

RESEARCH ARTICLE

Open Access

Pilus distribution among lineages of group b *streptococcus*: an evolutionary and clinical perspective

Amber Cody Springman^{1*}, David W Lacher², Emily A Waymire¹, Samantha L Wengert¹, Pallavi Singh¹, Ruth N Zadoks^{3,4}, H Dele Davies⁵ and Shannon D Manning^{1*}

Abstract

Background: Group B *Streptococcus* (GBS) is an opportunistic pathogen in both humans and bovines. Epidemiological and phylogenetic analyses have found strains belonging to certain phylogenetic lineages to be more frequently associated with invasive newborn disease, asymptomatic maternal colonization, and subclinical bovine mastitis. Pilus structures in GBS facilitate colonization and invasion of host tissues and play a role in biofilm formation, though few large-scale studies have estimated the frequency and diversity of the three pilus islands (PIs) across diverse genotypes. Here, we examined the distribution of pilus islands (PI) 1, 2a and 2b among 295 GBS strains representing 73 multilocus sequence types (STs) belonging to eight clonal complexes. PCR-based RFLP was also used to evaluate variation in the genes encoding pilus backbone proteins of PI-2a and PI-2b.

Results: All 295 strains harbored one of the PI-2 variants and most human-derived strains contained PI-1. Bovine-derived strains lacked PI-1 and possessed a unique PI-2b backbone protein allele. Neonatal strains more frequently had PI-1 and a PI-2 variant than maternal colonizing strains, and most CC-17 strains had PI-1 and PI-2b with a distinct backbone protein allele. Furthermore, we present evidence for the frequent gain and loss of genes encoding certain pilus types.

Conclusions: These data suggest that pilus combinations impact host specificity and disease presentation and that diversification often involves the loss or acquisition of PIs. Such findings have implications for the development of GBS vaccines that target the three pilus islands.

Keywords: *Streptococcus agalactiae*, Pilus, MLST, Molecular epidemiology

Background

Although group B *Streptococcus* (GBS, *Streptococcus agalactiae*) was originally described as a cause of mastitis in bovines, it has emerged as an important opportunistic pathogen in humans. GBS is typically a commensal in the urogenital and lower gastrointestinal tracts of healthy adults, and pregnant women can transmit the bacterium to their baby during childbirth. Newborns infected with GBS can develop life threatening infections including pneumonia, sepsis, and meningitis. GBS has also been shown to cause disease in the elderly and adults with

underlying medical conditions where skin and soft tissue infections, urinary tract infections, and bacteremia can result [1].

Molecular epidemiological studies utilizing multilocus sequence typing (MLST) have shown that the distribution of GBS lineages varies by source. Strains belonging to clonal complex (CC)-17 and CC-19, for example, more frequently caused newborn disease compared to strains of other CCs [2-4], with CC-17 strains causing more cases of meningitis and late-onset disease [2]. By contrast, the frequency of strains belonging to CCs 1, 23 and 19 was higher in asymptomatic pregnant women than newborns [3-5], though strains belonging to all three CCs have been frequently isolated from adult and elderly patients with invasive infections [6]. While a subset of CCs have been isolated from both humans and bovines, strains belonging

* Correspondence: manning71@msu.edu

[†]Deceased

¹Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA

Full list of author information is available at the end of the article

to CC-61 and CC-67 have been found exclusively in cattle [7-10]. Factors that dictate host specificity are poorly understood although several studies have shown that human- and bovine-derived strains have distinct genetic characteristics [7,8,11-13] that may facilitate adaptation to a particular species. Bovine strain FSL S3-026, for instance, was found to have a high frequency of insertion and strain-specific sequences that differed from eight human-derived genomes [13].

Surface adhesins and pili play important roles in GBS adaptation and host specificity. Three pilus islands, (PI)-1, PI-2a, and PI-2b, which encode distinct pilus structures that mediate interactions with host cells, have been identified [14]. Each PI encodes three structural proteins, a backbone protein (BP), two ancillary proteins (AP) and two pilus-specific class C sortase enzymes [15] that recognize LPXTG amino acid motifs on structural proteins and facilitate covalent attachment of these subunits to each other and the cell wall peptidoglycan [16,17]. Differences between PI-1 and the PI-2 variants have been noted [15]. PI-1 is a 16 kb element that integrates between genes *sag0633* and *sag0652* and is flanked by direct repeats, thereby facilitating horizontal gene transfer. PI-2a and PI-2b, however, integrate into one site between genes *sag1410* and *sag1403* and thus, only one or the other can be present in each strain.

In vitro models of GBS infection have shown that the APs initiate adherence to various tissues, whereas the BPs facilitate invasion and paracellular translocation of host cells [18-20]. Furthermore, PI-2a was suggested to be more important for biofilm formation [21,22] and the presence of the PI-2b protein, Spb1/SAN1518, was found to increase intracellular survival in macrophages [23]. *In vivo*, GBS pilus components are highly immunogenic and a pilus-vaccine containing the BP genes of PI-1 and PI-2b and the AP of PI-2a has been shown to elicit opsonophagocytic antibodies that confer protection in mice [24].

Given the role that pili play in GBS colonization and disease progression, the type of pilus likely impacts GBS colonization and invasion of host cells. Few studies, however, have characterized the distribution and genetic diversity of each PI in a large population of phylogenetically distinct GBS strains from various sources. Here, we screened for the presence of PI-1, PI-2a and PI-2b in 295 strains recovered from humans and bovines to examine the distribution of each PI across phylogenetic lineages resolved by MLST and identified associations with clinical phenotypes. Genetic variation was also examined in the PI-2 BPs to correlate alleles with phylogenetic lineages, host, and disease. A comprehensive analysis of PIs across diverse strain populations is important to guide current efforts aimed at developing pilus-based GBS vaccines.

Results

Phylogenetic analysis

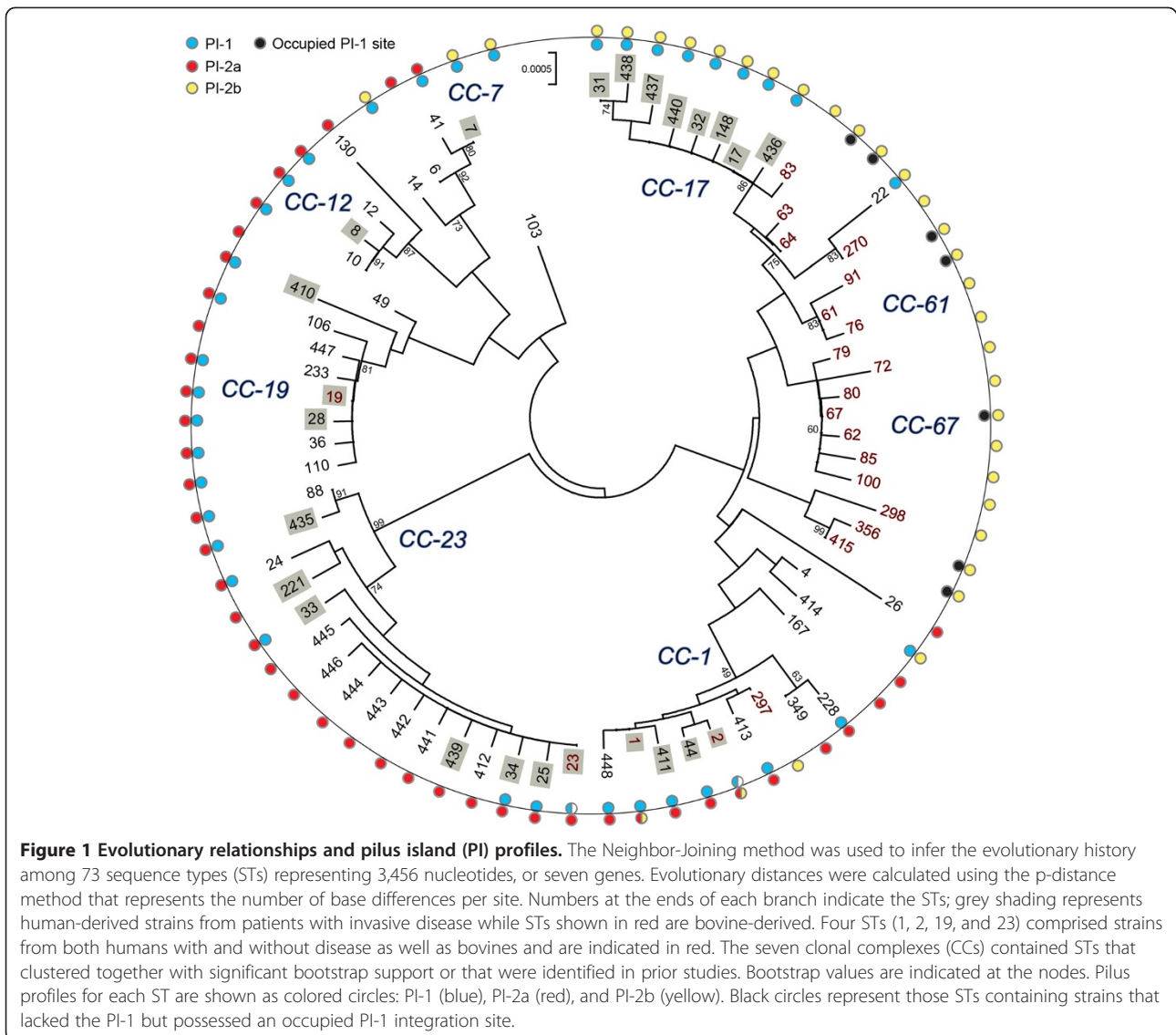
Application of MLST to the 295 strains grouped the 73 sequence types (STs) into eight clusters (Figure 1). Although CC-1 had low bootstrap support (49%), we considered it a cluster since our prior study [2] grouped ST-1 with the same STs included in this analysis. The difference in this study was due to the inclusion of the bovine-derived ST-297 strain. The same was true for CC-67, which comprised STs 62, 67, 80, 85, and 100 at 60% bootstrap support. Six singletons (STs 26, 49, 103, 167, 298, and 410) and four smaller clusters were also identified. Neighbor-net analysis provided evidence of recombination among the 73 STs (Figure 2).

Most strains represented CC-19 ($n = 88$; 30%), CC-17 ($n = 70$; 24%), CC-1 ($n = 36$; 12%), and CC-23 ($n = 30$; 10%). CC-23 was the most diverse with 16 unique STs, whereas CCs 17, 19 and 1 included nine, seven, and seven STs, respectively. By contrast, CCs 61 and 67 were exclusively comprised of bovine strains, while the remaining bovine strains belonged to three smaller clusters with low bootstrap values or CCs containing mostly human-derived strains. STs 1, 2, 19 and 23 had strains of both human and bovine origin.

Distribution of PIs across CCs and BP gene variation

Most strains ($n = 224$; 76%) contained PI-1 plus one of the two PI-2 variants, while 71 strains had PI-2a or PI-2b exclusively. PIs were correlated with specific lineages and closely related lineages had similar profiles. While PI-1 had a widespread distribution, the presence of PI-2a and PI-2b was non-random. Within CC's, little variation was observed in the frequency of PI-2a and PI-2b except in CCs 1 and 7, which had a range of PI profiles. PI-1 frequencies, however, varied within and across CCs, particularly in human strains (Figure 3). Most CC-23 strains ($n = 18$; 60%), for example, lacked PI-1, whereas virtually all CC-19 ($n = 88$; 100%) and CC-17 ($n = 69$; 99%) strains had PI-1 with one PI-2 variant. The only CC-17 strain without PI-1 (ST-83) originated from a bovine. Among strains of the same ST, multiple profiles were observed in two CCs. Within ST-1, all strains had PI-1/PI-2a ($n = 14$) or PI-2b ($n = 7$), while ST-2 strains had three profiles: PI-1/PI-2a ($n = 6$), PI-1/PI-2b ($n = 1$), and PI-2a only ($n = 1$). ST-23 strains had PI-2a with ($n = 4$) and without PI-1 ($n = 9$).

Nine PI-2a/PI-2b BP gene alleles were identified (Additional file 1: Figure S1) and varied across strains (Figure 4). Strains with PI-2a frequently had *gbs59* alleles 1 ($n = 89$; 30%) or 6 ($n = 32$; 11%) while strains with PI-2b had *san1519* alleles 2 ($n = 69$; 23%) or 3 ($n = 45$; 15%). Little variation was observed in *gbs59* among CC-19 strains and in *san1519* among CC-17, -61, and -67 strains. The remaining CCs were more diverse. CC-1 strains, for example, had five of six *gbs59* alleles.

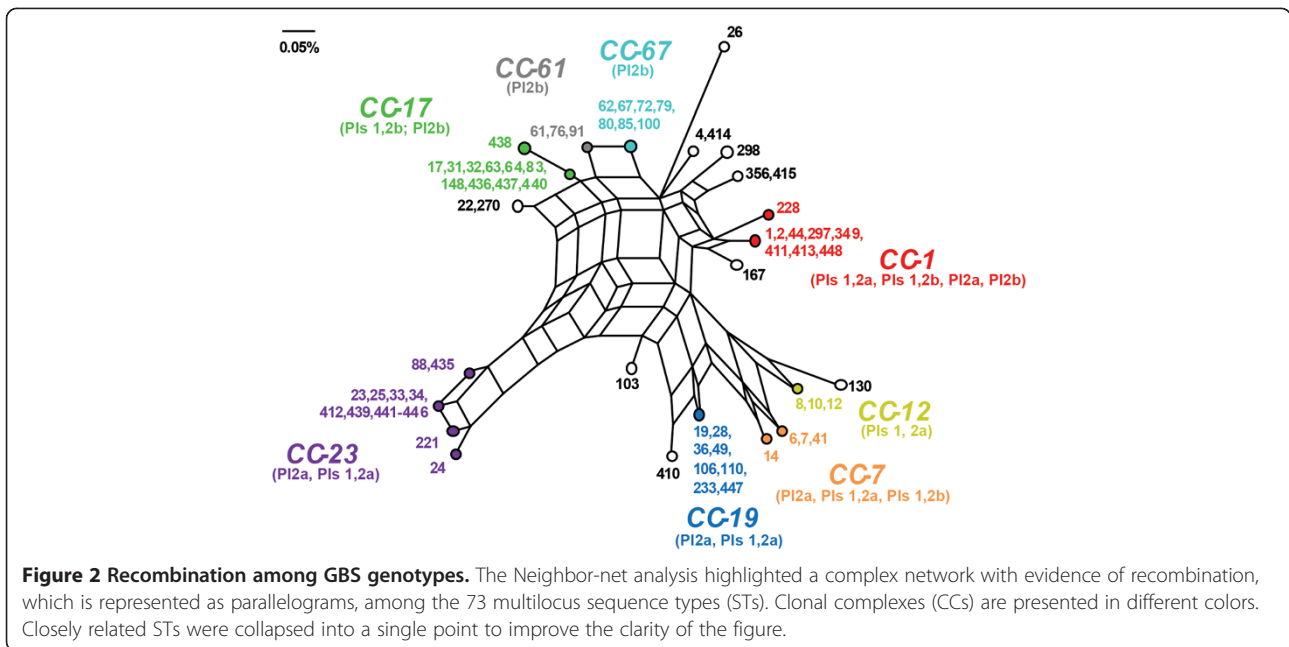


Epidemiological associations and host specificity

Bovine strains were less variable than human strains with respect to the presence of specific PIs. All bovine strains representing the 18 bovine-specific lineages lacked PI-1, though PI-1 was present in six of the seven bovine strains classified as STs 1, 2, 19, and 23 that contain mostly human-derived strains. Among the 45 PI-1-negative bovine strains, the integration site was occupied by a genetic element other than PI-1 in 18 (40%); the site was intact in the remaining 27. Because a subset of these strains had genomes available, the lack of PI-1 was confirmed in 10 of the 18 strains examined. More specifically, none of the PI-1 genes were detected in any of the strains even though *sal_0710* was detected, however, it was split between two separate contigs. Occupancy was not restricted to specific STs (Figure 1) and different strains representing bovine-specific STs 61, 67, 91, and 415 had both occupied and

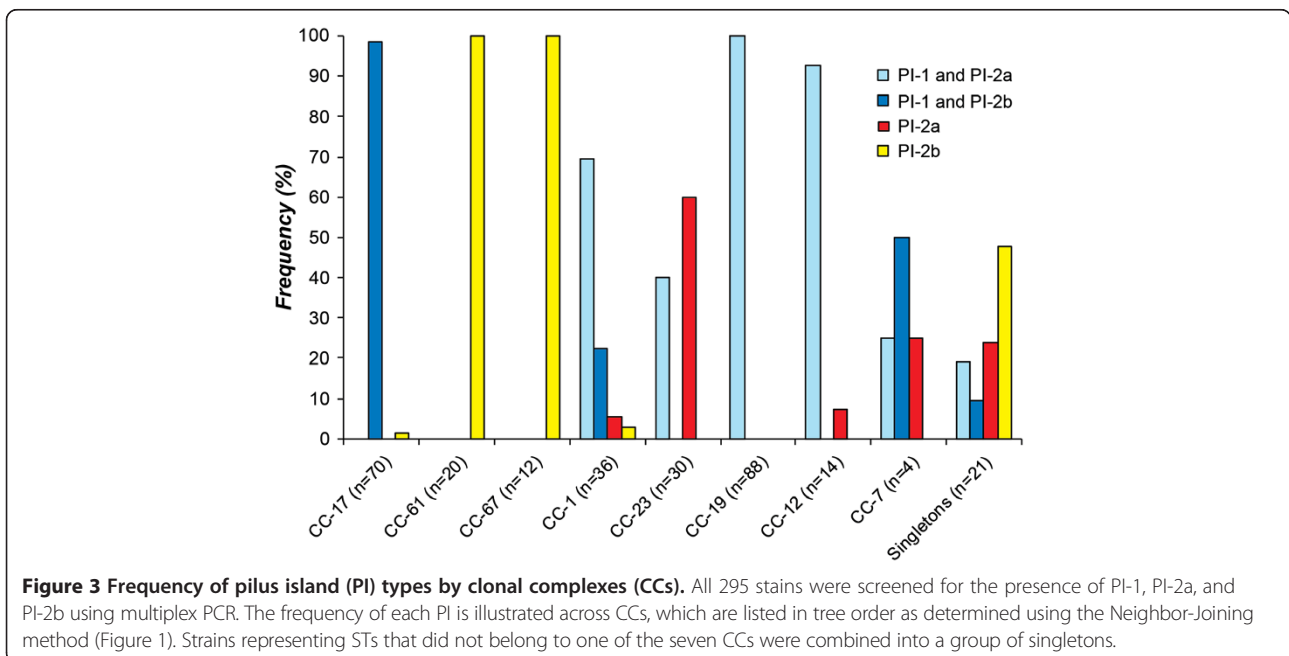
intact sites. All 26 human strains lacking PI-1, however, possessed an intact integration site. The three bovine strains of STs 23, 83 and 297, which lacked PI-1 and clustered with human strains belonging to CCs 23, 17, and 1, also had an intact integration site.

PI frequencies also varied by strain source. Among the 51 bovine strains, only six (12%) had PI-1 compared to 218 (89%) human strains. Indeed, human versus bovine strains were significantly more likely to have PI-1 as well as PI-2a (Table 1). Only seven (14%) of 51 bovine strains had PI-2a versus 163 (67%) of 244 human strains; six of these seven bovine strains also had PI-1. By contrast, the bovine strains were significantly more likely to have PI-2b than human strains and most (86%) possessed PI-2b exclusively. Among the human strains, differences in PI frequencies were observed by source. Invasive neonatal strains, for instance, were significantly more likely



to have PI-1 and one of the two PI-2 variants when compared to the maternal colonizing strains (Table 1). Specifically, 113 (57%) of the 199 strains with two pilus types were recovered from neonates while only 86 (43%) of maternal colonizing strains had both types. Further, the neonatal invasive strains were significantly more likely to have PI-1 with PI-2b than maternal colonizing strains, though the latter had significantly higher frequencies of PI-1 with PI-2a. No difference was observed in the frequency of PI-2a alone across strains.

Stratification by CCs indicated that the 70 CC-17 strains all contained PI-1/PI-2b except for one bovine strain (ST-83) that lacked PI-1. The frequency of strains with PI-1/PI-2b was higher in CC-17 strains relative to all other strains (Fisher's $p < 0.0001$) even after excluding bovine strains. A similar finding was observed for CC-19 strains, which were more likely to possess PI-1/PI-2a relative to all other strains (Fisher's $p < 0.0001$) regardless of *cps* (Additional file 1: Table S3). Among the human strains, however, there was no difference in the PI distribution



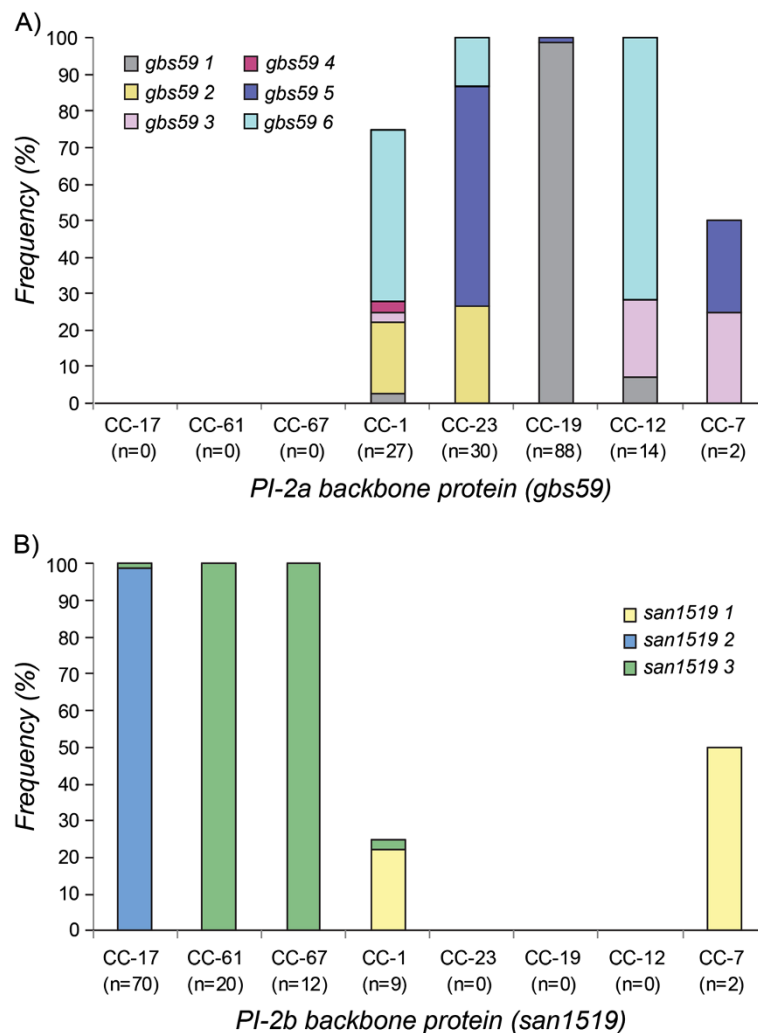


Figure 4 Frequency of pilus island (PI) backbone protein genes by clonal complex (CC). The distribution of **A)** six *gbs59* alleles specific for PI-2a is illustrated in 161 group B streptococcal strains and **B)** three *san1519* alleles specific for PI-2b in 113 strains belonging to the seven CCs. In each figure, the CCs are listed in tree order based on the Neighbor-Joining phylogeny (Figure 1). Singletons (n = 21) were excluded from this analysis.

among neonatal and colonizing strains of CC-17 or CC-19 since virtually all strains from each CC had the same profile even after stratifying by *cps*.

Differences in the allele distribution of the PI BP genes were also observed by source. The 44 bovine strains with PI-2b, for instance, had *san1519* allele 3, whereas only one PI-2b-positive human strain harbored this allele. Human strains more frequently had *san1519* alleles 2 (n = 69; 85%) and 1 (n = 11; 14%). After stratifying *san1519* alleles by source, strains from neonates more frequently had *san1519* allele 2 relative to maternal colonizing strains (Fisher's $p < 0.005$). No differences were observed in the *gbs59* allele distribution between PI-2a-positive human strains associated with asymptomatic colonization and neonatal disease.

PI acquisition and loss

To model PI-1 acquisition and loss, we mapped the distribution of PI-1 on a phylogenetic tree constructed in eBURST that predicts the ancestral genotypes among the predominant CCs. Three groups and three singletons were identified (Figure 5). PI acquisition and loss occurred frequently in human strains during the diversification of closely related genotypes. PI-1 loss was most common in strains of group 1 since four STs derived from a PI-1 and PI-2a-positive ST-1 strain lost PI-1, while PI-1 was maintained in those genotypes derived from ST-19. Similarly, ST-297, which was isolated from a bovine and is derived from ST-17, lacked PI-1 along with the bovine founder (ST-64) of group 2. Notably, some founding genotypes (e.g., STs 1, 23) were comprised of strains with multiple

Table 1 PI distributions among strains isolated from humans and bovines as well as neonates with disease (neonatal invasive) and pregnant women without disease (maternal colonizing)

Pilus island profile	Human-derived (n = 244)		Bovine-derived (n = 51)		Fisher's exact P-value
	n	(%)	n	(%)	
PI-1 and PI-2a (n = 143)	137	(56%)	6	(12%)	<0.00001
PI-1 and PI-2b (n = 81)	81	(33%)	0	(0%)	<0.00001
PI-2a only (n = 27)	26	(11%)	1	(2%)	0.06
PI-2b only (n = 44)	0	(0%)	44	(86%)	<0.00001

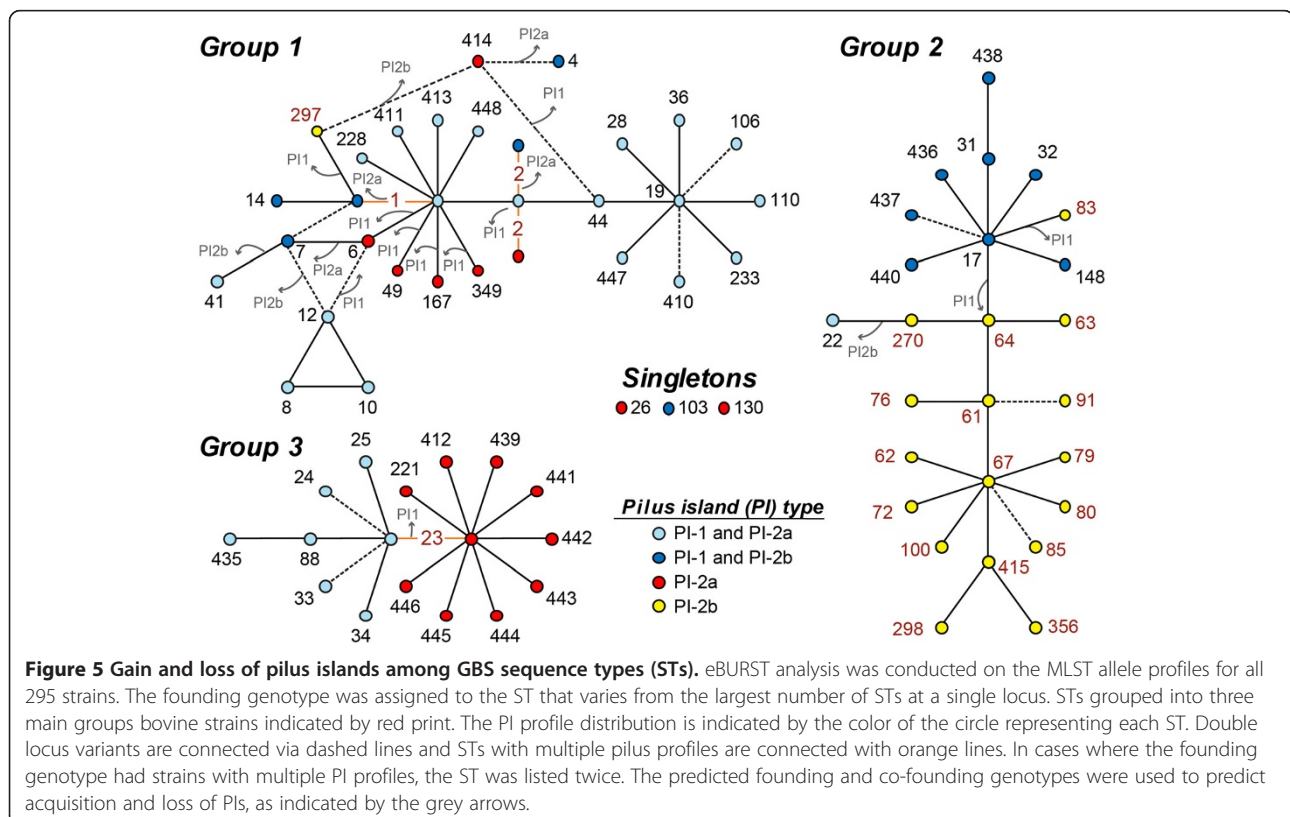
Pilus island profile	Maternal colonizing (n = 99)		Neonatal invasive (n = 120)		Chi square	P-value
	n	(%)	n	(%)		
PI-1 and PI-2a (n = 143)	66	(53%)	59	(47%)	6.8	0.009
PI-1 and PI-2b (n = 81)	20	(27%)	54	(73%)	14.8	0.0001
PI-2a only (n = 27)	13	(65%)	7	(35%)	3.5	0.06
PI-2b only (n = 44)	0	(0%)	0	(0%)	-	-

Note: The colonizing versus neonatal strain analysis excludes 76 strains that did not fall into either of the two categories. Percentages were calculated using the column as the denominator for the top half and row for the bottom half and frequencies were compared using the Likelihood Ratio Chi square (χ^2) and Fisher's Exact Test.

PI profiles. ST-1 strains, for instance, appear to have diversified into STs with four different PI profiles through the acquisition and loss of PI-1 as well as the exchange of PI-2a for PI-2b. Derivatives of ST-23 strains, however, have maintained one of two profiles following diversification.

Discussion

As was demonstrated previously, GBS strains from bovines and humans have distinct characteristics that reflect the independent divergence of these two strain populations [7-9,11-13]. The same is true for human-derived strains of different phylogenetic lineages. CC-17 strains,



for example, have unique virulence gene alleles [25,26] and PI profiles [27] relative to other CCs, which is likely important for virulence. This analysis of 295 diverse strains from multiple sources in North America provides additional support for these findings, further highlights the complexity of the GBS strain population, and identifies genetic characteristics correlated with strain origin.

The PI distribution observed in this study differs from distributions reported elsewhere in North America, Europe and South Africa [24,27,28]. This difference is largely due to the inclusion of bovine-derived strains in this study and reflects the impact of isolate selection on population level analyses. Most bovine strains had PI-2b exclusively, a profile that was also observed in bovine strains from other geographic locations [9,10] but only in a few human-derived strains [24,27,28]. The difference in PI frequencies between bovine and human strains suggests that pilus types contribute to host specificity. Indeed, most (88%) bovine strains lacked PI-1 unlike the human strains, which more frequently had PI-1 in combination with one of the PI-2 variants. Since 40% of the 45 bovine strains lacking PI-1 had an occupied integration site, it is likely that PI-1 confers an advantage in the human host and is not necessary for colonization in bovines. Interestingly, a PI-1 deletion mutant was found to reduce internalization by human-derived monocytes despite having no effect on attachment to A549 lung epithelial cells, VK2 vaginal cells, or ME180 cervical cells in a prior study [29]. It is therefore possible that PI-1 serves primarily to protect against human-derived phagocytic cells while other adherence factors are more important for GBS colonization of the genitourinary tract. Within bovine strains, PI-1 may represent a metabolic burden to the bacterium and be more susceptible to excision or may lack an accessible integration site that prevents PI-1 incorporation into the genome. BLAST results on the consensus sequences from the occupied integration site in two of the PI-1-negative bovine genomes (ANPW00000000 and ANPS01000000), for example, detected several genes from *Streptococcus dysgalactiae subsp. equisimilis*. A future comprehensive comparative genomics study, however, would be needed to better understand the level of diversity within this integration site in strains with and without PI-1.

A relationship was also observed between PI-1 and phylogenetic lineage. All strains of CCs 17 and 19, the complexes previously linked to neonatal infections [2-4], had PI-1 in combination with another PI-2 variant; most (95%) were serotype III. In human strains of all other lineages, however, many (27%) lacked PI-1 altogether suggesting that it is more important for colonization and disease progression in certain genetic backgrounds. As we have observed the same degree of diversity in many other GBS surface proteins [25,26], it is possible that

individual strains utilize different adherence mechanisms to colonize the host. Further stratification by the type of PI-2 variant demonstrated that 98% of neonatal CC-17 strains had PI-1 with PI-2b; none of the strains with this PI profile from other lineages originated from neonates, suggesting that PI-2b may be important for neonatal disease. Interestingly, all 53 *cpsIII* CC-17 strains contained *san1519* allele 2 encoding the PI-2b BP, the major component of the pilus structure [24], also suggesting a specific role for this allele in neonatal disease. Although the diversity of *san1519* is low, the allelic distribution varied among human and bovine strains with the latter exclusively carrying allele 3. Outside of CC-17, PI-1/2b-positive strains of CC-1 had *san1519* allele 1 and represented rare *cps* types (e.g., IV, VII, and VIII). The extensive genetic diversity seen across CCs reflects the independent divergence of these strain populations and highlights features that may influence host specificity and pathogenic potential. Additional studies are needed, however, to examine whether strains of different lineages and PI profiles have an enhanced ability to colonize and/or invade human epithelial cells. It would also be worthwhile to compare PI distributions among strains associated with uncomplicated infections such as urinary tract infections and wound infections since a prior study identified different STs to be associated with these types of infections [30].

Unlike *san1519*, the PI-2a BP gene, *gbs59*, was diverse in strains of lineages previously associated with maternal colonization (e.g., CC-1 and CC-23). Presumably, diversity within PI-2a enhances versatility and enhances the ability to colonize multiple hosts and niches. Support for this hypothesis comes from the reportedly high frequencies of CCs 1 and 23 in asymptomatic women [5] as well as their isolation from bovines [7,8,31] and other animal species [32,33]. As antigenic variation is important for evasion of host immune responses, the high level of diversity in *gbs59* may be the result of strong selective pressures encountered within different hosts. The presence of identical alleles among unrelated strains (Figure 4) also suggests that *gbs59* is a "hot spot" for recombination, while low sequence variability in *san1519* of PI-2b is evidence of a more constrained evolutionary history.

Because there is a clear correlation between phylogenetic lineage and PI profile, both vertical inheritance and horizontal gene transfer have likely contributed to the PI distribution observed. While the PI-2 variants are likely heritable, the presence of PI-1 is less predictable. For the 30 CC-23 strains examined, PI-1 was present in 12 (40%), which is considerably higher than the frequency detected in CC-23 strains from Spain [27], suggesting that there is considerable geographic variation in PI profiles. Such variation may be due to baseline frequencies of PI-1 in specific populations as it may be more susceptible to horizontal gene transfer, a plausible hypothesis since the island

is flanked by direct repeats and contains transposable elements [15]. The absence of PI-1 in CCs unrelated to CC-23 and in specific STs within CC-23 provides additional support for this hypothesis. Following horizontal gene transfer, PI-1 may remain incorporated into the chromosome in some strains, thereby resulting in an increased fitness and colonization potential. Alternatively, it may also be excised from others, which may be due to both host-specific pressures and bacterial stress responses. Indeed, increased horizontal gene transfer and mutation rates have been documented in other pathogens following exposure to certain stressors [34]. Because the GBS PIs are highly immunogenic [14,24], the loss of PI-1 could also provide a mechanism to evade the host immune responses, a process that could be advantageous to certain genotypes that are more prone to cause invasive disease or after exposure to new niche.

The eBURST analysis demonstrated that the neonatal invasive lineage, ST-17, is related to the ST-67 bovine lineage and suggests that PI-1 was either acquired in the ST-17 strain population or lost in the ST-67 bovine population. Although a close relationship was previously identified between STs 17 and 67 [7], it is important to note that eBURST results are greatly impacted by the number and type of STs included in any given analysis. More recent data of all STs available in the PubMLST database [35] suggest that ST-17 is part of eBURST group 1 with STs 19 and 1, which has subsequently diversified into several host-specific complexes including one containing ST-67 and other bovine-associated STs [33]. Further, it was suggested that the ST-17 subpopulation emerged via a series of evolutionary events including recombination among strains belonging to multiple clonal complexes [9] (Figure 2) as well as the acquisition of mobile genetic elements. This hypothesis is supported by our finding that many of the bovine strains were related to human strains containing PI-1 (e.g., ST 83 and 64, Figure 5) or had a PI-1 integration site occupied by another genetic element (e.g., STs 61, 64 and 67, Figure 5) unlike the human-derived strains. Those bovine strains with an occupied integration site may not be capable of acquiring PI-1, which may limit their ability to be transmitted to and sustained in the human host. Collectively, these data suggest that the human vs. bovine strains more frequently serve as eligible recipients for PI-1, and many human strains have lost PI-1 during diversification.

Although the factors that contributed to the emergence of GBS in human populations are not fully understood, acquisition of PI-1 through horizontal gene transfer may have facilitated this process. PI-1 likely increased the fitness and colonization potential of some strains within the human host, thereby allowing them to establish a niche within a pregnant mother, for instance, and enhancing the

likelihood of an opportunistic infection and subsequent transmission to a susceptible neonate. Additional studies, however, are required to test whether strains with different STs and PI profiles vary in their ability to colonize, persist, and invade host tissues relevant to the disease process. In the meantime, enhancing our understanding of PI distribution patterns and genetic diversity in strains from different sources and geographic locations is critical for future efforts aimed at the development of pilus-based GBS vaccines, which were effective in neonatal mice [24,27]. The variable presence of PI-1 among human strains and the possibility of PI-1 loss *in vivo* may limit protection elicited through a vaccine targeting PI-1 alone. Consequently, enhancing our understanding of PI distribution patterns and genetic diversity in strains from different sources and geographic locations is critical for future efforts aimed at the development of pilus-based GBS vaccines, which were effective in neonatal mice [24,27]. The variable presence of PI-1 among human strains and the possibility of PI-1 loss *in vivo* may limit protection elicited through a vaccine targeting PI-1 alone.

Conclusions

The analysis of 295 isolates from diverse sources demonstrated significant variation in the distribution of PI types across phylogenetic lineages and sources, suggesting that pilus combinations impact host specificity and disease outcomes. Moreover, we observed that diversification of specific GBS lineages within certain populations can involve the loss or acquisition of PIs. The variable presence of specific PIs has considerable implications for the development of GBS vaccines targeting these pili.

Methods

Bacterial population

A total of 295 bacterial isolates were included in the study. Most isolates were originally recovered from neonatal blood or cerebral spinal fluid (invasive isolates; $n = 120$) [36] and vaginal/rectal swabs of pregnant women (maternal colonizing isolates; $n = 89$) [37]. Approval to collect specimens was granted by the University of Calgary Ethics Board; informed consent was obtained prior to sample collection. Approval to characterize the de-identified bacterial isolates was provided by both the University of Calgary Ethics Board and Michigan State University Institutional Review Board.

Isolates were characterized by multilocus sequence typing to group isolates in to sequence types (STs) and clonal complexes (CCs). All CC-17 and -19 strains from 98 (out of 192) newborns [2] and 43 (out of 194) pregnant women [5] were included to compare PI frequencies among isolates from the same time and location. Fifty-one isolates from bovines with clinical or subclinical mastitis [7,8] were included to compare PI distributions to

human-derived isolates as was a reference set of 80 human-derived GBS strains of varying STs and serotypes [26]. Cultures were grown in Todd-Hewitt broth at 37°C with 5% CO₂ and capsule (*cps*) types were determined for a subset of strains as described [38].

Phylogenetic analysis

Seven housekeeping genes commonly used for MLST [3] were sequenced and a Neighbor joining phylogeny [39] with 1,000 bootstrap replications was constructed in MEGA5 [40]. Groups of three or more STs with >80% bootstrap support or that were defined in prior studies were considered to represent the CCs; all were originally uncovered by BURST [3]. Recombination was examined in SplitsTree4 [41], while eBURSTv3 [42] was used to identify ancestral genotypes and map PI acquisition and loss.

GBS PIs distribution and variation

PCR amplification of genes encoding sortase C [*sag647*, *sag1406* and *san1517*] and *adhP* was performed (Additional file 1: Table S1) and PI frequencies were compared by source and ST. For PI amplification by PCR, 2 mM dNTP was added to 25 mM MgCl₂, 10 mM primers, 10X buffer II, 1.5 U AmpliTaq Gold (Applied Biosystems), 15 ng/μl DNA and ddH₂O in a 25 μl final volume. Thermocycling conditions utilized an initial soak of 94°C for 10 min, followed by 35 cycles of: 92°C for 1 min, 53°C for 1 min, and 72°C for 30 sec; and a final step of 72°C for 5 min. Strains lacking PI-1 were screened using primers targeting *sal_0710*, which represents the integration site from GBS genome strain 515 (NZ_AAJP01000027) as described by Martins et al. [43]. Amplification of a 684 bp fragment indicated an intact integration site and no amplification indicated occupancy by a genetic element other than PI-1 [43]. The latter was confirmed by examining the occupied region in 12 published genomes, which included a subset of the PI-1-negative bovine strains examined as part of this study. The genomes included the following: ANPS01000000, ANPW00000000, ANQA00000000, ANPU00000000, ANPT00000000, ANPX01000000, ANPY00000000, ANQF00000000, ANCM00000000, ANPZ00000000, ANCK00000000, and ANCO00000000. BLAST was used to search for the ten known PI-1 genes, *sag0642-sag0651*, within the region along with the PI-1 integration site.

Variation within the BP gene of PI-1 was not examined as only 19 of 9,594 nucleotides varied across the six genome strains; however, *in silico* analysis of a subset of genomes [44-46] was performed to identify restriction enzymes (*PvuII* and *SspI*) capable of differentiating PI-2a and PI-2b BP genes, *gbs59* and *san1519* (Additional file 1: Table S2). For amplification of *gbs59*, PCR was performed in a 25 μl reaction with 10 mM of primers and LA *Taq*

(Takara Bio, Inc.) using the following conditions: initial soak at 94°C for 1 min, followed by 30 cycles of 94°C for 30 sec, 54°C for 30 sec, and 68°C for 3 min, and a final soak at 72°C for 10 min. Amplification of *san1519* used the same cycling conditions with a higher annealing temperature (55°C) and shorter extension time (1.5 min). *gbs59* was digested with *PvuII* (New England BioLabs, Inc.), while *SspI* (New England BioLabs, Inc.) was used for *san1519*.

Additional file

Additional file 1: Table S1. Comparison of pilus island type distributions among strains by group B streptococcal clonal complex (CC) and capsule (*cps*) type. **Table S2.** Pilus island (PI) multiplex PCR with gene targets, primer sequences, and expected size fragments. PCR targeting *sag647* (PI-1), *sag1406* (PI-2a), and *san1517* (PI-2b) was used to determine which PIs were present, while PCR-based restriction fragment length polymorphism (RFLP) analysis was used to amplify the PI-2 variant backbone protein (BP) genes, *gbs59* (PI-2a) and *san1519* (PI-2b). **Table S3.** PCR-based RFLP for backbone protein (BP) genes of pilus island (PI)-2a and PI-2b. Digestion of the PI-2a BP gene, *gbs59*, with *PvuII* yielded six major alleles, while *SspI* digestion of the PI-2b BP gene, *san1519*, yielded three alleles. The representative GenBank reference sequences for each variant are listed along with the average size of the expected fragments based on *in silico* analyses. **Figure S1.** Allelic variation in the backbone protein (BP) genes of the pilus island (PI) 2 variants. A) Neighbor-joining phylogeny of the PI-2a BP gene, *gbs59*, based on an *in silico* analysis of 23 published sequences available in GenBank. Six major alleles were identified with 1,273 differences in 2,163 nucleotides and sorted into two groups: group 1 contains alleles 1, 2, and 3, and group 2 contains alleles 4, 5, and 6. Bootstrap values based on 1000 replications are indicated at the nodes. B) Neighbor-joining phylogeny of three alleles of the PI-2b BP gene, *san1519*, based on an *in silico* analysis of three published sequences. *san1519* alleles 1 and 2 differ at 199 of 4,317 nucleotides, whereas alleles 2 and 3 differ at 54 sites. Strain FSL 53-026, indicated in red, represents a bovine strain.

Competing interests

The authors declare no competing interests.

Authors' contributions

ACS, SDM, HDD designed the study; ACS, EAW, SLW, PS performed the work and interpreted molecular and genomic data; ACS, DWL developed molecular assays; ACS, DWL, RNZ, HDD, SDM analyzed epidemiological and evolutionary data and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This paper is dedicated to Cody Springman, who worked so hard on this project and passed away just prior to publication. We thank Jacob Sinkoff and Cassandra Martin for technical support, Drs. Nicola Jones and Martin Wiedmann for providing the bovine strains, and the late Dr. Thomas S. Whittam for his guidance and support. This study was supported by the National Institutes of Health [grant number AI066081] and the Global Alliance to Prevent Prematurity and Stillbirth (GAPPS).

Author details

¹Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA. ²Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, MD, USA. ³Institute for Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK. ⁴Moredun Research Institute, Pentlands, UK. ⁵University of Nebraska Medical School, Omaha, NE, USA.

Received: 11 February 2014 Accepted: 13 June 2014
Published: 19 June 2014

References

1. Edwards MS, Baker CJ: **Group B streptococcal infections in elderly adults.** *Clin Infect Dis* 2005, **41**(6):839–847.
2. Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, Davies HD: **Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada.** *J Clin Microbiol* 2009, **47**(4):1143–1148.
3. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, Kunst F, Glaser P, Rusniok C, Crook DW, Harding RM, Bisharat N, Spratt BG: **Multilocus sequence typing system for group B *Streptococcus*.** *J Clin Microbiol* 2003, **41**(6):2530–2536.
4. Luan SL, Granlund M, Sellin M, Lagergard T, Spratt BG, Norgren M: **Multilocus sequence typing of Swedish invasive group B *Streptococcus* isolates indicates a neonatally associated genetic lineage and capsule switching.** *J Clin Microbiol* 2005, **43**(8):3727–3733.
5. Manning SD, Schaeffer KE, Springman AC, Lehotzky E, Lewis MA, Ouellette LM, Wu G, Moorer GM, Whittam TS, Davies HD: **Genetic diversity and antimicrobial resistance in group B *Streptococcus* colonizing young, nonpregnant women.** *Clin Infect Dis* 2008, **47**(3):388–390.
6. Salloum M, van der Mee-Marquet N, Valentin-Domelier AS, Quentin R: **Diversity of prophage DNA regions of *Streptococcus agalactiae* clonal lineages from adults and neonates with invasive infectious disease.** *PLoS One* 2011, **6**(5):e20256.
7. Bisharat N, Crook DW, Leigh J, Harding RM, Ward PN, Coffey TJ, Maiden MC, Peto T, Jones N: **Hyperinvasive neonatal group B *Streptococcus* has arisen from a bovine ancestor.** *J Clin Microbiol* 2004, **42**(5):2161–2167.
8. Sukhnanand S, Dogan B, Ayodele MO, Zadoks RN, Craver MP, Dumas NB, Schukken YH, Boor KJ, Wiedmann M: **Molecular subtyping and characterization of bovine and human *Streptococcus agalactiae* isolates.** *J Clin Microbiol* 2005, **43**(3):1177–1186.
9. Sorensen UB, Poulsen K, Ghezze C, Margarit I, Kilian M: **Emergence and global dissemination of host-specific *Streptococcus agalactiae* clones.** *mBio* 2010, **1**:3.
10. Yang Y, Liu Y, Ding Y, Yi L, Ma Z, Fan H, Lu C: **Molecular characterization of *Streptococcus agalactiae* isolated from bovine mastitis in Eastern China.** *PLoS One* 2013, **8**(7):e67755.
11. Bohnsack JF, Whiting AA, Martinez G, Jones N, Adderson EE, Detrick S, Blaschke-Bonkowski AJ, Bisharat N, Gottschalk M: **Serotype III *Streptococcus agalactiae* from bovine milk and human neonatal infections.** *Emerg Infect Dis* 2004, **10**(8):1412–1419.
12. Dogan B, Schukken YH, Santisteban C, Boor KJ: **Distribution of serotypes and antimicrobial resistance genes among *Streptococcus agalactiae* isolates from bovine and human hosts.** *J Clin Microbiol* 2005, **43**(12):5899–5906.
13. Richards VP, Lang P, Bitar PD, Lefebvre T, Schukken YH, Zadoks RN, Stanhope MJ: **Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*.** *Infect Gen Evol* 2011, **11**(6):1263–1275.
14. Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, Rosini R, Taddei AR, Mora M, Rappuoli R, Grandi G, Telford JL: **Genome analysis reveals pili in group B *Streptococcus*.** *Science* 2005, **309**(5731):105.
15. Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, Maione D, Taddei A, Santi I, Ghezze C, Brettoni C, Buccato S, Margarit I, Grandi G, Telford JL: **Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*.** *Mol Microbiol* 2006, **61**(1):126–141.
16. Dramsi S, Caliot E, Bonne I, Guadagnini S, Prevost MC, Kojadinovic M, Lalioui L, Poyart C, Trieu-Cuot P: **Assembly and role of pili in group B streptococci.** *Mol Microbiol* 2006, **60**(6):1401–1413.
17. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G: **Pili in gram-positive pathogens.** *Nat Rev Microbiol* 2006, **4**(7):509–519.
18. Pezzicoli A, Santi I, Lauer P, Rosini R, Rinaudo D, Grandi G, Telford JL, Soriani M: **Pilus backbone contributes to group B *Streptococcus* paracellular translocation through epithelial cells.** *J Infect Dis* 2008, **198**(6):890–898.
19. Maisey HC, Hensler M, Nizet V, Doran KS: **Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells.** *J Bacteriol* 2007, **189**(4):1464–1467.
20. Krishnan V, Gaspar AH, Ye N, Mandlik A, Ton-That H, Narayana SV: **An IgG-like domain in the minor pilin GBS52 of *Streptococcus agalactiae* mediates lung epithelial cell adhesion.** *Structure* 2007, **15**(8):893–903.
21. Rinaudo CD, Rosini R, Galeotti CL, Berti F, Necchi F, Reguzzi V, Ghezze C, Telford JL, Grandi G, Maione D: **Specific involvement of pilus type 2a in biofilm formation in group B *Streptococcus*.** *PLoS One* 2010, **5**(2):e9216.
22. Chattopadhyay Y, Mairey E, Mallet A, Dumenil G, Caliot E, Trieu-Cuot P, Dramsi S: **Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*.** *PLoS Pathog* 2009, **5**(5):e1000422.
23. Chattopadhyay D, Carey AJ, Caliot E, Webb RI, Layton JR, Wang Y, Bohnsack JF, Adderson EE, Ulett GC: **Phylogenetic lineage and pilus protein Spb1/SAN1518 affect opsonin-independent phagocytosis and intracellular survival of group B *Streptococcus*.** *Microbes Infect* 2011, **13**(4):369–382.
24. Margarit I, Rinaudo CD, Galeotti CL, Maione D, Ghezze 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: **Preventing bacterial infections with pilus-based vaccines: the group B *Streptococcus* paradigm.** *J Infect Dis* 2009, **199**(1):108–115.
25. Brochet M, Couve E, Zouine M, Vallaes T, Rusniok C, Lamy MC, Buchrieser C, Trieu-Cuot P, Kunst F, Poyart C, Glaser P: **Genomic diversity and evolution within the species *Streptococcus agalactiae*.** *Microbes Infect* 2006, **8**(5):1227–1243.
26. Springman AC, Lacher DW, Wu G, Milton N, Whittam TS, Davies HD, Manning SD: **Selection, recombination, and virulence gene diversity among group B streptococcal genotypes.** *J Bacteriol* 2009, **191**(17):5419–5427.
27. Martins ER, Andreu A, Melo-Cristino J, Ramirez M: **Distribution of pilus islands in *Streptococcus agalactiae* that cause human infections: insights into evolution and implication for vaccine development.** *Clin Vaccine Immunol* 2013, **20**(2):313–316.
28. Madzivhandila M, Adrian PV, Cutland CL, Kuwanda L, Madhi SA, PoPS Trial Team: **Distribution of pilus islands of group B *Streptococcus* associated with maternal colonization and invasive disease in South Africa.** *J Med Microbiol* 2013, **62**(Pt 2):249–253.
29. Jiang S, Park SE, Yadav P, Paoletti LC, Wessels MR: **Regulation and function of pilus island 1 in group B *Streptococcus*.** *J Bacteriol* 2012, **194**(10):2479–2490.
30. van der Mee-Marquet N, Fourny L, Arnault L, Domelier AS, Salloum M, Lartigue MF, Quentin R: **Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites.** *J Clin Microbiol* 2008, **46**(9):2906–2911.
31. Manning SD, Springman AC, Million AD, Milton NR, McNamara SE, Somsel PA, Bartlett P, Davies HD: **Association of group B *Streptococcus* colonization and bovine exposure: a prospective cross-sectional cohort study.** *PLoS One* 2010, **5**(1):e8795.
32. Bishop EJ, Shilton C, Benedict S, Kong F, Gilbert GL, Gal D, Godoy D, Spratt BG, Currie BJ: **Necrotizing fasciitis in captive juvenile *Crocodylus porosus* caused by *Streptococcus agalactiae*: an outbreak and review of the animal and human literature.** *Epidemiol Infect* 2007, **135**(8):1248–1255.
33. Delannoy CM, Crumlish M, Fontaine MC, Pollock J, Foster G, Dagleish MP, Turnbull JF, Zadoks RN: **Human *Streptococcus agalactiae* strains in aquatic mammals and fish.** *BMC Microbiol* 2013, **13**:41.
34. Foster PL: **Stress-induced mutagenesis in bacteria.** *Crit Rev Biochem Mol Biol* 2007, **42**(5):373–397.
35. Jolley KA, Chan M-S, Maiden MC: **mlstDB Net-distributed multi-locus sequence typing (MLST) database.** *BMC Bioinformatics* 2004, **5**(86):1–8.
36. Davies HD, Raj S, Adair C, Robinson J, McGeer A: **Population-based active surveillance for neonatal group B streptococcal infections in Alberta, Canada: implications for vaccine formulation.** *Pediatr Infect Dis J* 2001, **20**(9):879–884.
37. Davies HD, Adair C, McGeer A, Ma D, Robertson S, Mucenski M, Kowalsky L, Tyrell G, Baker CJ: **Antibodies to capsular polysaccharides of group B *Streptococcus* in pregnant Canadian women: Relationship to colonization status and infection in the neonate.** *J Infect Dis* 2001, **184**(3):285–291.
38. Manning SD, Lacher DW, Davies HD, Foxman B, Whittam TS: **DNA polymorphism and molecular subtyping of the capsular gene cluster of group B *Streptococcus*.** *J Clin Microbiol* 2005, **43**(12):6113–6116.
39. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**(4):406–425.
40. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**(8):1596–1599.

41. Huson DH, Bryant D: **Application of phylogenetic networks in evolutionary studies.** *Mol Biol Evol* 2006, **23**(2):254–267.
42. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG: **eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data.** *J Bacteriol* 2004, **186**(5):1518–1530.
43. Martins ER, Melo-Cristino J, Ramirez M: **Evidence for rare capsular switching in *Streptococcus agalactiae*.** *J Bacteriol* 2010, **192**(5):1361–1369.
44. Glaser P, Rusniok C, Buchrieser C, Chevalier F, Frangeul L, Msadek T, Zouine M, Couve E, Lalioui L, Poyart C, Trieu-Cuot P, Kunst F: **Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease.** *Mol Microbiol* 2002, **45**(6):1499–1513.
45. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, *et al*: **Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”.** *Proc Natl Acad Sci U S A* 2005, **102**(39):13950–13955.
46. Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D, Fedorova NB, Scanlan D, Khouri H, Mulligan S, Carty HA, Cline RT, Van Aken SE, Gill J, Scarselli M, *et al*: **Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*.** *Proc Natl Acad Sci U S A* 2002, **99**(19):12391–12396.

doi:10.1186/1471-2180-14-159

Cite this article as: Springman *et al*: Pilus distribution among lineages of group b streptococcus: an evolutionary and clinical perspective. *BMC Microbiology* 2014 **14**:159.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

