

Mobile Genetic Elements Provide Evidence for a Bovine Origin of Clonal Complex 17 of *Streptococcus agalactiae*[∇]

Geneviève Héry-Arnaud,¹ Guillaume Bruant,^{1†} Philippe Lanotte,^{1,2} Stella Brun,¹ Bertrand Picard,³ Agnès Rosenau,¹ Nathalie van der Mee-Marquet,^{1,2} Pascal Rainard,⁴ Roland Quentin,^{1,2} and Laurent Mereghetti^{1,2*}

Université François-Rabelais, IFR 136, Faculté de Médecine, EA 3854 Bactéries et Risque Materno-Foetal, Tours, France¹; Centre Hospitalier Universitaire, Tours, France²; Service de Microbiologie, Hôpital Avicenne, Assistance Publique-Hôpitaux de Paris, Bobigny, France³; and Institut National de la Recherche Agronomique, UR1282, Infectiologie Animale et Santé Publique, Nouzilly, France⁴

Received 8 November 2006/Accepted 12 May 2007

We sought an explanation for epidemiological changes in *Streptococcus agalactiae* infections by investigating the link between ecological niches of the bacterium by determining the prevalence of 11 mobile genetic elements. The prevalence of nine of these elements differed significantly according to the human or bovine origin of the isolate. Correlating this distribution with the phylogeny obtained by multilocus sequence analysis, we observed that human isolates harboring GBSi1, a clear marker of the bovine niche, clustered in clonal complex 17. Our results are thus consistent with the emergence of this virulent human clone from a bovine ancestor.

Epidemiological changes in the pattern of *Streptococcus agalactiae* infections over the last six decades from predominantly bovine infections (13) to a leading cause of human neonatal infections (25, 26) and, more recently, to a cause of invasive disease in elderly adults (6) require explanation. Genomic lineages grouping the most invasive isolates have been described by multilocus enzyme electrophoresis (17, 23), by pulsed-field gel electrophoresis (1, 24), and more recently by the analysis of the polymorphism in the allelic profiles of housekeeping genes (11) but are insufficient to account for the emergence of *S. agalactiae* from such diverse ecosystems. A link between the bovine and human niches has been proposed (8, 19) but remains a matter of debate. Indeed, it was recently suggested that human and bovine *S. agalactiae* isolates are largely unrelated (3) and represent distinct populations or subtypes (5, 27). Conversely, another study provided evidence to suggest that the major hyperinvasive sequence type 17 (ST-17) clone implicated in neonatal invasive diseases had arisen from bovine lineage ST-61 (2), whereas others suggested that the genome of the common ancestor was closer to that of human ST-17 strains than to that of the ST-61 strains of bovine origin (4). We hypothesized that mobile genetic elements (MGEs), such as insertion sequences (ISs) or group II introns, which drive bacterial

evolution (12, 30), might provide insight into changes in the epidemiology of *S. agalactiae* infections.

(This work was presented in part at the XVIth Lancefield International Symposium on Streptococci and Streptococcal Diseases, Cairns, Australia, 25 to 29 September 2005.)

We tested this hypothesis by analyzing a collection of 98 epidemiologically unrelated isolates, mostly collected in France (3 isolates from Germany and 1 isolate from Malaysia) between 1966 and 2003. Of the 98 isolates, 63 human isolates were from colonized individuals, adults with invasive diseases, and neonates with meningitis (52 of which were previously typed by multilocus sequence typing [MLST] and studied for the prevalence of IS1548, IS861, IS1381, ISSa4, and GBSi1) (10) and 35 bovine isolates were obtained from cows with clinical evidence of udder infection.

The prevalence of 11 MGEs (one group II intron [GBSi1] and 10 IS elements corresponding to six ISs [IS1548, IS861, IS1381, ISSa4, ISSag1, and ISSag2] and four transposase genes corresponding to remnants of IS elements [ISSag9, SAG0448, SAG1241, and SAG0610, relating to the identification of the open reading frames in strains A909 and 2605V/R] [28, 29]) was evaluated by PCR. IS1548, IS861, IS1381, ISSa4, ISSag1, ISSag2, and GBSi1 were detected as previously described (10). Other transposase genes were detected with the primers listed in Table 1. The prevalence of all the MGEs except IS861 and ISSag2 differed significantly as a function of the original reservoir (Fig. 1). This was particularly true for IS1381, IS1548, ISSag1, and ISSag2, which were mostly present in human isolates, and for GBSi1 and ISSa4, which were mostly present in bovine isolates (Fig. 1).

Multilocus sequence analysis, carried out with the standard MLST scheme of Jones et al. (11), resolved the 98 isolates into 47 sequence types (STs). A phylogenetic tree was generated,

* Corresponding author. Mailing address: EA 3854 “Bactéries et Risque Materno-Foetal,” Laboratoire de Bactériologie, Faculté de Médecine de Tours, Université François-Rabelais, 2 bd. Tonnellé, 37032 Tours, France. Phone: 33 247478056. Fax: 33 247473812. E-mail: laurent.mereghetti@med.univ-tours.fr.

† Present address: Groupe de Recherche sur les Maladies Infectieuses du Porc, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, Canada.

[∇] Published ahead of print on 25 May 2007.

TABLE 1. Primers used for the amplification of transposase genes from the *S. agalactiae* isolates

Target	GenBank accession no. ^a	Sequence of primer (5'–3')		T_m^b (°C)	Amplicon size (bp)
		Forward	Reverse		
ISSag9	NC_004116	GACCGTAATGACATCCAATCTGG	GGCTTGAAATGGACACGGTT	50	600
SAG0448	NC_004116	CCACAGAATTACTTAACTTCCTAGC	CGACCAACCTTATTGAATCACT	50	994
SAG1241 (<i>orfA</i>)	NC_004116	CTCTTGTTAACCTTTATCAAACCTGG	CCTAAAAAAGCAAGTGCCA	50	218
SAG0610	NC_004116	TACTCCACATTAGCTAAAGAACAGG	GGCCTGATTTAACCGAGATT	50	394

^a NC_004116 refers to strain 2603 V/R (29).

^b T_m , melting temperature.

based on the nucleotide sequences of the supergene obtained by concatenation of the seven loci for the 98 isolates, using MEGA software version 3.1 (www.megasoftware.net/) with the neighbor-joining (NJ) method (14). Clonal complexes (CCs) (isolates sharing six or seven identical alleles) were defined by eBURST analysis (<http://eburst.mlst.net/>) (7). Seven main NJ phylogenetic divisions were identified, closely matching the seven main CCs identified by eBURST analysis (Fig. 2): subclonal complex 19 (subCC19) ($n = 21$), subCC2 ($n = 11$), CC10 ($n = 12$), CC7 ($n = 6$), CC17 ($n = 13$), CC67 ($n = 15$), and CC23 ($n = 10$). Three of these CCs were “host specific”: CC17 and subCC19 consisted almost exclusively of human isolates (100% and 95% of the isolates, respectively), whereas CC67 was almost exclusively composed of bovine isolates (93.8% of the isolates). Moreover, the NJ dendrogram shows that human lineage CC17 arose from bovine lineage CC67 (Fig. 2), consistent with the assertions of Bisharat et al. (2).

A factorial analysis of correspondence was performed on the whole data set, with the 11 MGEs as active variables and the three main clonal complexes (CC19, CC17, and CC67) and the bovine or human origins of the strains as illustrative variables. Computations were performed with SPAD.N software (Centre International de Statistiques et Informatique Appliquées, St. Mandé, France) as previously described (22). The projections of the active and illustrative variables and of the strains on the F1/F2 plane accounting for 40.97% of the total

variance (Fig. 3) separated two groups of variables and of strains. GBSi1 and ISSa4 (and to a lesser degree SAG0448 and SAG1241), bovine origin, CC17 and CC67, and 94.3% of bovine isolates were projected onto the negative values of F1. IS1548, IS1381, ISSag9, and ISSag1 (and to a lesser degree SAG0610), human origin, CC19, and 79.4% of human isolates were projected onto the positive values of F1. These findings confirm those of Luan et al. in that the two major human *S. agalactiae* lineages are each marked by one MGE: subCC19 (defined as CC19 by Luan et al. [16]) is characterized by IS1548, whereas CC17 is characterized by GBSi1. All but one of the human isolates projected with the bovine isolates and variable CC67 onto the negative values of F1 belonged to CC17, highlighting the stronger linkage of CC17 with the bovine lineage CC67 than with any of the other human lineages (Fig. 3). Moreover, our data also suggest that the presence of GBSi1 in the bovine reservoir predates its presence in the human reservoir because of the following. (i) All our bovine isolates harbored GBSi1, whereas this element was present in less than one-third of human isolates (almost all of which belonged to CC17). (ii) The bovine isolates had higher copy numbers and more-complex patterns for GBSi1 than the human isolates did (data not shown), consistent with an increase in IS pattern complexity over time (18). MGEs therefore provide evidence for a bovine ancestor of the human clonal lineage CC17, all the isolates of which harbor GBSi1. This statement is consistent with the pro-

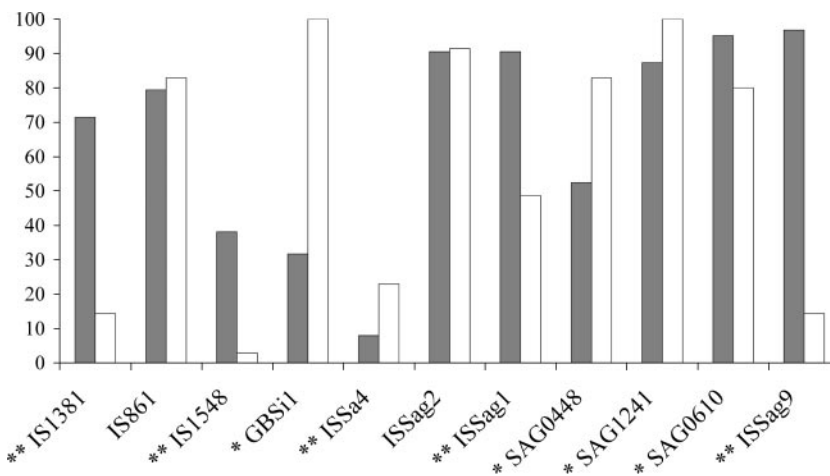
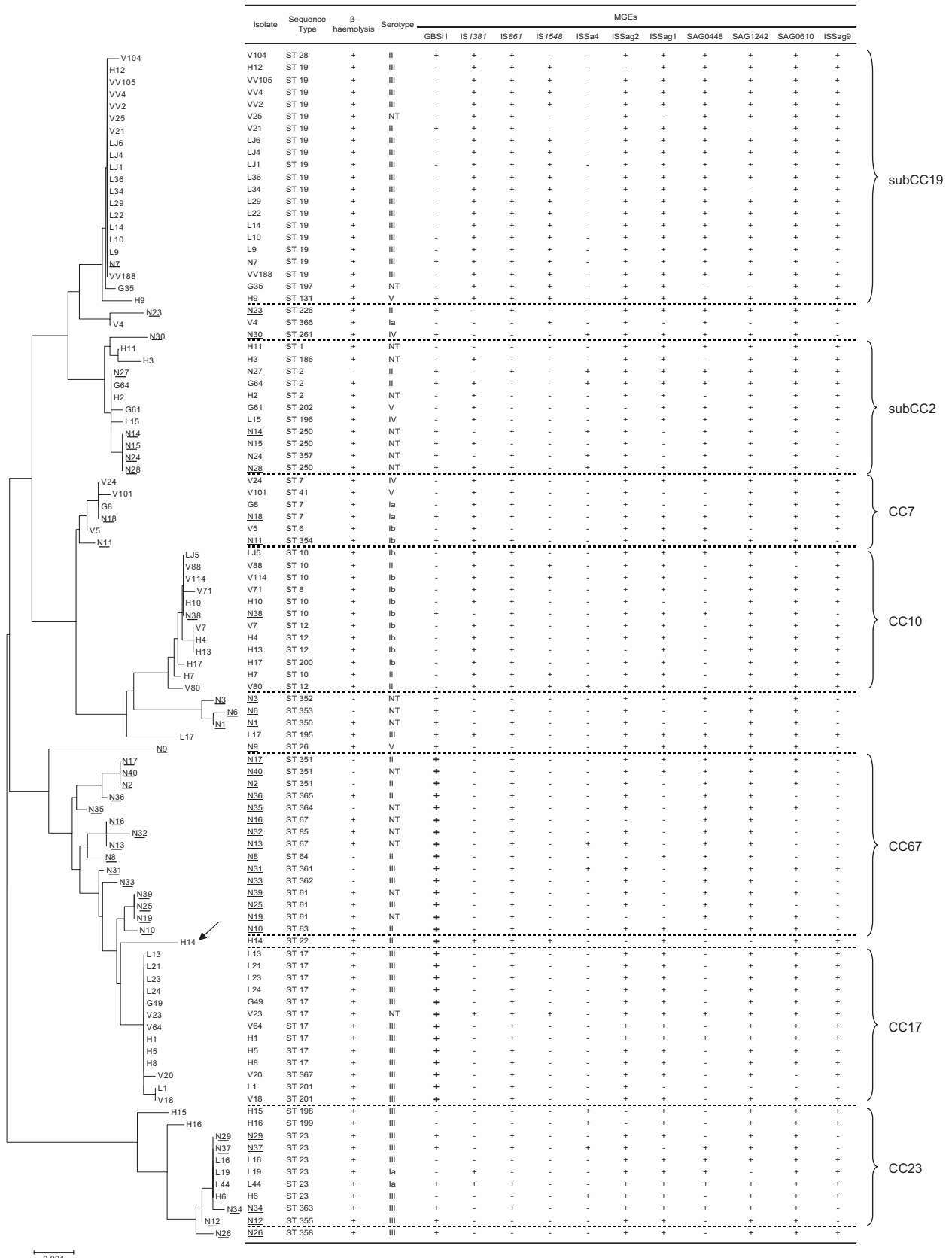


FIG. 1. Prevalence of the MGEs within the human (filled bars) and bovine (open bars) *S. agalactiae* isolates. Significant differences (without correction for bias due to multiple comparisons) are indicated by one asterisk for P values below 0.05 and two asterisks for P values below 0.001.



0.001

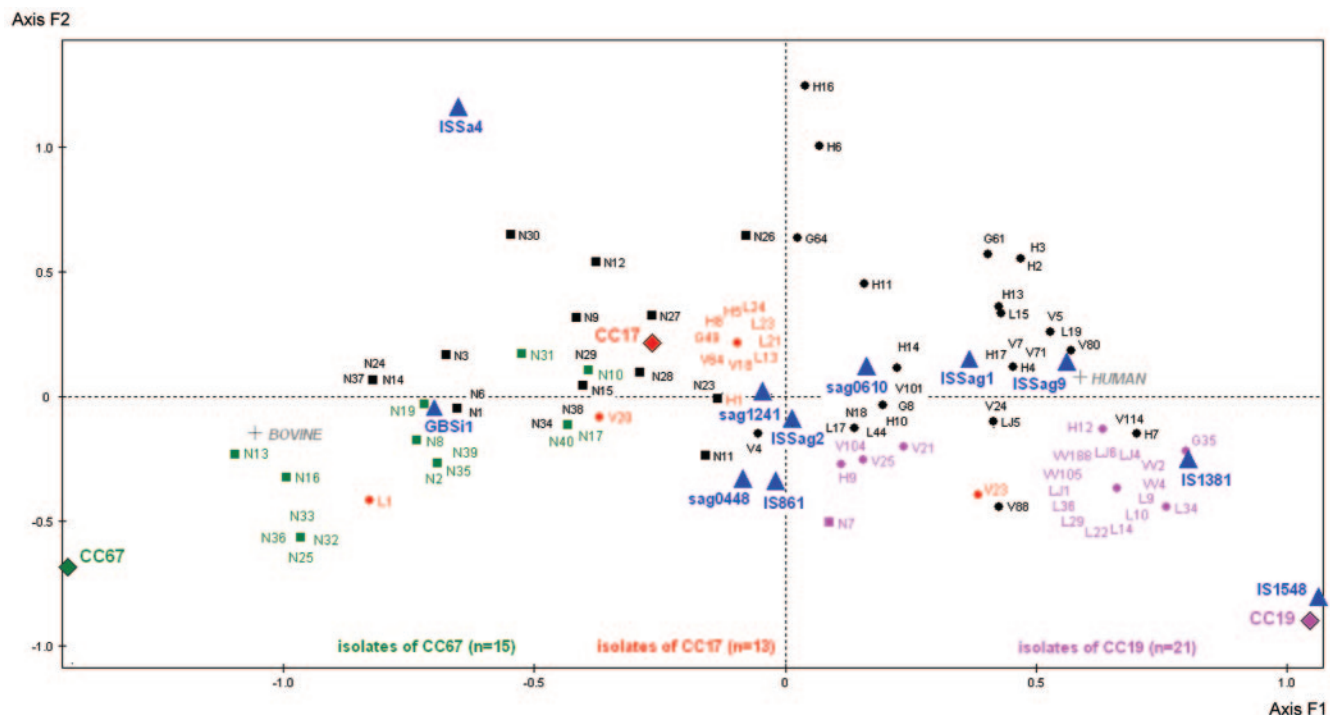


FIG. 3. Factorial analysis of correspondence of 98 *S. agalactiae* isolates based on the MGE data. Eleven active variables (the MGEs [▲] IS1381, IS861, IS1548, GBSi1, ISSa4, ISSag2, ISSag1, SAG0448, SAG1241, SAG0610, and ISSag9), five illustrative variables (the main clonal complexes CC19 [red diamond], CC17 [pink diamond], and CC67 [green diamond] and the human [circles] and bovine [squares] origins of the isolates), and the 98 isolates are projected onto the F1/F2 plane. This plane, obtained by computation, is defined by the two principal axes of the analysis; the first axis, F1, explains most of the variance, and the second axis, F2 (orthogonal to F1), explains most of the variance not explained by F1. Together, F1 and F2 accounted for 40.97% of the total variance. Isolates of CC17 ($n = 13$) are shown in red, isolates of CC67 ($n = 15$) are shown in green, and isolates of subCC19 ($n = 21$) are shown in pink.

posed evolutionary scheme of MGE acquisition, in which GBSi1 is thought to have been acquired recently by human isolates (10).

The particular distribution of MGEs among bovine and human isolates (Fig. 3) is also consistent with a physical barrier reducing the exchange of genetic material between the bovine and human reservoirs. The human isolate H14, which occupies a prominent position on the NJ dendrogram between the bovine CC67 and the human CC17 (Fig. 2), displayed two particular features. First, the copy number of GBSi1 in this isolate was much higher than that in other isolates (human isolates had a median of 3 copies of GBSi1, whereas H14 harbored 12 copies [data not shown]). Parkhill et al. suggested that IS expansion occurs during evolutionary bottlenecks (20, 21). They argued that in *Yersinia pestis*, IS expansion coincided with a change in the ecological niche of the species from a commensal organism of the gut to a systemic pathogen (21). Second, isolate H14 harbored both IS1548

and GBSi1. In our study, as in previous studies (9, 15), very few isolates harbored both GBSi1 and IS1548 (Fig. 2). Granlund et al. explained this mutually exclusive distribution by the presence of a closed insertion site for the two MGEs on the bacterial chromosome (9), but this distribution could also be explained by IS1548 and GBSi1 originating from two different ecological niches. Thus, human clone CC17 seems to have evolved from the bovine lineage containing GBSi1 and has retained the GBSi1 genetic marker from its bovine donor.

In conclusion, our results provide at least a partial explanation of the emergence of *S. agalactiae* as a major cause of neonatal meningitis, despite its original identification as the causal agent of bovine mastitis. Assuming that CC17 evolved from a bovine ancestor, as shown both by MLST analysis and by the presence of the bovine marker GBSi1, the isolates of this CC may have genetic features different from those of other *S. agalactiae* isolates that have adapted to their human hosts

FIG. 2. Relatedness of the 98 isolates of *S. agalactiae* analyzed by the NJ method, using MEGA v3.1 software and the Kimura two-parameter mutation model of genetic distance. The nucleotide differences between the supergene sequences obtained by multilocus sequence analysis method of Jones et al. (11) were used to infer phylogenetic relationships between isolates. We used eBURST v3 software to group the 98 isolates into CCs. For each isolate, the ST, beta-hemolysis, serotype, and presence (+) or absence (-) of the 11 MGEs are indicated. CCs are also indicated on the right of the figure (several isolates were not attributed a CC number by the Start software and are considered "singletons"). The prominent position of isolate H14 is indicated by a black arrow. The human isolates are named L or LJ (meningocerebral fluid), H (adult), V or VV (vaginal carriage), or G (gastric fluid); the bovine isolates are named N (bovine isolates are underlined). NT, nontypeable.

over a longer period of time. These features probably concern genes not present in the core genome, as recently defined by Tettelin et al., but that have been acquired by horizontal exchange due to the environmental gene pool available in other ecosystems (28). Indeed, we recently reported a link between prophage DNA fragments and *S. agalactiae* isolates from neonatal cases of meningitis, suggesting the presence of additional genetic material in these strains (10).

We thank Nicolas Jones, who manages the MLST website database, for assigning new allele numbers and new sequence type numbers. We also thank Stéphanie Gouriou for her help in using the SPAD.N software.

REFERENCES

1. Bidet, P., N. Brahimi, C. Chalas, Y. Aujard, and E. Bingen. 2003. Molecular characterization of serotype III group B streptococcus isolates causing neonatal meningitis. *J. Infect. Dis.* **188**:1132–1137.
2. 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.
3. 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.
4. 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.
5. Dogan, B., Y. H. Schukken, C. Santisteban, and K. J. Boor. 2005. Distribution of serotypes and antimicrobial resistance genes among *Streptococcus agalactiae* isolates from bovine and human hosts. *J. Clin. Microbiol.* **43**:5899–5906.
6. Edwards, M. S., M. A. Rench, D. L. Palazzi, and C. J. Baker. 2005. Group B streptococcal colonization and serotype-specific immunity in healthy elderly persons. *Clin. Infect. Dis.* **40**:352–357.
7. Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* **186**:1518–1530.
8. Finch, L. A., and D. R. Martin. 1984. Human and bovine group B streptococci: two distinct populations. *J. Appl. Bacteriol.* **57**:273–278.
9. 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.
10. Héry-Arnaud, G., G. Bruant, P. Lanotte, S. Brun, A. Rosenau, N. van der 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.
11. 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.
12. Kazazian, H. H., Jr. 2004. Mobile elements: drivers of genome evolution. *Science* **303**:1626–1632.
13. Keefe, G. P., I. R. Dohoo, and E. Spangler. 1997. Herd prevalence and incidence of *Streptococcus agalactiae* in the dairy industry of Prince Edward Island. *J. Dairy Sci.* **80**:464–470.
14. Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* **5**:150–163.
15. Luan, S. L., M. Granlund, and M. Norgren. 2003. An inserted DNA fragment with plasmid features is uniquely associated with the presence of the GBSi1 group II intron in *Streptococcus agalactiae*. *Gene* **312**:305–312.
16. 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.
17. 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. USA* **86**:4731–4735.
18. Naas, T., M. Blot, W. M. Fitch, and W. Arber. 1995. Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. *Mol. Biol. Evol.* **12**:198–207.
19. Norcross, N. L., and N. Oliver. 1976. The distribution and characterization of group B streptococci in New York State. *Cornell Vet.* **66**:242–248.
20. Parkhill, J., M. Sebahia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtmann, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**:32–40.
21. Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebahia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**:523–527.
22. Picard, B., J. S. Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion, and E. Denamur. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* **67**:546–553.
23. Quentin, R., H. Huet, F. S. Wang, P. Geslin, A. Goudeau, and R. K. Selander. 1995. Characterization of *Streptococcus agalactiae* strains by multilocus enzyme genotype and serotype: identification of multiple virulent clone families that cause invasive neonatal disease. *J. Clin. Microbiol.* **33**:2576–2581.
24. Rolland, K., C. Marois, V. Siquier, B. Cattier, and R. Quentin. 1999. Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *hylB* gene analysis. *J. Clin. Microbiol.* **37**:1892–1898.
25. Schrag, S. J. 2004. The past and future of perinatal group B streptococcal disease prevention. *Clin. Infect. Dis.* **39**:1136–1138.
26. Schuchat, A. 2001. Group B streptococcal disease: from trials and tribulations to triumph and trepidation. *Clin. Infect. Dis.* **33**:751–756.
27. 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.
28. 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. USA* **102**:13950–13955.
29. 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. USA* **99**:12391–12396.
30. Toro, N. 2003. Bacteria and Archaea group II introns: additional mobile genetic elements in the environment. *Environ. Microbiol.* **5**:143–151.