Emergence and Global Dissemination of Host-Specific *Streptococcus agalactiae* Clones

Uffe B. Skov Sørensen,[a] Knud Poulsen,[a] Claudia Ghezzo,[b] Immaculada Margarit,[b] and Mogens Kilian[a]

Department of Medical Microbiology and Immunology, Aarhus University, Aarhus C, Denmark,[a] and Novartis Vaccines and Diagnostics, Siena, Italy[b]

ABSTRACT To examine the global diversity of *Streptococcus agalactiae* (group B streptococci [GBS]) and to elucidate the evolutionary processes that determine its population genetics structure and the reported changes in host tropism and infection epidemiology, we examined a collection of 238 bovine and human isolates from nine countries on five continents. Phylogenetic analysis based on the sequences of 15 housekeeping genes combined with patterns of virulence-associated traits identified a genetically heterogeneous core population from which virulent lineages occasionally emerge as a result of recombination affecting major segments of the genome. Such lineages, like clonal complex 17 (CC17) and two distinct clusters of CC23, are exclusively adapted to either humans or cattle and successfully spread globally. The recent emergence and expansion of the human-associated and highly virulent sequence type 17 (ST17) could conceivably account, in part, for the increased prevalence of neonatal GBS infections after 1960. The composite structure of the *S. agalactiae* genome invalidates phylogenetic inferences exclusively based on multilocus sequence typing (MLST) data and thereby the previously reported conclusion that the human-associated CC17 emerged from the bovine-associated CC67.

IMPORTANCE Group B streptococci (GBS) (*Streptococcus agalactiae*) have long been recognized as important causes of mastitis in cattle. After 1960, GBS also became the most prevalent cause of invasive and often fatal infections in newborns. At the same time, GBS are carried by a substantial proportion of healthy individuals. The aims of this study were to elucidate the genetic mechanisms that lead to diversification of the GBS population and to examine the relationship between virulence and host preference of evolutionary lineages of GBS. Genetic analysis of GBS isolates from worldwide sources demonstrated epidemic clones adapted specifically to either the human or bovine host. Such clones seem to emerge from a genetically heterogeneous core population as a result of recombination affecting major segments of the genome. Emergence and global spread of certain clones explain, in part, the change in epidemiology of GBS disease and may have implications for prevention.

Address correspondence to Mogens Kilian, kilian@microbiology.au.dk.

C attle and humans are the main reservoirs for the pyogenic *Streptococcus agalactiae*, which only occasionally has been isolated from other animal species (1). In 1933 to 1938, Rebecca Lancefield described *S. agalactiae* isolates from bovine sources as group B streptococci (GBS) and divided the group into a number of serotypes; the number of serotypes was later increased to a total of 10. GBS are still documented as one of the major causes of subclinical mastitis in dairy cattle in some countries (2, 3). Lancefield also recognized GBS as a human pathogen, but before 1960, reports on the clinical significance of this bacterium for humans were only sporadic (4). After 1960, GBS became the leading cause of invasive neonatal infections in developed countries (5, 6). The reason for this change in human disease incidence, which does not appear to be associated with increased attention or improved diagnostic techniques, remains unknown (4, 7). GBS rarely cause infections in healthy adults; however, occasionally it may cause morbidity in the elderly, in pregnant women, and in patients with underlying predisposing conditions (8).

S. agalactiae is also a common member of the human intestinal and vaginal microbiota. The reported carriage rates among both women and men vary between 10 and 36% (5, 8), and the cumulative carriage rate over a 1-year period may exceed 50% (9). Newborns of mothers who carry GBS are often colonized during delivery (10).

In a pioneering population genetics study of GBS performed by multilocus enzyme electrophoresis (MLEE), Musser et al. (11) demonstrated two evolutionary clusters among serotype III GBS isolates from diseased and asymptomatic neonates in the United States. These two clusters, one of which (electrophoretic type 1 [ET 1]) was significantly more associated with invasive disease than the other, were subsequently confirmed by multilocus sequence typing (MLST) and assigned to clonal complex 17 (CC17) and CC19, respectively (12). The results of several studies support the hypothesis that CC17 consists of hypervirulent GBS adapted to the human host (13–15). It has been speculated that GBS isolates of bovine and human origins constitute separate populations (14, 16, 17). Accordingly, no close relatedness between bovine and human serotype III GBS isolates was observed in some studies (18,

19), while others suggested that the neonatal invasive CC17 has arisen from a bovine ancestor (14, 20).

Recently, Brochet and coworkers analyzed the genomes of eight human isolates of GBS and interestingly demonstrated that these genomes had been shaped by conjugal transfer of large DNA segments (21). According to the MLST database, a total of 482 sequence types (STs) of GBS that constitute a limited number of complexes were recognized as of February 2010 (http://pubmlst.org/sagalactiae/), but information on isolates of bovine origin is largely missing.

The goals of the genetic analyses presented in this paper were to examine the diversity of the GBS population of both human and bovine origin at the global level, to elucidate the evolutionary processes that determine its population structure and the observed differences in virulence potential, and to examine characteristics and patterns of host tropism. We show that no simple monophyletic model for the evolution of GBS exists. Rather, our results support the conclusion that the genome of GBS is shaped by recombination involving large genome segments and indicate that some of the clones may have emerged recently and disseminated globally.

RESULTS

Initial characterization and sorting of isolates. Of the 267 GBS isolates obtained from 14 laboratories in nine countries on five continents, 128 were from human patients and healthy carriers and 139 were isolates from cattle with mastitis. All human isolates were considered epidemiologically independent. Initial sequencing and analysis of the genes encoding initiation factor B (*infB*), superoxide dismutase A (*sodA*), and C5a peptidase (*scpB*) showed that bovine isolates obtained by sampling different quarters on the same animal and isolates obtained from the same farm were identical and considered to be doublets. Therefore, 29 of the 139 bovine isolates were excluded from further analysis. The sources and origins of the remaining 238 isolates are listed in Table S1 in the supplemental material.

Initial cluster analysis of the 238 isolates based on sequences of the two housekeeping genes *infB* and *sodA*, previously shown to be useful markers of GBS lineages (18, 19, 22), revealed 12 clusters of two or more strains (provisionally labeled clusters A to L) and 10 singletons (Fig. 1). This preliminary analysis identified a strong host association of isolates belonging to most clusters. Thus, 41 out of 44 isolates belonging to the division consisting of clusters H to L and two associated singletons were from cattle, whereas 35 out of 36 isolates in cluster E were of human origin. In contrast, the major cluster of 117 isolates (cluster A) included a mixture of bovine and human isolates.

To obtain further information on the degree of genetic and phenotypic heterogeneity of the individual lineages identified by the preliminary cluster analysis, we examined the 238 isolates for molecular capsule type, presence of pilus islands (pilus island 1 [PI-1], PI-2a, or PI-2b), presence and sequence of the *scpB* gene, presence or absence of the group II intron GBSi1 (GBS intron 1) and of the β-antigen/β-protein gene (*bac*), and selected phenotypic traits. The results summarized in Table S2 in the supplemental material show that clusters that were homogeneous with regard to host association also showed a high degree of homogeneity in genetic and phenotypic traits. For example, all 36 isolates of the human-associated group E had *cps* type III and harbored PI-2b and the group II intron GBSi1 but lacked the *bac* gene and the
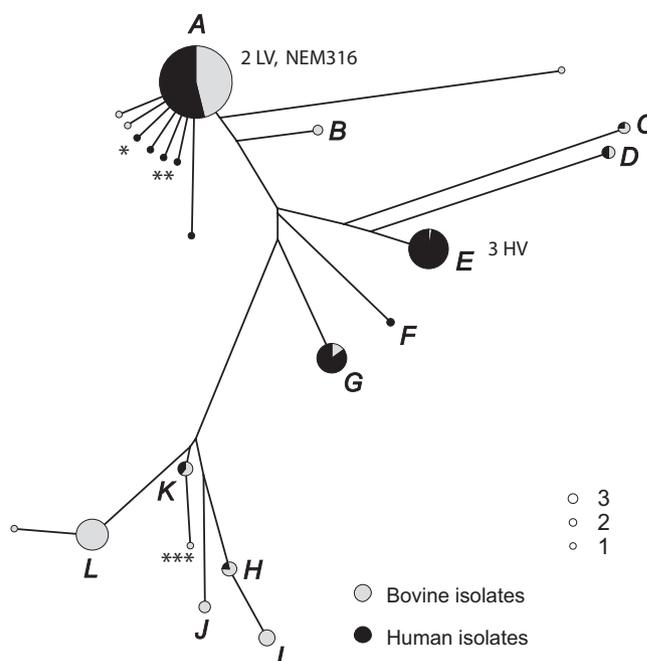
FIG 1 Preliminary clustering of bovine and human GBS isolates. The tree was constructed using the minimum evolution criterion. The analysis was based on concatemers generated from sequences of internal fragments of two genes, *infB* and *sodA*. A total of 238 GBS isolates were included in the analysis; of these, 228 were assigned to groups (groups A to L) of two or more isolates, and 10 were singletons (indicated by the smallest circles). The size of the circle reflects the number of isolates in the group. The sizes of groups consisting of one isolate (singleton), two isolates, and three isolates are shown in the symbol key. HV and LV indicate the locations of five serotype III isolates representing high-virulence and low-virulence clones (three HV clones and two LV clones), respectively, described by Musser et al. (11). Summaries of characteristics for isolates assigned to the 12 groups are listed in Table S2 in the supplemental material. The 55 representative isolates were selected within the groups (52 isolates) and among "outsiders" (three isolates marked by asterisks).

ability to ferment lactose. In contrast, isolates of the other predominantly human cluster, cluster G, had *cps* type Ia and carried PI-2a but lacked PI-1. The 44 isolates of bovine clusters H to L lacked the *scpB* gene (except for three human isolates) and the *bac* gene, carried PI-2b (two of the three exceptional human isolates carried PI-2a) and the group II intron GBSi1, and, with few exceptions, fermented lactose. Capsule gene types varied among the individual bovine clusters H to L. The heterogeneity of cluster A was confirmed by both genetic and phenotypic results. Notably, *cps* types Ib and IV and the *bac* gene encoding β-antigen were exclusively associated with cluster A.

As exemplified by the difference between cluster E and the group of clusters H to L mentioned above, the presence of the *scpB* gene was strongly associated with isolates of human origin. This gene, which encodes C5a peptidase, was detected in 125 of the 128 human isolates (98%) and only in 48 of the 110 bovine strains (44%). Sequencing of the 173 *scpB* genes that were detected by Southern blot analysis revealed five distinct alleles, of which allele A, found in one-third of the human isolates of group E, encodes a protein without C5a peptidase activity but with retained affinity for fibronectin (23). Likewise, the ability to ferment lactose was strongly associated with origin. Of the 110 bovine isolates of the collection, 101 (92%) fermented lactose, in contrast to only 16 (13%) out of 128 isolates of human origin.
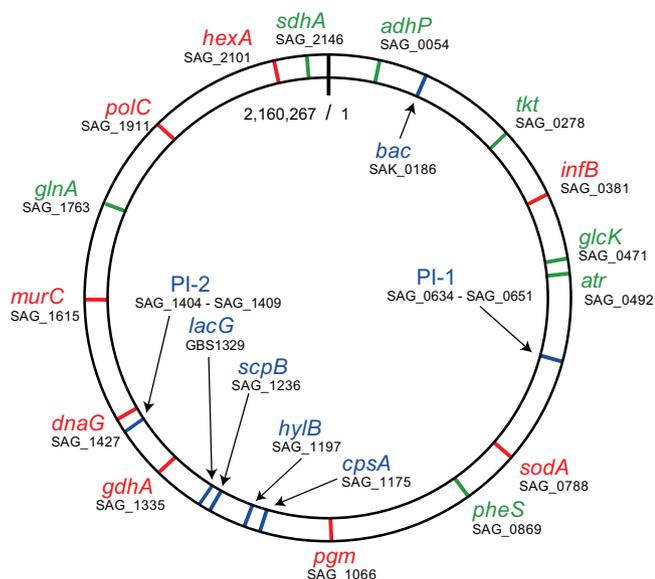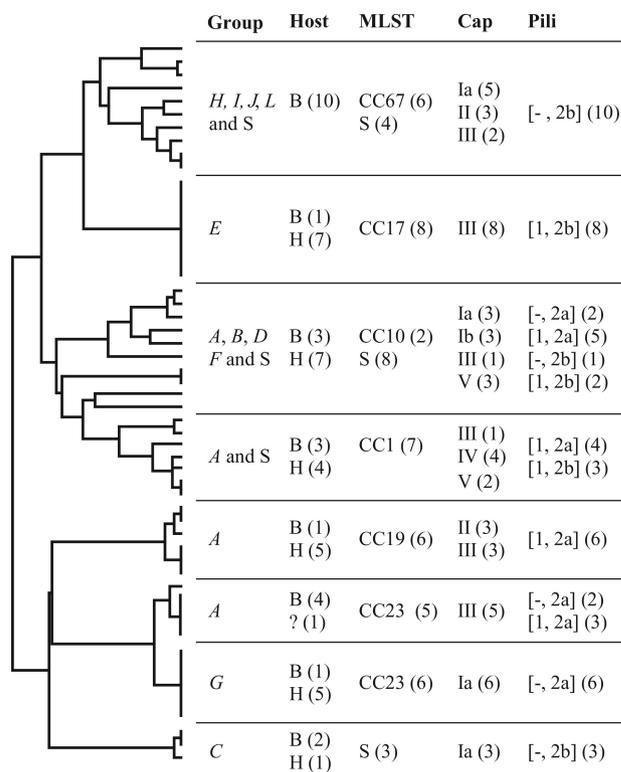
FIG 2 Schematic map of the *S. agalactiae* genome showing the positions of pilus islands (PI-1 and PI-2) and of genes examined in the present study. Genes used in conventional MLST analysis are indicated in green. The eight additional genes are indicated in red. Other genes, which were not included in the clustering analyses, are shown in blue. Locus tags and positions are given for the genes of strain 2603V/R (25), except for the genes missing in this GBS strain: those encoding β-antigen (*bac*) and 6-phospho-beta-galactosidase (*lacG*). The locus tags and positions for these two genes are given for strains A909 and NEM316, respectively. The complete genome sequences of the GBS strains NEM316, 2603V/R, and A909 have been determined (24–26), and all three show conserved gene order (synteny).

FIG 3 Extended characterization of 55 representative *S. agalactiae* isolates. The dendrogram shown to the left was constructed by a clustering analysis using the minimum evolution criterion based on concatemers generated from sequenced internal fragments of 15 genes. The table to the right of the dendrogram summarizes the main results for each cluster. For details, see Fig. S1 in the supplemental material. The table shows the provisional group assignment (Fig. 1; see Fig. S1 and Table S2 in the supplemental material), host, MLST result, *cps* type (Cap), and pilus island distribution. The host is indicated as follows: B, bovine; H, human. The clonal complex (e.g., CC67) and the presence of singletons (S) inferred by conventional MLST are shown in the MLST column. The pilus island distribution (Pili column) is indicated within brackets as follows: 1, presence of PI-1; [minus], absence of PI-1; 2a, presence of PI-2a; 2b, presence of PI-2b. The number of isolates is shown in parentheses in the host, MLST, Cap, and Pili columns.

As shown in Table S2 in the supplemental material, the presence or absence of pilus islands and their allelic variants correlated with the clusters identified. While all 238 isolates carried PI-2, PI-1 was missing in 54% of the bovine isolates (59/110) and in 24% of the human isolates (31/128). Variant PI-2a was exclusively associated with clusters A, F, G, and K and was detected in both human and bovine isolates.

Overall agreement between clusters identified by different approaches. On the basis of the results of the analyses described above, we selected 55 strains as representative of the major lineages of the GBS population for further detailed analysis. This analysis was based on sequences of fragments of 15 housekeeping genes, including *infB* and *sodA*, the seven genes used in the MLST GBS scheme (*adhP*, *atr*, *glcK*, *glnA*, *pheS*, *sdhA*, and *tkt*) (12), and six additional genes (*dnaG*, *gdhA*, *hexA*, *murC*, *pgm*, and *polC*) (for details, see Table S3 in the supplemental material). The six extra genes were chosen on the basis of published genome sequences (24–26) to ensure coverage of different parts of the GBS genome (Fig. 2). The *scpB* gene was not included in this analysis, because the gene was missing in most bovine strains (see Table S2 in the supplemental material).

Cluster analysis based on the 15 genes identified a total of 33 allelic profiles among the 55 strains (Fig. 3). Restricting analysis to the conventional seven MLST loci allowed us to assign ST numbers to these 55 strains. As expected, this analysis resulted in lower resolution and reduced the number of allelic profiles/sequence types (STs) to 24. eBURST analysis assigned 14 out of the 24 STs to recognized clonal complexes, while 10 STs were singletons (see Materials and Methods).

Figure 3 summarizes the information on each of the 55 strains related to a tree generated by analysis of concatemers of sequenced fragments of 15 housekeeping genes (for details, see Fig. S1 in the supplemental material). Apart from the expected increased resolution obtained by adding genes to the analysis, an overall congruence was observed between clusters identified by analysis of 2, 7, or 15 genes. One important example was that the analysis based on 15 genes subdivided sequence type 23 (ST23) into two well-defined clusters comprising human *cps* type Ia isolates and bovine *cps* type III isolates. It is noteworthy that these two clusters were distinguished by both *infB* and *sodA* sequences and by distinct *gdh* alleles (see Fig. S1 in the supplemental material). This confirms the heterogeneity among ST23 strains previously observed by DNA array hybridizations (27) and shows that clinically important GBS populations of different serotypes and host specificity are not necessarily distinguished by the current MLST scheme. The positions of the 24 STs distinguished by MLST analysis in relation to all 472 STs recognized by 1 February 2010 are shown in Fig. S2 in the supplemental material and demonstrate that all major parts of the

GBS population are represented by the 55 strains selected for detailed analysis.

Replacements of large genome segments distort phylogenetic signals. In spite of the overall agreement of clusters identified by analyses based on different numbers of loci, the comparison revealed notable differences in the inferred mutual relationships (see Fig. S3 in the supplemental material). Most notably, the relationships between CC23, CC19, and CC1 inferred by the two dendrograms (based on 7 and 15 genes) differed. To trace phylogenic signals in the actual sequences, we then performed multilocus sequence analysis (MLSA) on concatemers of the 15 gene sequences ordered according to their position in the genome and constituting a total of 6,731 nucleotides. Alignment of the concatenated sequences revealed 150 polymorphic sites (2.2%), 83% of which were phylogenetically informative.

To evaluate the contribution of homologous recombination as a potential explanation for the noncongruent dendrograms, we performed three different analyses. First, analysis of the association between allelic variants in the 15 gene loci showed evidence of significant linkage disequilibrium. One example is the association between *sodA* allele F and *infB* allele A. Out of the 238 isolates in the collection, 117 had both of these alleles (group A, Fig. 1 and Table S2 in the supplemental material), 102 had none of them (groups C, E to J, and L and three singletons), and only 19 isolates possessed one of the two alleles alone (groups B, D, and K and seven singletons). These frequencies are significantly different from the expected frequencies if randomly associated ($P < 0.001$) and suggest a clonal population structure. Second, a neighbor net analysis based on the concatenated sequences depicted a network of numerous splits which, in contrast to the significant linkage disequilibrium, is indicative of recombination, although an overall tree-like structure is maintained (see Fig. S4 in the supplemental material). Third, we further subjected fragments of the concatamers of 15 genes to separate analysis. The concatemers were divided into two halves, i.e., nucleotides 1 to 3269, representing eight gene loci located in the first half of the circular genome (corresponding to positions 72071 through 1076765 in the genome of strain 2603V/R), and nucleotides 3270 to 6731, representing seven gene loci located in the second half of the genome (positions 1343429 through 2129503) (Fig. 4). Neighbor-joining trees generated from the respective concatemers representing the two halves of the genome exhibited pronounced differences in topology (Fig. 4). For example, analysis based on genes located in the second half of the genome showed that CC19 strains were most closely related to CC23 strains (mean genetic distance, 0.0036) and most remote from CC1 (mean genetic distance, 0.0090). In contrast, analysis based on the genes located in the first half of the genome showed the opposite relationships, i.e., CC19 and CC23 were the most distantly related of all clusters (mean genetic distance, 0.0069), whereas CC19 and CC1 merged into one single cluster (mean genetic distance, 0.0020). Thus, although each of the trees was supported by high bootstrap values (Fig. 4), none of them reflected the phylogeny inferred from analysis of sequences sampled from the entire genome (Fig. 3). The lack of congruence of trees based on genes in the two halves of the genome indicates that different segments of the GBS genome evolved independently. A similar lack of congruence was observed when trees based on the seven traditional loci used in MLST and on the 15 loci were compared (see Fig. S3 in the supplemental material).

To analyze the contribution of homologous recombination to the population structure of GBS in further detail, we made an inventory of the polymorphic sites recognized in all 16 gene loci sequenced, including *scpB*. The variable sites for the 55 representative isolates were tabulated sequentially according to their position in the genome (see Fig. S5 in the supplemental material). Obvious examples of polymorphisms due to recent single-nucleotide substitutions within clusters were rare. However, the allelic variation between clusters showed a pronounced mosaic-like pattern. Notably, long, apparently continuous stretches of identical alleles were shared by isolates assigned to different clusters. For example, isolates in CC19 have a large block of alleles identical to a block of alleles in CC1 isolates, followed by a block of alleles with identity to a block in CC17 isolates and a block of alleles with identity to a block in CC23 isolates. This suggests that very large fragments of the genome are involved in the recombinational events. Likewise, ST221, some ST23 isolates (represented by, for example, isolate 29 and 5 other strains), and ST2 isolates of CC1 have a mosaic of shared alleles (see Fig. S5 in the supplemental material) that suggest several independent recombinational events. Together, the results show that successful recombination events affecting large areas of the genome have confounded the relationships between GBS clusters.

DISCUSSION

S. agalactiae is a member of the commensal microbiota of the human intestinal and genitourinary tracts and is also an important pathogen in humans and cattle and is occasionally found in other species. The different spectra of relationships to the respective hosts and of the types of infections caused by GBS raise the question of whether different GBS lineages are adapted to specific hosts and to a commensal or pathogenic life style. Detailed knowledge of the evolutionary processes that may lead to differences in host tropism and virulence may also provide an explanation for the emergence of GBS as a dominant pathogen in neonates after 1960 (28). Our population genetics study, which was designed to answer these questions, demonstrated that recombination involving large fragments of the GBS genome occasionally result in host-specific clones that successfully disseminate globally and cause disease.

When particular genetic variants of bacteria are sequestered in geographically or ecologically isolated subpopulations, association signals that do not reflect the entire population may occur. The majority of reported population genetic studies of GBS were based on relatively small or geographically restricted collections of isolates often obtained from either humans or cattle. In this context, it is noteworthy that the prevalence of some clonal complexes that are considered more virulent varies at different geographic locations (13, 27). In an attempt to overcome this potential bias, we collected both human and bovine isolates from geographically widespread areas of the world and omitted isolates with obvious epidemiological relationships. As in most studies, the overrepresentation of isolates from humans or cattle with infections may have resulted in an incomplete picture of the natural population of GBS. Thus, our data may reflect the more virulent part of the GBS population, and potential relationships between some lineages may have been missed.

At a first glance, the population genetics structure of GBS as reflected in a dendrogram based on 15 gene loci appears clonal with deep branches (Fig. 3). Cluster analysis using different mod-
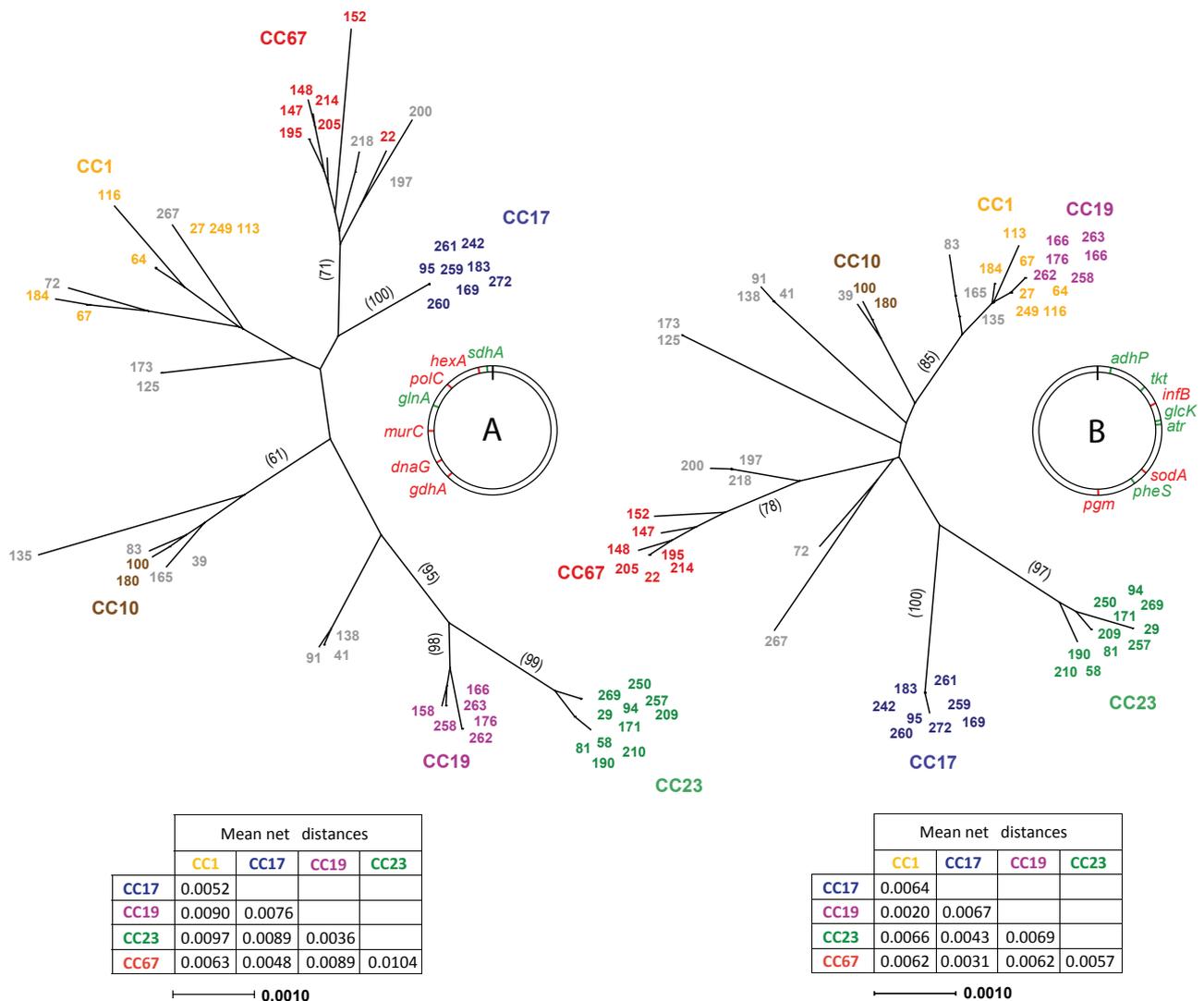
FIG 4 Discordance between topologies of two trees obtained for the same set of GBS isolates based on MLSA schemes, including genes located in different genome sectors. The neighbor-joining tree obtained for the 55 representative GBS isolates using an MLSA scheme based on concatemers of seven genes located in sequential order in the last half of the genome (A) was compared with the tree obtained from the same strains using an MLSA scheme based on concatemers of eight genes located in sequential order in the first half of the genome (B). Relative genome positions of the genes included in the respective MLSA schemes are indicated in the genome diagrams. The strains were assigned to clonal complexes by MLST/eBURST analysis, and strain numbers were color coded accordingly. The average net distances between the clonal complexes are shown in the matrices. Distances were calculated from concatemers based on the two MLSA schemes. The numbers in parentheses are bootstrap values (1,000 replicates).

els unambiguously assigned most of the isolates to six major clusters equivalent to the recognized clonal complexes CC1, CC10, CC17, CC19, CC23, and CC67. Clonality was further supported by strong linkage disequilibrium between particular alleles in independent loci.

Clonal population structures occur when recombination is absent or when the recombination rate is low compared to the mutation rate. In such cases, phylogenetic trees constructed from different portions of the same data set would largely show congruence. Nevertheless, comparative phylogenetic analysis of two halves of the GBS genome based on our data set of 15 housekeeping genes showed strikingly different tree topologies (Fig. 4) in apparent conflict with the linkage disequilibrium of alleles at some loci mentioned above. Relatively few *de novo* mutations were de-

tected, and isolates assigned to distinct clusters contained identical alleles and, in some cases, exhibited long, apparently continuous stretches of identical alleles in several gene loci covering particular genome segments (see Fig. S5 in the supplemental material). This mosaic of long segments in the GBS genome is strong evidence of recombination and provides population-based support for the recent observation by Brochet and coworkers (21) based on *in silico* analysis of eight GBS genome sequences that large conjugal replacements of the chromosome play an important role in the evolution of GBS. The apparent conflict with the observed linkage disequilibrium can be explained by the linked transfer of multiple genes located in the affected genome segments.

On the basis of our findings, we propose that GBS consist of a large and diverse core population, the majority of which act as

commensals in humans and possibly in cattle and certain other animal species. This is reflected in CC1, CC10, and associated singletons that could not be assigned with confidence to any of the six clusters because they combine genetic markers of several of these (see Fig. S1 and Fig. S5 in the supplemental material). This core population of GBS has an almost unlimited pool of genes in agreement with the open pangenome concept of Tettelin et al. (26) and shows lack of correlation between capsular type, genotype, host tropism, and other properties as a result of frequent recombination resulting in the composite genome structure described by Tettelin et al. (26) and Brochet et al. (21, 27). Occasional emerging recombinant clones will possess properties that allow them to successfully disseminate (clonal expansion) and to cause disease in particular hosts. CC17, which includes the highly virulent serotype III defined by Musser et al. (11), is a typical example of a homogeneous "epidemic clone" with rapid global dissemination and successful adaptation to a special habitat, human neonates. In agreement with this scenario, the genetic diversity within the core population is extensive, whereas CC17 and the two fractions of CC23 show very limited diversity, indicating that they emerged recently from the core population (Fig. 3; see Table S2 in the supplemental material). In contrast, the diversity of the cluster constituting CC67, which is adapted to the bovine host, is significantly higher, presumably reflecting a longer evolutionary history. Like CC17, CC67 may have emerged after "en bloc" transfer of major chromosomal sections but, unlike CC17, subsequently diversified through additional recombination events affecting minor genome fragments (see Fig. S5 in the supplemental material). The latter process may also explain the diversification of CC23 into two separate lineages associated with humans and cattle.

Emergence and clonal expansion of epidemic clones, such as ST17, may conceivably account for a part of the striking increase in neonatal morbidity due to GBS after 1960. On the basis of MLST data, it has been proposed that ST17 and ST67 emerged from a common bovine ancestor (14). The alleged close genetic relationship suggested by four out of seven identical MLST gene alleles was supported by the results of hybridization experiments with eight short sequence tags that were identified on the basis of sequences unique to CC17 among human serotype III isolates (19) and on the distribution of mobile genetic elements, such as GBSi1 (20). However, according to our results, bovine CC67 and human CC17 differed in alleles of seven out of eight additional housekeeping genes examined, in the presence of the C5a peptidase gene *scpB*, in the ability to ferment lactose, and in the presence of pilus island PI-1 (see Fig. S1 in the supplemental material). Although both clusters carried GBSi1 in agreement with previous reports (20, 29, 30), this was the case for half of the isolates assigned to other clusters (see Fig. S1 in the supplemental material). Collectively, our data show that CC17 and CC67 are more distantly related than predicted by analysis based on the seven MLST gene loci and do not support the conclusion that the highly virulent human CC17 emerged recently from bovine CC67. Our data rather support the evolutionary scenario that CC17 step by step emerged by recombination between strains belonging to different clonal complexes, one of which may have been CC67. Multiple occasions of conjugation and fragmentation of the genomes followed by homologous recombination conceivably explain the mosaic-like patterns of relationships between strains of individual clonal complexes. This conclusion should be substantiated by future analysis of complete genome sequences.

The findings of this study redefine the significance of several properties usually associated with GBS virulence. All isolates assigned to the successful human pathogenic cluster CC17 carried the *scpB* gene, which is considered a putative virulence factor in *Streptococcus pyogenes* and in GBS. However, 33% of the isolates in CC17 (group E [see Table S2 in the supplemental material]) carried the A allele that encodes a truncated protein without C5a peptidase activity but with retained affinity for fibronectin (23). The fact that isolates in CC17 appear to have enhanced virulence suggests that fibronectin adhesion activity, rather than the ability to cleave C5a, is a pathogenicity factor in neonatal infections. As shown in Table S2 in the supplemental material, the A allele of *scpB* has spread within the GBS population, indicating that possession of this allele may provide a selective advantage. In GBS, the *scpB* gene is located on a composite transposon together with *lmb*, which encodes a laminin-binding surface protein (31). Presumably, this genetic element has been acquired from *S. pyogenes* by horizontal gene transfer. The fact that our collection included three strains isolated in the 1930s that harbor the *scpB* gene shows that acquisition of the C5a peptidase gene by GBS took place long before the significant increase in neonatal GBS infections in the 1960s. Isolates assigned to the bovine cluster CC67 all lacked the *scpB* gene, which indicates that none of the two features of the C5a peptidase mentioned are essential for colonization or infection of cattle (32). This is in agreement with the specific induction of *scpB* by human serum, but not bovine serum (33).

Another striking difference between isolates obtained from humans and cattle is the ability to ferment lactose. The bovine isolates were obtained either from cattle with mastitis or from milk samples. The feature may be one mechanism of the adaptation of GBS to the bovine udder, because lactose constitutes approximately 5% of bovine milk by weight. One notable lactose-fermenting strain was strain NEM316, whose genome has been sequenced.

The β-antigen is a protein associated with the cell wall that binds human IgA and plasma protein factor H (34). The gene encoding this protein was detected in less than 20% of the isolates and was not present among isolates assigned to CC17 or CC67. Accordingly, the β-antigen does not appear to be an essential virulence factor in GBS, or alternatively, the same function is associated with a different protein in some strains.

Recently, pili involved in adhesion to host tissues during colonization were demonstrated in GBS (35). The general presence of PI-2 genes irrespective of host association supports their significance in colonization of both the human and bovine host and underscores their potential as a component of a vaccine against GBS infections as previously suggested (36). PI-1 genes, in contrast, were absent in several lineages.

It has been suggested that diversity in capsular types is achieved mainly through horizontal transfer of genes encoding enzymes and other proteins involved in the synthesis and assembly of saccharide components of the individual capsular types (37) and that capsule switching, as a result, occurs frequently in the GBS population (12, 26, 30, 38). Nevertheless, among the major evolutionary lineages revealed by our population genetics analysis, three lineages, CC17 and the two subclusters within CC23 (Fig. 3), consisted almost exclusively of isolates belonging to single capsular types (i.e., either *cps* type III or Ia) consistent with their general

genetic homogeneity and recent emergence. While this finding supports short-term stability in the capsular type, the multiple types expressed by isolates belonging to more heterogenic and presumably evolutionarily older lineages support the hypothesis that capsule switching is a regular phenomenon within the diverse core of the GBS population (Fig. 3).

There are several important practical consequences of our findings. MLST has been generally accepted as the standard method of genotyping, but the inability of the method to distinguish two clinically separate populations of CC23 indicates that the number of loci selected for the GBS scheme is insufficient in this case. Moreover, the fact that the MLST loci are located primarily in one half of the genome (Fig. 2) combined with the conjugal transfer of unusually large parts of the GBS genome may result in flawed conclusions if the MLST genes are used for phylogenetic analyses. For example, according to the MLST data, CC1 and CC19 are the closest neighbors, whereas a more comprehensive sampling of the genome shows that they are as distant as any pair of complexes in the population (Fig. 3; see Fig. S3 in the supplemental material). Similar problems arose when the MLST data were used for eBURST analysis. Our data illustrate that eBURST analysis based on the seven MLST loci links groups that are otherwise unrelated according to a more comprehensive coverage of the genome. For example, it has been concluded that CC17 may include both serotype III and serotype II isolates (39). However, our analysis of extended allele profiles clearly demonstrates that CC17 is a very homogenous cluster of *cps* type III isolates only.

In conclusion, our study provides new information on the population structure of *S. agalactiae* and on the genetic mechanisms behind the emergence of host-specific clones that successfully spread globally. Conceivably, emergence of the human-associated highly virulent ST17, which is still highly conserved, partly accounts for the increased prevalence of neonatal GBS infections after 1960. The composite structure of the *S. agalactiae* genome invalidates phylogenetic inferences based on MLST data, and the results obtained are at variance with the previously reported conclusion that the human-associated CC17 and the bovine-associated CC67 emerged from a recent common ancestor.

MATERIALS AND METHODS

Bacterial isolates and culture conditions. The strain collection included 267 GBS isolates of bovine and human origin. The strains were requested from 14 recognized laboratories in nine countries (see Table S1 in the supplemental material), and all strains obtained were included in the study without preconditions. The human isolates were recovered from episodes of invasive disease, from vaginal samples obtained from asymptomatic women, and from healthy neonatal carriers. The bovine isolates were from cattle with clinical and subclinical mastitis in dairy herds.

Strain NEM316 (24) (CIP 82.45 = ATCC 12403 = Lancefield serotype III reference strain D136C = our strain 81), whose genome has been sequenced, is usually listed as an isolate from a fatal case of neonatal septicemia. However, its human origin is confirmed neither by Rebecca Lancefield's publications nor by the files at the Lancefield Collection of Streptococcus Strains, Laboratory of Bacterial Pathogenesis and Immunology, Rockefeller University (Vincent Fischetti, personal communication). Apparently, Lancefield received the strain from Leonard Colebrook (original designation, 3 cole 106) (http://www.straininfo.net), and we consider its origin unknown.

Isolates were cultured as described and species designation was confirmed by a combination of standard tests (40).

Phenotypic analyses. Fermentation of lactose and hydrolysis of salicin were examined in phenol red broth base supplemented with meat extract (0.3%) and the respective carbohydrates at a concentration of 1% (w/vol). Test tubes were inoculated, incubated at 37°C, and observed for 7 days. Hyaluronate lyase (hyaluronidase) activity was tested on BHI plates (3.7% brain heart infusion agar and 2% Noble agar [BD Difco, BD Diagnostics, Franklin Lakes, NJ] [pH 6.8] supplemented with 0.02% [wt/vol] hyaluronic acid [Sigma-Aldrich, St. Louis, MO] and 0.1% [wt/vol] bovine serum albumin [Sigma-Aldrich]) (41). Bacteria were point inoculated on the plates, which were then incubated for 18 h at 37°C and "developed" by flooding the surface with 2 N acetic acid. Intact hyaluronic acid forms a turbid precipitate with albumin at low pH, and hyaluronidase activity results in a clear zone surrounding the bacterial growth.

Southern blot analysis. The restriction fragment length polymorphism (RFLP) typing of selected genes was performed as described previously (42). The probe for the C5a peptidase gene (*scpB*) area was a 3.5-kb PCR fragment amplified by the ForW and RevW primers as described previously (42, 43). The probe used for the β-antigen gene (*bac*) was a 1.3-kb PCR fragment of the gene corresponding to positions 112 to 1395 in the open reading frame (ORF) (42, 44).

Sequencing of gene fragments. Multilocus sequence typing (MLST) was based on sequencing of the internal fragments of seven housekeeping genes, namely, *adhP*, *atr*, *glcK*, *glnA*, *pheS*, *sdhA*, and *tkt*, as described previously (12). The seven genes were amplified by PCR using Ready-To-Go PCR beads (Amersham Pharmacia Biotech), and amplicons were sequenced on both strands (for details, see Table S3 in the supplemental material). Assignment to particular alleles and sequence types (STs) were performed at the *S. agalactiae* MLST website (http://pubmlst.org/sagalactiae/). Similarly, internal fragments (450 to 860 bp) of nine additional genes, *dnaG*, *gdhA*, *hexA*, *infB* (22), *murC*, *scpB* (45), *sodA* (46), *pgm*, and *polC*, were amplified and sequenced on both strands with the primers listed in Table S3 in the supplemental material. Sequences were aligned, trimmed, and compared by using the BioEdit version 7.0.5.2 (http://www.mbio.ncsu.edu/BioEdit/) and MEGA4 (http://www.megasoftware.net/mega4) (47) programs. Sequences of novel alleles identified in this study were deposited in the GenBank database (http://www.ncbi.nlm.nih.gov). Designations and accession numbers for the novel alleles are listed in Table S4 in the supplemental material.

MLST and assignment to clonal groups. The most recent allelic profiles, including those of the nine new STs detected in this study, were downloaded from the *S. agalactiae* MLST website and analyzed by the eBURST approach (version 3 [http://eburst.mlst.net]). The isolates were assigned to clonal complexes named according to the inferred founding genotype, i.e., CC17, CC19, CC23, etc., as originally defined by Maynard Smith et al. (48): "Clonal complexes are composed of two parts, a 'consensus group', i.e., a group of bacteria that are identical at all seven loci, and 'single locus variants' (SLVs) which are identical to a consensus group at six loci but which differ at the seventh." Isolates that did not fulfill these criteria were regarded as singletons. Hence, a clonal complex was defined more rigorously than the relaxed definition commonly used for BURST lineages (12), which lately have been considered equivalent to clonal complexes (15).

Molecular capsule typing. Capsular gene types (*cps*) were determined by the PCR and sequencing procedure developed by Kong et al. (49). Capsular gene type-specific sequences were amplified by PCR using eight primer pairs. *cps* types Ia, Ib, and III to VI were identified directly according to amplification of a PCR product with a single type-specific primer set, while *cps* types II and VII were identified by sequencing of the amplicon (49). GBS isolates of *cps* type VIII fail to yield a 790-bp PCR amplicon at the 3′ end of *cpsE-cpsF* and 5′ end of *cpsG* (49). As the method described by Kong et al. is unable to differentiate types V and IX, isolates positive for *cps* type V in the PCR were tested in a PCR specific for serotype IX (50).

Pilus typing. Oligonucleotide primers for amplification of the pilus-encoding genes representing pilus islands PI-1, PI-2a, and PI-2b (51) were designed on the basis of the genome sequences of isolates 2603V/R (25)

and COH1 (26). The primers are listed in Table S5 in the supplemental material. PCRs were performed by Pwo *Taq* polymerase as recommended by the manufacturer (Roche). The following amplification conditions were used: (i) 5 min at 94°C; (ii) 25 cycles, with 1 cycle consisting of 30 s at 94°C, 50 s at 50°C, and 1 min at 72°C for 1 min; (iii) extension step of 7 min at 72°C. Each reaction mixture contained the following: 2 μl (10 ng/μl) genomic DNA, 50 pmol of each oligonucleotide, 2.5 μl of buffer plus MgSO$_4$ (10×), 4 μl deoxynucleoside triphosphates (dNTPs) (10 mM each), 0.5 μl Pwo polymerase, and H$_2$O up to 25 μl.

Detection of the group II intron GBSi1. PCR was performed to detect the presence of GBSi1 among the isolates as described previously (52). Ready-To-Go PCR beads and primer spafo1 combined with primer sparev2 were used in a PCR to amplify a 1.4-kb product from the group II intron GBSi1.

Phylogenetic analysis. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 4 (47). Bootstrap analyses were based on 1,000 replicates. Furthermore, multivariate tables were made from the allelic profiles, and distance matrices were generated from the tables by using the online program START2 (Sequence Type Analysis and Recombinational Tests 2) (version 0.5.14 [http://pubmlst.org/software/analysis/start2]). Alleles designated by letters (*scpB* and *sodA* genes) were replaced with the corresponding allele numbers before the data were inserted in the START2 program. Neighbor net networks were generated by using the SplitsTree program version 4.2 (http://www.splitstree.org) (53).

Statistical analysis. For 2 × 2 contingency tables, probabilities (two sided) were calculated by the Fisher exact test with the InStat program (version 3.06 for Windows; GraphPad Software, Inc., San Diego, CA).

ACKNOWLEDGMENTS

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00178-10/-/DCSupplemental

FIG S1, EPS file, 1.817 MB.
FIG S2, EPS file, 1.097 MB.
FIG S3, EPS file, 1.049 MB.
FIG S4, EPS file, 0.951 MB.
FIG S5, EPS file, 1.744 MB.
TABLE S1, DOC file, 0.071 MB.
TABLE S2, DOC file, 0.84 MB.
TABLE S3, DOC file, 0.133 MB.
TABLE S4, DOC file, 0.038 MB.

TABLE S5, DOC file, 0.040 MB.

REFERENCES

1. **Elliott, J. A., R. R. Facklam, and C. B. Richter.** 1990. Whole-cell protein patterns of nonhemolytic group B, type Ib, streptococci isolated from humans, mice, cattle, frogs, and fish. J. Clin. Microbiol. **28**:628–630.
2. **Keefe, G. P.** 1997. *Streptococcus agalactiae* mastitis: a review. Can. Vet. J. **38**:429–437.
3. **Merl, K., A. Abdulmawjood, C. Lämmler, and M. Zschock.** 2003. Determination of epidemiological relationships of *Streptococcus agalactiae* isolated from bovine mastitis. FEMS Microbiol. Lett. **226**:87–92.
4. **Anthony, B. F., and D. M. Okada.** 1977. The emergence of group B streptococci in infections of the newborn infant. Annu. Rev. Med. **28**:355–369.
5. **Trijbels-Smeulders, M. A., L. A. Kollee, A. H. Adriaanse, J. L. Kimpen, and L. J. Gerards.** 2004. Neonatal group B streptococcal infection: incidence and strategies for prevention in Europe. Pediatr. Infect. Dis. J. **23**:172–173.
6. **Schuchat, A.** 1999. Group B streptococcus. Lancet **353**:51–56.
7. **Ross, P. W.** 1984. Group-B streptococcus—profile of an organism. J. Med. Microbiol. **18**:139–166.
8. **Schuchat, A.** 1998. Epidemiology of group B streptococcal disease in the United States: shifting paradigms. Clin. Microbiol. Rev. **11**:497–513.
9. **Hansen, S. M., N. Uldbjerg, M. Kilian, and U. B. S. Sørensen.** 2004. Dynamics of *Streptococcus agalactiae* colonization in women during and after pregnancy and in their infants. J. Clin. Microbiol. **42**:83–89.
10. **Melchers, W. J. G., J. M. J. E. Bakkers, M. Toonen, van Kuppeveld, F. J. M., M. Trijbels, and J. A. A. Hoogkamp-Korstanje.** 2003. Genetic analysis of *Streptococcus agalactiae* strains isolated from neonates and their mothers. FEMS Immunol. Med. Microbiol. **36**:111–113.
11. **Musser, J. M., S. J. Mattingly, R. Quentin, A. Goudeau, and R. K. Selander.** 1989. Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B Streptococcus) causing invasive neonatal disease. Proc. Natl. Acad. Sci. U. S. A. **86**:4731–4735.
12. **Jones, N., J. F. Bohnsack, S. Takahashi, K. A. Oliver, M. S. Chan, F. Kunst, P. Glaser, C. Rusniok, D. W. Crook, R. M. Harding, N. Bisharat, and B. G. Spratt.** 2003. Multilocus sequence typing system for group B streptococcus. J. Clin. Microbiol. **41**:2530–2536.
13. **Lin, F. Y., A. Whiting, E. Adderson, S. Takahashi, D. M. Dunn, R. Weiss, P. H. Azimi, J. B. Philips III, L. E. Weisman, J. Regan, P. Clark, G. G. Rhoads, C. E. Frasch, J. Troendle, P. Moyer, and J. F. Bohnsack.** 2006. Phylogenetic lineages of invasive and colonizing strains of serotype III group B streptococci from neonates: a multicenter prospective study. J. Clin. Microbiol. **44**:1257–1261.
14. **Bisharat, N., D. W. Crook, J. Leigh, R. M. Harding, P. N. Ward, T. J. Coffey, M. C. Maiden, T. Peto, and N. Jones.** 2004. Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor. J. Clin. Microbiol. **42**:2161–2167.
15. **Jones, N., K. A. Oliver, J. Barry, R. M. Harding, N. Bisharat, B. G. Spratt, T. Peto, and D. W. Crook.** 2006. Enhanced invasiveness of bovine-derived neonatal sequence type 17 group B streptococcus is independent of capsular serotype. Clin. Infect. Dis. **42**:915–924.
16. **Jensen, N. E., and F. M. Aarestrup.** 1996. Epidemiological aspects of group B streptococci of bovine and human origin. Epidemiol. Infect. **117**:417–422.
17. **Martinez, G., J. Harel, R. Higgins, S. Lacouture, D. Daignault, and M. Gottschalk.** 2000. Characterization of *Streptococcus agalactiae* isolates of bovine and human origin by randomly amplified polymorphic DNA analysis. J. Clin. Microbiol. **38**:71–78.
18. **Sukhnanand, S., B. Dogan, M. O. Ayodele, R. N. Zadoks, M. P. Craver, N. B. Dumas, Y. H. Schukken, K. J. Boor, and M. Wiedmann.** 2005. Molecular subtyping and characterization of bovine and human *Streptococcus agalactiae* isolates. J. Clin. Microbiol. **43**:1177–1186.
19. **Bohnsack, J. F., A. A. Whiting, G. Martinez, N. Jones, E. E. Adderson, S. Detrick, A. J. Blaschke-Bonkowsky, N. Bisharat, and M. Gottschalk.** 2004. Serotype III *Streptococcus agalactiae* from bovine milk and human neonatal infections. Emerg. Infect. Dis. **10**:1412–1419.
20. **Hery-Arnaud, G., G. Bruant, P. Lanotte, S. Brun, A. Rosenau, N. Mee-Marquet, R. Quentin, and L. Mereghetti.** 2005. Acquisition of insertion sequences and the GBSi1 intron by *Streptococcus agalactiae* isolates correlates with the evolution of the species. J. Bacteriol. **187**:6248–6252.
21. **Brochet, M., C. Rusniok, E. Couve, S. Dramsi, C. Poyart, P. Trieu-Cuot,**

F. Kunst, and P. Glaser. 2008. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. U. S. A. **105**:15961–15966.

22. Hedegaard, J., M. Hauge, J. Fage-Larsen, K. K. Mortensen, M. Kilian, H. U. Sperling-Petersen, and K. Poulsen. 2000. Investigation of the translation-initiation factor IF2 gene, *infB*, as a tool to study the population structure of *Streptococcus agalactiae*. Microbiology **146**:1661–1670.

23. Tamura, G. S., J. R. Hull, M. D. Oberg, and D. G. Castner. 2006. High-affinity interaction between fibronectin and the group B streptococcal C5a peptidase is unaffected by a naturally occurring four-amino-acid deletion that eliminates peptidase activity. Infect. Immun. **74**:5739–5746.

24. Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couve, L. Lalioui, C. Poyart, P. Trieu-Cuot, and F. Kunst. 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. Mol. Microbiol. **45**:1499–1513.

25. Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi, and C. M. Fraser. 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. U. S. A. **99**:12391–12396.

26. Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. DeBoy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R Rappuoli, and C. M. Fraser. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. U. S. A. **102**:13950–13955.

27. Brochet, M., E. Couve, M. Zouine, T. Vallaeys, C. Rusniok, M. C. Lamy, C. Buchrieser, P. Trieu-Cuot, F. Kunst, C. Poyart, and P. Glaser. 2006. Genomic diversity and evolution within the species *Streptococcus agalactiae*. Microbes Infect. **8**:1227–1243.

28. Johri, A. K., L. C. Paoletti, P. Glaser, M. Dua, P. K. Sharma, G. Grandi, and R. Rappuoli. 2006. Group B Streptococcus: global incidence and vaccine development. Nat. Rev. Microbiol. **4**:932–942.

29. Granlund, M., L. Oberg, M. Sellin, and M. Norgren. 1998. Identification of a novel insertion element, IS1548, in group B streptococci, predominantly in strains causing endocarditis. J. Infect. Dis. **177**:967–976.

30. Luan, S. L., M. Granlund, M. Sellin, T. Lagergard, B. G. Spratt, and M. Norgren. 2005. Multilocus sequence typing of Swedish invasive group B streptococcus isolates indicates a neonatally associated genetic lineage and capsule switching. J. Clin. Microbiol. **43**:3727–3733.

31. Franken, C., G. Haase, C. Brandt, J. Weber-Heynemann, S. Martin, C. Lämmler, A. Podbielski, R. Lutticken, and B. Spellerberg. 2001. Horizontal gene transfer and host specificity of beta-haemolytic streptococci: the role of a putative composite transposon containing scpB and lmb. Mol. Microbiol. **41**:925–935.

32. Lindahl, G., M. Stalhammar-Carlemalm, and T. Areschoug. 2005. Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. Clin. Microbiol. Rev. **18**:102–127.

33. Gleich-Theurer, U., S. Aymanns, G. Hass, S. Mauerer, J. Vogt, and B. Spellerberg. 2009. Human serum induces streptococcal C5a peptidase expression. Infect. Immun. **77**:3817–3825.

34. Areschoug, T. F., M. F. Stalhammar-Carlemalm, I. F. Karlsson, and G. Lindahl. 2002. Streptococcal beta protein has separate binding sites for human factor H and IgA-Fc. J. Biol. Chem. **277**:12642–12648.

35. Lauer, P., C. D. Rinaudo, M. Soriani, I. Margarit, D. Maione, R. Rosini, A. R. Taddei, M. Mora, R. Rappuoli, G. Grandi, and J. L. Telford. 2005. Genome analysis reveals pili in Group B *Streptococcus*. Science **309**:105.

36. Margarit, I., C. D. Rinaudo, C. L. Galeotti, D. Maione, C. Ghezzo, E. Buttazzoni, R. Rosini, Y. Runci, M. Mora, S. Buccato, M. Pagani, E. Tresoldi, A. Berardi, R. Creti, C. J. Baker, J. L. Telford, and G. Grandi. 2009. Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm. J. Infect. Dis. **199**:108–115.

37. Cieslewicz, M. J., D. Chaffin, G. Glusman, D. Kasper, A. Madan, S. Rodrigues, J. Fahey, M. R. Wessels, and C. E. Rubens. 2005. Structural and genetic diversity of group B streptococcus capsular polysaccharides. Infect. Immun. **73**:3096–3103.

38. Davies, H. D., N. Jones, T. S. Whittam, S. Elsayed, N. Bisharat, and C. J. Baker. 2004. Multilocus sequence typing of serotype III group B streptococcus and correlation with pathogenic potential. J. Infect. Dis. **189**:1097–1102.

39. Bohnsack, J. F., A. Whiting, M. Gottschalk, D. M. Dunn, R. Weiss, P. H. Azimi, J. B. Philips III, L. E. Weisman, G. G. Rhoads, and F. Y. Lin. 2008. Population structure of invasive and colonizing strains of *Streptococcus agalactiae* from neonates of six U.S. Academic Centers from 1995 to 1999. J. Clin. Microbiol. **46**:1285–1291.

40. Kilian, M.. 2005. *Streptococcus* and *Lactobacillus*, p. 833–881. *In* S. P. Borriello, P. R. Murray, and G. Funke (ed.), Topley and Wilson's microbiology and microbial infections, 10th ed. Hodder Arnold, London, United Kingdom.

41. Smith, R. F., and N. P. Willett. 1968. Rapid plate method for screening hyaluronidase and chondroitin sulfatase-producing microorganisms. Appl. Microbiol. **16**:1434–1436.

42. Hauge, M., C. Jespersgaard, K. Poulsen, and M. Kilian. 1996. Population structure of *Streptococcus agalactiae* reveals an association between specific evolutionary lineages and putative virulence factors but not disease. Infect. Immun. **64**:919–925.

43. Cleary, P. P., J. Handley, A. N. Suvorov, A. Podbielski, and P. Ferrieri. 1992. Similarity between the group B and A streptococcal C5a peptidase genes. Infect. Immun. **60**:4239–4244.

44. Jerlstrom, P. G., G. S. Chhatwal, and K. N. Timmis. 1991. The IgA-binding beta antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions. Mol. Microbiol. **5**:843–849.

45. Chmouryguina, I., A. Suvorov, P. Ferrieri, and P. P. Cleary. 1996. Conservation of the C5a peptidase genes in group A and B streptococci. Infect. Immun. **64**:2387–2390.

46. Gaillot, O., C. Poyart, P. Berche, and P. Trieu-Cuot. 1997. Molecular characterization and expression analysis of the superoxide dismutase gene from *Streptococcus agalactiae*. Gene **204**:213–218.

47. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596–1599.

48. Smith, J. M., E. J. Feil, and N. H. Smith. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. Bioessays **22**:1115–1122.

49. Kong, F., S. Gowan, D. Martin, G. James, and G. L. Gilbert. 2002. Serotype identification of group B streptococci by PCR and sequencing. J. Clin. Microbiol. **40**:216–226.

50. Kong, F., L. M. Lambertsen, H. C. Slotved, D. Ko, H. Wang, and G. L. Gilbert. 2008. Use of phenotypic and molecular serotype identification methods to characterize previously nonserotypeable group B streptococci. J. Clin. Microbiol. **46**:2745–2750.

51. Rosini, R., C. D. Rinaudo, M. Soriani, P. Lauer, M. Mora, D. Maione, A. Taddei, I. Santi, C. Ghezzo, C. Brettoni, S. Buccato, I. Margarit, G. Grandi, and J. L. Telford. 2006. Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. Mol. Microbiol. **61**:126–141.

52. Granlund, M., F. Michel, and M. Norgren. 2001. Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B streptococci. J. Bacteriol. **183**:2560–2569.

53. Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**:254–267.