://doi.org/10.1093/infdis/jis605

12. Chen VL, Avci FY, Kasper DL (2013) A maternal vaccine against group B Streptococcus: past, present, and future. Vaccine 31(Suppl 4):D13–D19. https://doi.org/10.1016/j.vaccine.2012.12.080

13. Heath PT (2011) An update on vaccination against group B streptococcus. Expert Rev Vaccines 10(5):685–694. https://doi.org/10.1586/erv.11.61

14. Heath PT (2016) Status of vaccine research and development of vaccines for GBS. Vaccine 34(26):2876–2879. https://doi.org/10.1016/j.vaccine.2015.12.072

15. Buurman ET, Timofeyeva Y, Gu J, Kim JH, Kodali S, Liu Y, Mininni T, Moghazeh S, Pavliakova D, Singer C, Singh S, Handke LD, Lotvin J, Prasad AK, Scully IL, Donald RGK, Jansen KU, Anderson AS (2019) A novel hexavalent capsular polysaccharide conjugate vaccine (GBS6) for the prevention of neonatal group B Streptococcal infections by maternal immunization. J Infect Dis 220(1):105–115. https://doi.org/10.1093/infdis/jiz062

16. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Linares J, Jacobs MR, Kristinsson KG, Beall BW, Klugman KP, Parkhill J, Hakenbeck R, Bentley SD, Brueggemann AB (2013) Pneumococcal capsular switching: a historical perspective. J Infect Dis 207(3):439–449. https://doi.org/10.1093/infdis/jis703

17. Martins ER, Melo-Cristino J, Ramirez M (2010) Evidence for rare capsular switching in Streptococcus agalactiae. J Bacteriol 192(5):1361–1369. https://doi.org/10.1128/JB.01130-09

18. Neemuchwala A, Teatero S, Athey TB, McGeer A, Fittipaldi N (2016) Capsular switching and other large-scale recombination events in invasive sequence type 1 group B Streptococcus. Emerg Infect Dis 22(11):1941–1944. https://doi.org/10.3201/eid2211.152064

19. Seale AC, Koech AC, Sheppard AE, Barsosio HC, Langat J, Anyango E, Mwakio S, Mwarumba S, Morpeth SC, Anampiu K, Vaughan A, Giess A, Mogeni P, Walusuna L, Mwangudzah H, Mwanzui D, Salim M, Kemp B, Jones C, Mturi N, Tsofa B, Mumbo E, Mulewa D, Bandika V, Soita M, Owiti M, Onzere N, Walker AS, Schrag SJ, Kennedy SH, Fegan G, Crook DW, Berkley JA (2016) Maternal colonization with Streptococcus agalactiae and associated stillbirth and neonatal disease in coastal Kenya. Nat Microbiol 1(7):16067. https://doi.org/10.1038/nmicrobiol.2016.67

20. Meehan M, Cunney R, Cafferkey M (2014) Molecular epidemiology of group B streptococci in Ireland reveals a diverse population with evidence of capsular switching. Eur J Clin Microbiol Infect Dis 33(7):1155–1162. https://doi.org/10.1007/s10096-014-2055-5

21. Nagano N, Koide S, Hayashi W, Taniguchi Y, Tanaka H, Maeyama Y, Suzuki M, Kimura K, Arakawa Y, Nagano Y (2019) Population-level transition of capsular polysaccharide types among sequence type 1 group B Streptococcus isolates with reduced penicillin susceptibility during a long-term hospital epidemic. Int J Antimicrob Agents 53(3):203–210. https://doi.org/10.1016/j.ijantimicag.2018.11.003

22. Teatero S, McGeer A, Low DE, Li A, Demczuk W, Martin I, Fittipaldi N (2014) Characterization of invasive group B Streptococcus strains from the Greater Toronto Area, Canada. J Clin Microbiol 52(5):1441. https://doi.org/10.1128/JCM.03554-13

23. Margarit I, Rinaudo CD, Galeotti CL, Maione D, Ghezzo C, Buttazzoni E, Rosini R, Runci Y, Mora M, Buccato S, Pagani M, Tresoldi E, Berardi A, Creti R, Baker CJ, Telford JL, Grandi G (2009) Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm. J Infect Dis 199(1):108–115. https://doi.org/10.1086/595564

24. Dramsi S, Caliot E, Bonne I, Guadagnini S, Prevost MC, Kojadinovic M, Lalioui L, Poyart C, Trieu-Cuot P (2006) Assembly and role of pili in group B streptococci. Mol Microbiol 60(6):1401–1413. https://doi.org/10.1111/j.1365-2958.2006.05190.x

25. Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, Maione D, Taddei A, Santi I, Ghezzo C, Brettoni C, Buccato S, Margarit I, Grandi G, Telford JL (2006) Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae. Mol Microbiol 61(1):126–141. https://doi.org/10.1111/j.1365-2958.2006.05225.x

26. Springman AC, Lacher DW, Waymire EA, Wengert SL, Singh P, Zadoks RN, Davies HD, Manning SD (2014) Pilus distribution among lineages of group b streptococcus: an evolutionary and clinical perspective. BMC Microbiol 14:159. https://doi.org/10.1186/1471-2180-14-159

27. Feng Y, Zou S, Chen H, Yu Y, Ruan Z (2020) BacWGSTdb 2.0: a one-stop repository for bacterial whole-genome sequence typing and source tracking. Nucleic Acids Res. https://doi.org/10.1093/nar/gkaa821

28. Ruan Z, Feng Y (2016) BacWGSTdb, a database for genotyping and source tracking bacterial pathogens. Nucleic Acids Res 44(D1):D682–D687. https://doi.org/10.1093/nar/gkv1004

29. Manning SD, Lewis MA, Springman AC, Lehotzky E, Whittam TS, Davies HD (2008) Genotypic diversity and serotype distribution of group B streptococcus isolated from women before and after delivery. Clin Infect Dis 46(12):1829–1837. https://doi.org/10.1086/588296

30. Bjornsdottir ES, Martins ER, Erlendsdottir H, Haraldsson G, Melo-Cristino J, Kristinsson KG, Ramirez M (2016) Changing epidemiology of group B streptococcal infections among adults in Iceland: 1975-2014. Clin Microbiol Infect 22(4):379 e379–379 e316. https://doi.org/10.1016/j.cmi.2015.11.020

31. Cieslewicz MJ, Chaffin D, Glusman G, Kasper D, Madan A, Rodrigues S, Fahey J, Wessels MR, Rubens CE (2005) Structural and genetic diversity of group B streptococcus capsular polysaccharides. Infect Immun 73(5):3096–3103. https://doi.org/10.1128/IAI.73.5.3096-3103.2005

32. Poyart C, Tazi A, Reglier-Poupet H, Billoet A, Tavares N, Raymond J, Trieu-Cuot P (2007) Multiplex PCR assay for rapid and accurate capsular typing of group B streptococci. J Clin Microbiol 45(6):1985–1988. https://doi.org/10.1128/JCM.00159-07

33. Kapatai G, Patel D, Efstratiou A, Chalker VJ (2017) Comparison of molecular serotyping approaches of Streptococcus agalactiae from genomic sequences. BMC Genomics 18(1):429. https://doi.org/10.1186/s12864-017-3820-5

34. Khodaei F, Najafi M, Hasani A, Kalantar E, Sharifi E, Amini A, Aghazadeh M (2018) Pilus-encoding islets in S. agalactiae and its association with antibacterial resistance and serotype distribution. Microb Pathog 116:189–194. https://doi.org/10.1016/j.micpath.2018.01.035

35. Basic-Hammer N, Vogel V, Basset P, Blanc DS (2010) Impact of recombination on genetic variability within Staphylococcus aureus clonal complexes. Infect Genet Evol 10(7):1117–1123. https://doi.org/10.1016/j.meegid.2010.07.013

36. Lan R, Reeves PR, Octavia S (2009) Population structure, origins and evolution of major Salmonella enterica clones. Infect Genet Evol 9(5):996–1005. https://doi.org/10.1016/j.meegid.2009.04.011

37. Turrientes MC, Gonzalez-Alba JM, del Campo R, Baquero MR, Canton R, Baquero F, Galan JC (2014) Recombination blurs phylogenetic groups routine assignment in Escherichia coli: setting the record straight. PLoS One 9(8):e105395. https://doi.org/10.1371/journal.pone.0105395

38. Martins ER, Pessanha MA, Ramirez M, Melo-Cristino J, Portuguese Group for the Study of Streptococcal I (2007) Analysis of group B streptococcal isolates from infants and pregnant women in Portugal revealing two lineages with enhanced invasiveness. J Clin Microbiol 45(10):3224–3229. https://doi.org/10.1128/JCM.01182-07

39. Poyart C, Reglier-Poupet H, Tazi A, Billoet A, Dmytruk N, Bidet P, Bingen E, Raymond J, Trieu-Cuot P (2008) Invasive group B streptococcal infections in infants, France. Emerg Infect Dis 14(10):1647–1649. https://doi.org/10.3201/eid1410.080185

40. Lamagni TL, Keshishian C, Efstratiou A, Guy R, Henderson KL, Broughton K, Sheridan E (2013) Emerging trends in the epidemiology of invasive group B streptococcal disease in England and Wales, 1991-2010. Clin Infect Dis 57(5):682–688. https://doi.org/10.1093/cid/cit337

41. Martins ER, Melo-Cristino J, Ramirez M, Portuguese Group for the Study of Streptococcal I (2012) Dominance of serotype Ia among group B Streptococci causing invasive infections in nonpregnant adults in Portugal. J Clin Microbiol 50(4):1219–1227. https://doi.org/10.1128/JCM.05488-11

42. Alhhazmi A, Hurteau D, Tyrrell GJ (2016) Epidemiology of invasive group B Streptococcal disease in Alberta, Canada, from 2003 to 2013. J Clin Microbiol 54(7):1774–1781. https://doi.org/10.1128/JCM.00355-16

43. Teatero S, Ferrieri P, Martin I, Demczuk W, McGeer A, Fittipaldi N (2017) Serotype distribution, population structure, and antimicrobial resistance of group B Streptococcus strains recovered from colonized pregnant women. J Clin Microbiol 55(2):412–422. https://doi.org/10.1128/JCM.01615-16

44. Lipsitch M (1999) Bacterial vaccines and serotype replacement: lessons from Haemophilus influenzae and prospects for Streptococcus pneumoniae. Emerg Infect Dis 5(3):336–345. https://doi.org/10.3201/eid0503.990304

45. Madzivhandila M, Adrian PV, Cutland CL, Kuwanda L, Madhi SA, Po PSTT (2013) Distribution of pilus islands of group B streptococcus associated with maternal colonization and invasive disease in South Africa. J Med Microbiol 62(Pt 2):249–253. https://doi.org/10.1099/jmm.0.052951-0