JB Journals.ASM.org

# Whole-Genome Comparison Uncovers Genomic Mutations between Group B Streptococci Sampled from Infected Newborns and Their Mothers

**Alexandre Almeida,[a,b,c] Adrien Villain,[d] Caroline Joubrel,[e,f,g,h,i] Gérald Touak,[e,f] Elisabeth Sauvage,[a,b] Isabelle Rosinski-Chupin,[a,b] Claire Poyart,[e,f,g,h,i] Philippe Glaser[a,b,d]**

Institut Pasteur, Unité de Biologie des Bactéries Pathogènes à Gram-Positif, Paris, France[a]; CNRS UMR3525, Paris, France[b]; Université Pierre et Marie Curie, Paris, France[c]; Institut Pasteur, Plateforme de Bio-Analyse Génomique, Paris, France[d]; Service de Bactériologie, Centre National de Référence des Streptocoques, Groupe Hospitalier Paris Centre Cochin-Hôtel Dieu-Broca, Assistance Publique Hôpitaux de Paris, Paris, France[e]; DHU Risques et Grossesse, Assistance Publique Hôpitaux de Paris, Paris, France[f]; INSERM, U1016, Paris, France[g]; CNRS (UMR 8104), Paris, France[h]; Université Paris Descartes, Sorbonne Paris Cité, Paris, France[i]

## ABSTRACT

***Streptococcus agalactiae*** **(group B *Streptococcus* or GBS), a commensal of the human gut and genitourinary tract, is a leading cause of neonatal infections, in which vertical transmission from mother to child remains the most frequent route of contamination. Here, we investigated whether the progression of GBS from carriage to disease is associated with genomic adaptation. Whole-genome comparison of 47 GBS samples from 19 mother-child pairs uncovered 21 single nucleotide polymorphisms (SNPs) and seven insertions/deletions. Of the SNPs detected, 16 appear to have been fixed in the population sampled whereas five mutations were found to be polymorphic. In the infant strains, 14 mutations were detected, including two independently fixed variants affecting the *covRS* locus, which is known to encode a major regulatory system of virulence. A one-nucleotide insertion was also identified in the promoter region of the highly immunogenic surface protein Rib gene. Gene expression analysis after incubation in human blood showed that these mutations influenced the expression of virulence-associated genes. Additional identification of three mutated strains in the mothers' milk raised the possibility of the newborns also being a source of contamination for their mothers. Overall, our work showed that GBS strains in carriage and disease scenarios might undergo adaptive changes following colonization. The types and locations of the mutations found, together with the experimental results showing their phenotypic impact, suggest that those in a context of infection were positively selected during the transition of GBS from commensal to pathogen, contributing to an increased capacity to cause disease.**

## IMPORTANCE

**Group B *Streptococcus* (GBS) is a major pathogen responsible for neonatal infections. Considering that its colonization of healthy adults is mostly asymptomatic, the mechanisms behind its switch from a commensal to an invasive state are largely unknown. In this work, we compared the genomic profile of GBS samples causing infections in newborns with that of the GBS colonizing their mothers. Multiple mutations were detected, namely, within key virulence factors, including the response regulator CovR and surface protein Rib, potentially affecting the pathogenesis of GBS. Their overall impact was supported by differences in the expression of virulence-associated genes in human blood. Our results suggest that during GBS's progression to disease, particular variants are positively selected, contributing to the ability of this bacterium to infect its host.**

*Streptococcus agalactiae*, or group B *Streptococcus* (GBS), is currently regarded as one of the leading causes of neonatal sepsis and meningitis worldwide (1–4). Known as a commensal of the digestive and genitourinary tracts of 10 to 30% of the human population, GBS is also a significant source of disease in immunocompromised adults and in the elderly (4–7). Apart from causing human infections, GBS is an etiological agent of bovine mastitis and invasive disease in fish (8–10). First identified in the 1930s as a cause of neonatal mortality, it became a widespread concern in many developed countries in the second half of the 20th century (11–13). In the United States, throughout the 1990s, 7,600 cases of infection in newborns leading to 310 deaths per year were observed, whereas most recently, the global burden of GBS disease was estimated to be 0.53 per 1,000 live births (4, 14). Furthermore, GBS infections can lead to additional complications. Particularly regarding neonatal meningitis, studies have shown some degree of cognitive impairment in 15 to 20% of affected infants, at least until 5 years after birth (15, 16).

Neonatal infections are classified as either early-onset disease (EOD) when they occur in newborns within the first 6 days of life or late-onset disease (LOD) when they are diagnosed from 7 days

**Received** 2 June 2015 **Accepted** 5 August 2015

**Accepted manuscript posted online** 17 August 2015

**Citation** Almeida A, Villain A, Joubrel C, Touak G, Sauvage E, Rosinski-Chupin I, Poyart C, Glaser P. 2015. Whole-genome comparison uncovers genomic mutations between group B streptococci sampled from infected newborns and their mothers. J Bacteriol 197:3354–3366. doi:10.1128/JB.00429-15.

**Editor:** V. J. DiRita

Address correspondence to Claire Poyart, claire.poyart@cch.aphp.fr, or Philippe Glaser, pglaser@pasteur.fr.

A.V. and C.J. contributed equally to this work.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.00429-15.



doi:10.1128/JB.00429-15

up to approximately 3 months of age (4). Intrapartum antibioprophylaxis for parturients at risk of GBS infection was able to reduce the incidence of EOD, but it did not affect the rate of LOD (3). In EOD, a GBS strain has colonized the maternal vaginal tract and is subsequently transferred to the baby during or just before birth. The most frequent form of infection is thought to be through aspiration, by the baby, of GBS-contaminated amniotic or vaginal fluid from the mother. This causes an initial dissemination of the bacteria in the respiratory epithelium and in the child's bloodstream, leading to clinical manifestations usually in the form of pneumonia or sepsis (4, 17, 18). Late-onset cases typically result in meningitis, but the mode of transmission remains unclear. LOD may result from community- or hospital-acquired contamination after birth, even though the mother is likely the main source of infection. GBS strains can be grouped into 10 different serotypes according to their polysaccharide capsule (19). Epidemiological data have established a strong link between capsular serotype (CPS) III and a substantial proportion of EOD incidences, as well as the majority of LOD infections (20–23). Within these strains, multilocus sequence typing (MLST) analyses have inferred sequence type 17 (ST17) to be the most frequently associated lineage, which led to its current label as a "hypervirulent lineage" (22, 24).

To investigate the emergence of GBS neonatal infections, population and evolutionary studies have detailed a recent origin of the main human-adapted clonal complexes, with recombination of large chromosomal regions playing a key role in their genomic diversification (25). By Bayesian phylogenetic inference, the widespread use of tetracycline in the 1960s was reported as one of the main driving forces behind the selection of pathogenic clones (26). Nevertheless, not all tetracycline-resistant lineages have disseminated; thus, additional colonization and virulence properties seem to have been responsible for the expansion of certain clones still causing disease today. In this regard, GBS pathogenesis is related to several virulence factors, and their role is to promote the ability of the bacteria to infect and damage the host. Some notable examples include the hypervirulent GBS adhesin (HvgA), the β-hemolysin encoded by the *cyl* operon, the C5a peptidase, and surface proteins of the alpha-like family (17, 18, 27–30). Moreover, the expression of major virulence factors is regulated by the two-component system CovRS (control of virulence regulator and sensor). Inactivation of this system, first described in group A *Streptococcus* (GAS), was proven to enhance the pathogenesis of GAS by derepressing the transcription of virulence-associated genes (31). CovR is also a key regulator of virulence in GBS, depending on the site or mode of infection in the host (32, 33).

With the advent of high-throughput sequencing, novel insights into the *in vivo* evolution of many bacterial pathogens have been obtained. For instance, studies on *Staphylococcus aureus*, *Burkholderia pseudomallei*, and *Pseudomonas aeruginosa* have highlighted the selection of a diverse range of genomic mutations during chronic infection (34–36). Furthermore, whole-genome sequencing was recently used to investigate the within-host diversity and transmission of methicillin-resistant *S. aureus* in a veterinary context (37). Yet, specifically in GBS, the exact extent of genomic identity and virulence potential between strains colonizing women and those acquired by their babies through direct transmission remains to be elucidated. Comparative genomic analyses of GBS strains from carriage and disease would help gain a better understanding of the switch from commensal to pathogen and infer crucial virulence and genetic determinants responsible for neonatal GBS diseases.

In this work, we aimed to determine whether specific mutations are selected throughout the course of disease. We performed whole-genome sequencing of a unique collection of GBS samples from infected neonates and from their mothers in either a carriage or a disease context. Interestingly, the majority of the mutations found appear to have been fixed in their population, revealing a selective signal for nonsynonymous substitutions located within functionally significant genes. Thus, we suggest that the disease-associated variants underwent a positive selection process that might have contributed to the increased virulence of these mutated strains. However, by comparing each mutated region with publicly available GBS genomes as an outgroup, we also identified mutant alleles in samples obtained from carriage. Mutant variants found in GBS-positive milk samples from three mothers are compatible with a bidirectional transmission of GBS in which the infants contaminated their mothers through breastfeeding.

## MATERIALS AND METHODS

**Bacterial strains and culture conditions.** A total of 47 GBS samples from 19 mother-child pairs were used in this study (Table 1; see Table S1 in the supplemental material). Biological samples were collected at 12 different hospitals throughout France by following routine clinical laboratory procedures in which an indeterminate number of GBS colonies were sampled and subsequently transferred to the Centre National de Référence des Streptocoques (CNR-Strep). These samples were cultured either on Todd-Hewitt (TH) agar plates at 37°C or on Columbia agar plates with 5% horse blood at 37°C with 5% $CO_2$ and stored in 20% glycerol–TH broth at −80°C.

**Sequencing, *de novo* assembly, and epidemiological analysis.** Chromosomal DNA was extracted with the DNeasy blood and tissue kit (Qiagen) by following the manufacturer's instructions. Genomes were sequenced by the Illumina HiSeq 2000 and MiSeq sequencing platforms with single-read runs of 101 and 150 nucleotides, respectively (see Fig. S1A in the supplemental material). Reads were filtered by the fqCleaner tool according to the following parameters. Each read was trimmed off to remove nonconfident bases at the 5′ and 3′ ends with a quality score below 20. All reads shorter than 30 bp were discarded. Reads with <80% confident bases were excluded, as were artifactual and duplicated reads. Genome sequences were assembled by the Velvet software (38) with an optimized k-mer value, a minimum coverage of 10, and a contig length of at least 200 bp. STs were determined by MLST by extracting the sequences of the seven genes of the GBS MLST system (*adhP*, *pheS*, *atr*, *glnA*, *sdhA*, *glcK*, and *tkt*) (24) and comparing them with the known STs from the GBS MLST web server (http://pubmlst.org/sagalactiae/) through the Center for Genomic Epidemiology web tool (http://www.genomicepidemiology.org/). Capsular types were determined by molecular serotyping (39) and confirmed by BLASTn similarity search of the nucleotide sequences of the 10 *cps* loci of the known GBS serotypes.

**Variant detection.** Within each pair of samples, several approaches were taken to ensure the most reliable detection of the existing genomic variation within the population and to circumvent the limitations of traditionally used bioinformatic pipelines (see Fig. S1 in the supplemental material). First, for each pair, both the child and mother filtered reads were mapped to the assembled genome of the mother sample by Burrows-Wheeler Aligner (BWA) (40). Variant calling was performed with Genome Analysis ToolKit (41) by taking into account the published recommendations (42, 43). Single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) were prefiltered and selected according to the following parameters adapted from Lieberman et al. (44): at least 15 reads aligning in both directions, at least a 100-bp distance from the contig boundaries of the reference assembly, an average base quality greater than 19, an average mapping quality above 34, and a *P* value above 0.05 with

**TABLE 1** *S. agalactiae* samples used in this study

| Strain | Pair | Hospital[a] | Location | Yr[b] | Time (days) after birth[c] | Disease | Origin | Serotype | MLST |
|---|---|---|---|---|---|---|---|---|---|
| Mother 1 | 1 | A | Paris | 2012 | 0 | NA[d] | Vaginal fluid | III | ST17 |
| Child 1 | | A | Paris | 2012 | 0 | EOD | Blood culture | III | ST17 |
| Mother 2 | 2 | B | Clamart | 2012 | 110 | NA | Milk | Ia | ST23 |
| Child 2 | | B | Clamart | 2012 | 104 | LOD | Blood culture | Ia | ST23 |
| Mother 3 | 3 | A | Paris | 2008 | 0 | NA | Urine | Ia | ST23 |
| Child 3 | | A | Paris | 2008 | 1 | EOD | Cerebrospinal fluid | Ia | ST23 |
| Mother 4 | 4 | A | Paris | 2008 | 0 | NA | Vaginal fluid | III | ST17 |
| Child 4 | | A | Paris | 2008 | 6 | EOD | Blood culture | III | ST17 |
| Mother 5-1 | 5 | A | Paris | 2008 | 4 | NA | Vaginal fluid | III | ST17 |
| Mother 5-2 | | A | Paris | 2008 | 23 | NA | Milk | III | ST17 |
| Child 5 | | A | Paris | 2008 | 22 | LOD | Blood culture | III | ST17 |
| Mother 6 | 6 | C | Colombes | 2008 | 1 | NA | Urine | Ia | ST24 |
| Child 6-1 | | C | Colombes | 2008 | 1 | EOD | Blood culture | Ia | ST24 |
| Child 6-2 | | C | Colombes | 2008 | 0 | EOD | Gastric fluid | Ia | ST24 |
| Mother 7 | 7 | D | Chartres | 2008 | 29 | NA | Blood culture | V | ST1 |
| Child 7 | | E | Montmorency | 2008 | 18 | LOD | Blood culture | III | ST17 |
| Mother 8 | 8 | A | Paris | 2008 | 0 | NA | Vaginal fluid | III | ST17 |
| Child 8-1 | | A | Paris | 2008 | 1 | EOD | Cerebrospinal fluid | III | ST17 |
| Child 8-2 | | A | Paris | 2008 | 1 | EOD | Blood culture | III | ST17 |
| Mother 9 | 9 | F | Colombes | 2008 | 0 | NA | Amniotic fluid | III | ST23 |
| Child 9 | | F | Colombes | 2008 | 1 | EOD | Blood culture | III | ST23 |
| Mother 10-1 | 10 | A | Paris | 2009 | 16 | NA | Milk | Ia | ST23 |
| Mother 10-2 | | A | Paris | 2009 | −3 | NA | Vaginal fluid | Ia | ST23 |
| Child 10 | | A | Paris | 2009 | 16 | LOD | Blood culture | Ia | ST23 |
| Mother 11 | 11 | A | Paris | 2009 | −1 | NA | Vaginal fluid | III | ST23 |
| Child 11-1 | | A | Paris | 2009 | 21 | LOD | Blood culture | III | ST17 |
| Child 11-2 | | A | Paris | 2009 | 35 | LOD | Blood culture | III | ST17 |
| Mother 12 | 12 | G | Paris | 2009 | 0 | NA | Blood culture | III | ST27 |
| Child 12-1 | | G | Paris | 2009 | 0 | EOD | Gastric fluid | III | ST27 |
| Child 12-2 | | G | Paris | 2009 | 0 | EOD | Blood culture | III | ST27 |
| Mother 13 | 13 | H | Amiens | 2010 | 7 | NA | Milk | III | ST17 |
| Child 13 | | H | Amiens | 2010 | 5 | EOD | Cerebrospinal fluid | III | ST17 |
| Mother 14 | 14 | I | Montpellier | 2010 | 33 | NA | Milk | III | ST17 |
| Child 14 | | I | Montpellier | 2010 | 29 | LOD | Cerebrospinal fluid | III | ST17 |
| Mother 15 | 15 | C | Colombes | 2010 | −1 | NA | Vaginal fluid | V | ST1 |
| Child 15-1 | | C | Colombes | 2010 | 0 | EOD | Gastric fluid | V | ST1 |
| Child 15-2 | | C | Colombes | 2010 | 0 | EOD | Blood culture | V | ST1 |
| Mother 16 | 16 | J | Frejus | 2011 | 0 | NA | Placenta | III | ST17 |
| Child 16 | | J | Frejus | 2011 | 0 | EOD | Gastric fluid | III | ST17 |
| Mother 17 | 17 | C | Colombes | 2011 | 0 | NA | Placenta | III | ST17 |
| Child 17 | | C | Colombes | 2011 | 0 | EOD | Blood culture | III | ST17 |
| Mother 18 | 18 | K | Le Kremlin Bicêtre | 2011 | 0 | NA | Blood culture | Ia[e] | ST24 |
| Child 18 | | K | Le Kremlin Bicêtre | 2011 | 0 | EOD | Blood culture | Ia[e] | ST24 |

(Continued on following page)

**TABLE 1** (Continued)

| Strain | Pair | Hospital[a] | Location | Yr[b] | Time (days) after birth[c] | Disease | Origin | Serotype | MLST |
|---|---|---|---|---|---|---|---|---|---|
| Mother 19 | 19 | L | Compiègne | 2011 | 7 | NA | Vaginal fluid | III | ST17 |
| Child 19-1 | | L | Compiègne | 2011 | 5 | EOD | Cerebrospinal fluid | III | ST17 |
| Child 19-2 | | L | Compiègne | 2011 | 5 | EOD | Blood culture | III | ST17 |
| Child 19-3 | | L | Compiègne | 2011 | 6 | EOD | Urine | III | ST17 |

[a] One-letter code representing the hospital at which the sample was collected.
[b] Year of sampling.
[c] Time elapsed after the birth of the child from the corresponding pair.
[d] NA, nonapplicable.
[e] Deletion of the *cpsIad*, *cpsE*, *cpsF*, and *cpsG* genes from the *cps* operon was found.

Fisher's exact test supporting a null hypothesis that the variant frequency is the same for reads in both directions (see Fig. S1B in the supplemental material). After this initial filtering, if a variant was detected in <97% of the reads in the mapping of the child sample or >3% of the reads in the mapping of the mother population, it was suspected of corresponding to a polymorphic variant. Therefore, in these cases, an isolated colony from the putative mixed sample was sequenced and used as an isogenic control. False-positive polymorphisms were inferred as those found to be heterogeneous (with an allele frequency between 3% and 97%) in the mapping of the control sample against its assembled genome. Filtered SNPs were then classified as either fixed or polymorphic according to whether they were present in more or less than 97% of the reads, respectively, in at least one of the samples of each pair (Table 2; see Table S2 in the supplemental material). No frequency estimate was calculated for the indels because of a greater probability of mapping and sequencing errors.

To extend this analysis to the identification of genomic islands or other genetic material acquired by the GBS population in the child, the unmapped reads recovered from each pair were assembled by the Velvet

**TABLE 2** Genomic mutations detected in the population of each pair of samples

| Sample[a] | MLST | Type | Classification[b] | Locus ID[c] | Function[d] | BLAST[e] |
|---|---|---|---|---|---|---|
| Mother 2 | ST23 | SNP | Nonsense (Q7X), fixed | *gbs1596* | Intermediary metabolism | Glyoxalase |
| Mother 2 | ST23 | SNP | Synonymous, fixed | *gbs0198* | Intermediary metabolism | Polynucleotide phosphorylase, alpha chain |
| Mother 4 | ST17 | SNP | Missense (K88E), polymorphic | *SAG0163_02655* | Information pathways | Site-specific recombinase |
| Mother 9 | ST23 | Indel | Noncoding[h] | *gbs0476* (3′ UTR) | Unknown | Putative membrane protein |
| Child 9 | ST23 | SNP | Missense (R79H), polymorphic | *gbs1259* | Intermediary metabolism | Metallo-beta-lactamase superfamily protein |
| Mother 10-1 | ST23 | SNP | Noncoding, fixed | *gbs1986* | Cellular processes | ABC transporter |
| Child 11-2[f] | ST17 | Indel | Frameshift | *gbs1427* | Information pathways | KH domain protein |
| Child 11-2[f] | ST17 | SNP | Missense (A175D), fixed | *gbs1950* | Cellular processes | Phosphate ABC transporter |
| Mother 12 | ST27 | SNP | Noncoding, fixed | *gbs1946* (5′ UTR) | Cellular processes | Glucose-specific PTS enzyme IIABC |
| Child 12[g] | ST27 | SNP | Missense (A49V), fixed | *gbs0668* | Intermediary metabolism | D-lactate dehydrogenase |
| Child 12[g] | ST27 | SNP | Missense (G814S), fixed | *gbs1038* | Cellular processes | Hypothetical ABC transporter permease |
| Child 12-1 | ST27 | SNP | Noncoding, polymorphic | *gbs1946* (5′ UTR) | Cellular processes | Glucose-specific PTS enzyme IIABC |
| Child 12-1 | ST27 | Indel | In-frame deletion[h] | *gbs1377* | Intermediary metabolism | Homocysteine *S*-methyltransferase |
| Child 12-1 | ST27 | Indel | Large deletion[h] | | Unknown | Phage terminase |
| Child 12-2 | ST27 | SNP | Noncoding, fixed | *gbs1672* (5′ UTR) | Information pathways | Two-component response regulator (CovR) |
| Child 12-2 | ST27 | SNP | Missense (R186K), fixed | *gbs0231* | Cellular processes | Efflux protein |
| Mother 14 | ST17 | SNP | Noncoding, fixed | *gbs0330* (5′ UTR) | Information pathways | Transcriptional regulator (MarR family) |
| Mother 14 | ST17 | SNP | Missense (G30V), fixed | *gbs1234* | Cellular processes | NeuD protein |
| Child 14 | ST17 | Indel | Duplication[h] | *gbs1958* | Information pathways | Transcriptional regulator (MerR family) |
| Mother 15 | ST1 | Indel | Noncoding[h] | *gbs0419* (5′ UTR) | Intermediary metabolism | GDXG lipolytic enzyme family protein |
| Mother 15 | ST1 | SNP | Missense (I183T), fixed | *gbs1230* | Intermediary metabolism | Glycerol-3-phosphate acyltransferase (PlsY) |
| Mother 15 | ST1 | SNP | Noncoding, fixed | | Unknown | *Anaerococcus prevotii* plasmid |
| Mother 18 | ST24 | SNP | Synonymous, polymorphic | *gbs0189* | Cellular processes | Trehalose-specific PTS enzyme IIABC |
| Mother 18 | ST24 | SNP | Missense (P396S), polymorphic | *gbs1460* | Cellular processes | Polysaccharide biosynthesis protein |
| Mother 19 | ST17 | SNP | Missense (P80S), fixed | *gbs2047* | Information pathways | RecA protein |
| Child 19[g] | ST17 | SNP | Nonsense (E96X), fixed | *gbs1159* | Intermediary metabolism | Phosphotransacetylase |
| Child 19[g] | ST17 | SNP | Missense (A115V), fixed | *gbs1672* | Information pathways | Two-component response regulator (CovR) |
| Child 19[g] | ST17 | Indel | Noncoding[h] | *GBSCOH1_0416* | Cellular processes | Rib protein |

[a] Sample of the corresponding pair in which the mutant allele was most frequently detected.
[b] Classification of the mutation detected. For nonsynonymous substitutions, the amino acid change is shown in parentheses. A mutation was considered fixed if at least 97% of the filtered reads supported the mutation in at least one of the samples of the pair (see Table S2 in the supplemental material). No frequency estimate was calculated for indels.
[c] Locus ID of the mutated gene: *SAG0163_02655* corresponding to GBS MRI Z1-215 (GenBank accession no. NZ_ANET01000026), *GBSCOH1_0416* corresponding to GBS COH1 (GenBank accession no. HG939456), and the rest corresponding to GBS NEM316 (GenBank accession no. AL732656).
[d] Functional category of each affected gene based on the SagaList database of NEM316.
[e] Annotation based on BLASTn homology search.
[f] Mutations were called by mapping against child 11-1.
[g] Mutations were detected in all child samples of the corresponding pair.
[h] No homology with any annotated GBS genomes in the NCBI database.

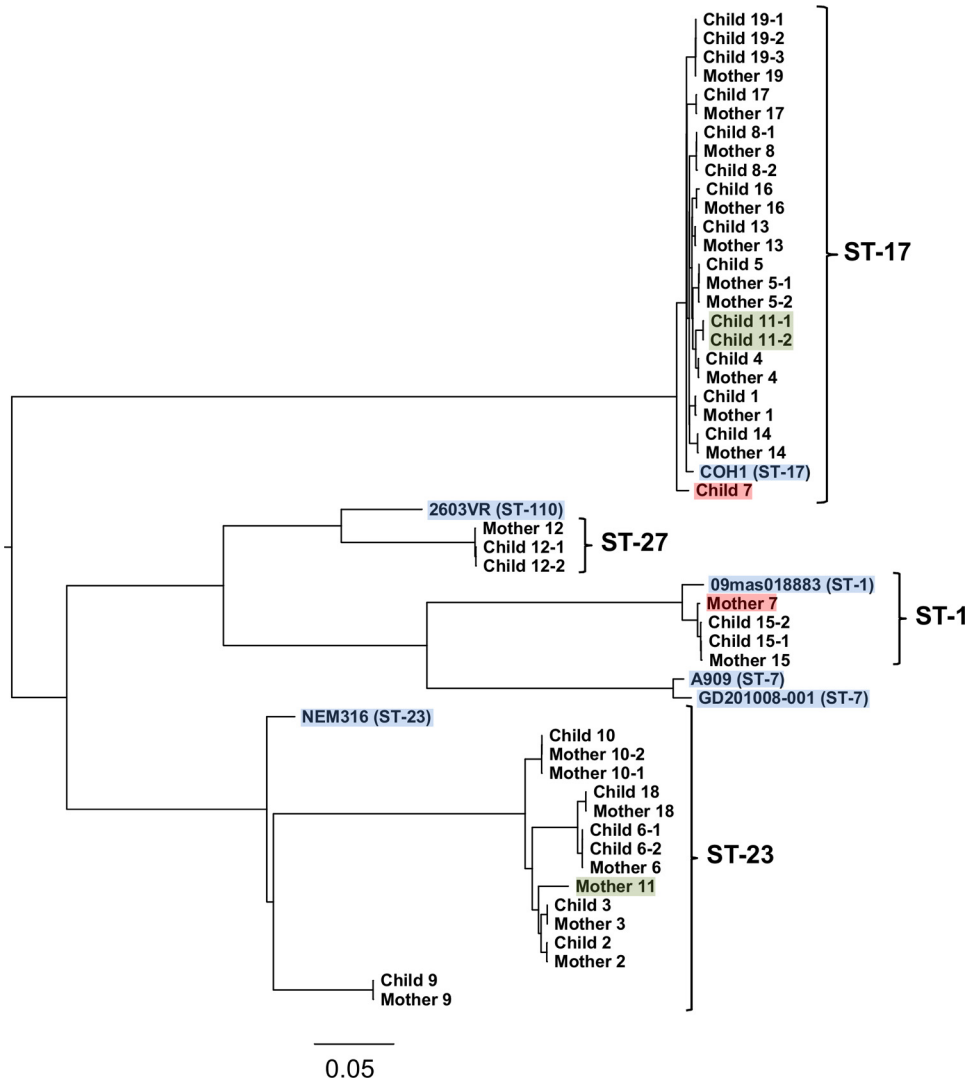


**FIG 1** Whole-genome phylogeny of mother-child GBS pairs. Maximum-likelihood phylogenetic tree built with snpTree (49) based on genomic SNPs. The unrooted tree depicts the phylogenetic relationship of each strain in our study with the published complete GBS genomes in the NCBI database (highlighted in blue), i.e., 09mas018883 (GenBank accession no. HF952104), 2603V/R (GenBank accession no. AE009948), A909 (GenBank accession no. CP000114), COH1 (GenBank accession no. HG939456), GD201008-001 (GenBank accession no. CP003810), and NEM316 (GenBank accession no. AL732656). Strain names are as indicated in Table 1 (see also Table S1 in the supplemental material). Strains from pairs 7 and 11 are highlighted in red and green, respectively.

software (see Fig. S1B in the supplemental material). The same k-mer, coverage, and contig length values input for the assembly of the strains from the corresponding pair were used. Contigs assembled from the unmapped reads were blast searched against their own assembled genome to exclude false-positive matches.

The same strategy of read mapping and variant detection was used with the strains of each pair against their closest complete genome within the NCBI database (see Fig. S1C in the supplemental material). Pairs 1, 4, 5, 8, 13, 14, 16, 17, and 19 were mapped against the genome of GBS COH1 (ST17); pairs 2, 3, 6, 9, 10, and 18 were mapped against the genome of GBS NEM316 (ST23); pair 12 was mapped against the genome of GBS 2603V/R (ST110); and pair 15 was mapped against the genome of GBS 09mas018883 (ST1) (Fig. 1).

Moreover, the *breseq* computational pipeline (45) was used to identify the gain or loss of small and large sequence fragments between the mother and child populations sampled. Mapping was performed with the filtered sequencing reads from the child sample against the assembled genome of the mother sample in each pair (see Fig. S1D in the supplemental material). Large deletions or insertions with a coverage value below 10 and within repeated regions or genes with multiple copy numbers were not considered. Each mutated gene was blast searched against the 300 sequenced GBS genomes in the NCBI database available in April 2015. On the basis of the identification of the transcription start sites mapped on the genome of GBS NEM316 (46), the relative position of each SNP located within noncoding regions was ascertained. Functional categories of the mutated genes were inferred on the basis of the SagaList annotation created from the GBS NEM316 sequencing project (47, 48).

**Phylogenetic inference.** To determine the strain most closely related to each isolate, we built a phylogenetic tree by using snpTree 1.1 (49). With this software, each assembled and complete GBS genome was initially aligned against GBS NEM316 with the Nucmer application from the MUMmer software package (version 3.23), and genomic SNPs were detected with the show-snps option. A filtering process was subsequently applied to discard SNPs <100 bp apart and within 100 bp of the contigs' boundaries. Finally, the resulting genomic SNPs were used to build a maximum-likelihood tree (49) (Fig. 1).

We inferred the wild-type and mutated alleles of each variable position by reconstructing their ancestral sequence with the most closely related publicly available GBS genomes (see Fig. S2 in the supplemental material). Sequences were aligned with ClustalW, and maximum-likelihood phylogenetic trees were built with MEGA 6.06 (50) by using a general time-reversible substitution model with a gamma distribution rate across sites and a proportion of invariant sites. In this work, we assumed that the mutation event occurred from the wild-type allele to the mutant allele, even though we cannot formally exclude the very unlikely possibility of a reversion of the mutated variant.

**RNA extraction and RT-qPCR.** Bacterial strains were precultured overnight in TH broth at 37°C in standing cultures without agitation. For incubation in human blood, a modified form of the protocol of Oggioni et al. (51) was used. Briefly, bacterial dilutions were first grown to the exponential growth phase (optical density at 600 nm of 0.4 to 0.5) in 10 ml of TH broth. Subsequently, 5 ml of each culture was pelleted and stored at −80°C while the other 5 ml was pelleted, washed twice with phosphate-buffered saline, and resuspended in 5 ml of whole human blood from one donor. Blood cultures were then incubated for 1 h at 37°C under agitation. Afterwards, samples were subjected to low-speed centrifugation (200 × *g* for 5 min) to remove human cells and the supernatant containing the bacteria was then centrifuged at 5,000 × *g* for 10 min. Bacterial pellets obtained were stored at −80°C. Total RNAs were extracted with a phenol/TRIzol-based purification method as previously described (32). RNA quality was assessed by electrophoretic analysis with an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA), and reverse transcription (RT) was performed with the Superscript II indirect cDNA kit (Invitrogen, Life Technologies). Primer pairs were designed with Primer-BLAST from the NCBI website to obtain a predicted amplicon size of 100 to 200 bp (see Table S3 in the supplemental material) and quantitative PCR (qPCR) was carried out with SYBR green PCR kits (Applied Biosystems, Life Technologies). Relative gene expression was quantified by a standard-curve-based method in which regression analysis was performed by using serial dilutions of a positive amplification control. Calculated values given in arbitrary units were normalized with the expression of the housekeeping gene *gyrA*. Each assay was performed in triplicate with three independent cultures. A two-tailed *t* test was carried out to determine whether expression differences were statistically significant.

**D-Lactate quantification and carbohydrate metabolism assays.** To measure the concentration of D-lactate in the supernatant of each culture, we used the D-Lactate Colorimetric Assay kit (Sigma-Aldrich) according to the manufacturer's instructions. The API 50 CH test (bioMérieux) was used to compare the abilities of GBS strains from each target pair to ferment 49 different carbohydrates. Each strain was initially grown for 18 to 24 h on TH agar plates. Afterwards, a quantity of cells amounting to two bacterial loops was suspended in 10 ml of API 50 CHL medium. The culture suspensions were then loaded onto API 50 CH test strips in accordance with the manufacturer's instructions, and the tubes were covered with mineral oil. The test strips were incubated at 37°C, and results were recorded 24 and 48 h later.

**Nucleotide sequence accession numbers.** Sequencing reads from each run and the corresponding genome assemblies have been deposited in the EMBL nucleotide sequence database (http://www.ebi.ac.uk/ena) under study accession number PRJEB6691. For the accession numbers of the individual files, see Table S1 in the supplemental material.

## RESULTS

**Characterization and genotyping of GBS pairs.** For this study, a total set of 47 GBS samples were obtained from 19 pairs of infected newborns and their mothers (Table 1; see Table S1 in the supplemental material). One to three samples were collected per individual, corresponding to various sites colonized by GBS at different time points relative to the child's birth. A total of 11 samples were collected from carriage-related sites in the mother (amniotic fluid, placenta, and vaginal fluid), in contrast to 10 from other contaminated samples (blood, milk, and urine). Conversely, 22 blood, cerebrospinal fluid, and urine samples were obtained from the newborns, together with only 4 carriage samples of gastric fluid. Of the 19 pairs analyzed, 13 corresponded to EOD cases and 6 corresponded to LOD, with strains isolated 16 to 110 days after the infant's birth (Table 1).

After whole-genome sequencing, CPSs and MLST profiles were determined by BLASTn similarity searches of the assembled genomes of the samples (Table 1). CPS III and, more specifically, hypervirulent clone ST17 were the most abundant lineages identified in our work, accounting for 31 (66%) and 25 (53%) strains, respectively. GBS samples obtained from LOD belonged exclusively to ST17 ($n = 4$) and ST23 ($n = 2$), while ST17 accounted for four (80%) of five strains responsible for meningitis. From the genotyping of the 19 sets of mother-child samples analyzed, 17 pairs (89%) corresponded to identical CPSs and STs, whereas strains from pairs 7 and 11, both obtained from LOD, belonged to distinct lineages. In these two pairs, the infants were infected by ST17 strains and their mothers were colonized by ST1 (pair 7) or ST23 (pair 11). No genomic comparison of the mother and child populations of these two pairs was performed, considering that they belong to distantly related lineages (Fig. 1 and Table 1). Even so, two samples were retrieved from the infant corresponding to pair 11, isolated from blood samples cultured at 21 (child 11-1) and 35 (child 11-2) days after birth. Since the two isolates belong to the same ST, we compared them by using the first population sampled as a reference to track any genomic changes occurring throughout the course of infection.

**Comparative analysis reveals genomic variations within functionally relevant regions.** To analyze the population diversity of each pair of GBS samples (Table 1; see Table S1 in the supplemental material), different strategies based on read mapping and *de novo* assembly were applied as described in Materials and Methods and as depicted in Fig. S1 in the supplemental material. The purpose was to identify genomic differences between mother and child GBS populations while taking into account the possibility of within-host variation. A phylogenetic tree was built with all 19 GBS pairs, together with the most closely related strains whose complete genome sequences are publicly available. A total of 6,122 genomic SNPs were used (Fig. 1). Even though pairs 4, 5, and 8 were all ST17 strains collected from the same hospital during the same year, their phylogenetic distribution suggests that these pairs are no closer to each other than to the other ST17 pairs collected in different time periods from different hospitals. Moreover, mother and child strains from pairs 7 and 11 are clearly shown to belong to divergent lineages (Fig. 1).

No genomic differences were detected in eight (42%) pairs (1, 3, 5, 6, 8, 13, 16, and 17), encompassing seven incidences of EOD and one of LOD. Whole-genome comparison of the nine genetically diverse pairs of samples (2, 4, 9, 10, 12, 14, 15, 18, and 19), as well as the two samples obtained from child 11, resulted in the identification of a total of 21 SNPs and seven indels, ranging from 1 to 1,132 bp (Table 2; see Table S2 in the supplemental material). Determination of the frequency of the reads supporting each of the SNPs detected within the population allowed the identification of 16 fixed mutations and five polymorphic sites present in less than 97% (between 3 and 91%) of the population sampled (see Table S2). BLASTn similarity search of the flanking regions of each mutated site against published genomes in the NCBI database allowed the characterization of each variant identified (Table

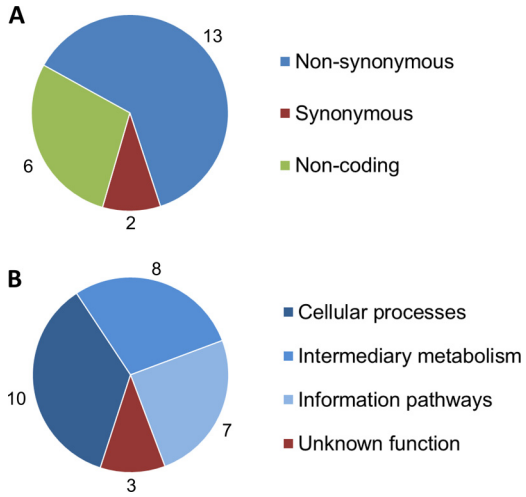


**FIG 2** Characteristics of the mutations identified. (A) Classification of the SNPs identified in 10 of the 19 pairs of GBS samples analyzed by comparative genomics. (B) Functional category of each mutated region based on homology search of the flanking gene of each site, as annotated in the SagaList database of GBS NEM316. Both the type of mutation and the functional category are represented by different colors, as shown in the key. Each value depicted is the number of mutations identified belonging to each category.

2; see Table S2 in the supplemental material). Thirteen SNPs were classified as nonsynonymous (11 missense and 2 nonsense), but only 2 were classified as synonymous (Fig. 2A). Furthermore, one out-of-frame deletion (1 bp) was detected, together with one in-frame deletion (18 bp), a 46-bp out-of-frame duplication, and a large 1,132-bp deletion encompassing a phage terminase. Surprisingly, nine of the mutations (six SNPs and three indels) occurred in noncoding sequences, including five in 5′ untranslated regions (UTRs) and one in a 3′ UTR. Overall, on the basis of the functional classification of all the mutated regions, deduced from the SagaList database of GBS NEM316 (47, 48), 10 were related to cell envelope and cellular processes, 8 were related to intermediary metabolism, 7 were related to information pathways, and 3 were related to unknown functions (Fig. 2B).

In order to identify the most likely ancestral sequence of the common ancestor of each GBS pair and determine the mutated strain, we compared each variable locus with the 300 sequenced GBS genomes available in the NCBI database as an outgroup. These comprise carriage and invasive isolates of human origin, as well as strains from bovine, fish, and other animal hosts. We infer the wild-type allele as the one identical to all of the genetically closest genomes and the mutant variant as the one differing from these sequences (see Fig. S2 in the supplemental material). For all but four of the mutations we detected, the same variant was present in each of the 300 NCBI strains. A total of 14 mutations were deduced to have occurred within nine isolates obtained from the mother and 14 mutations within eight samples from the child (Table 2).

Focusing on the most relevant mutations identified in the GBS populations sampled, assessment of pair 2, from a case of LOD, revealed two mutations in the mother strains isolated from milk: one SNP within a gene coding for a glyoxalase (*gbs1596*, according to GBS NEM316 [48]), a detoxification system of reactive aldehydes, and another in a region encoding a polynucleotide phosphorylase (*gbs0198*). This enzyme is conserved among bacteria and has been shown to regulate the expression of virulence factors in *S. aureus*, *Francisella tularensis*, and *Yersinia* spp. (52–54).

In pair 12 (EOD), only one mutation was present in the mother GBS population, obtained from a blood culture, in the 5′ UTR of a glucose-specific phosphotransferase system (PTS) enzyme IIABC gene (*gbs1946* [*ptsG*]) that takes part in the phosphorylation and transport of glucose (55). In the gastric fluid sample from the child (child 12-1), six polymorphic variants were detected in various proportions (see Table S2 in the supplemental material), including an additional mutation in the 5′ UTR of *ptsG* (Table 2; see Table S2). In contrast, four mutations appear to have been positively selected specifically in the GBS population obtained from the blood (child 12-2). Interestingly, one of the mutations is located in the 5′ UTR of the gene encoding the response regulator of virulence CovR (*gbs1672* [*covR*]). Since the most common etiological pathway of EOD caused by GBS is aspiration or ingestion of GBS-contaminated fluids (4, 17, 18), our observations within this pair suggest that multiple GBS variants were ingested by the baby through the digestive tract but only one prevailed and was selected in the child's bloodstream.

Regarding pair 14 (LOD), a duplication of 46 bp in the MerR family of transcriptional regulators gene (*gbs1958*) was identified in the newborn sample. These regulators are responsible for the activation of genes associated with transport and multidrug resistance in response to various environmental stresses, as reported in *Bacillus subtilis* and other bacterial species (56). Conversely, two mutations were detected in the mother sample: one SNP in the 5′ UTR of a gene encoding a transcriptional regulator of the multiple antibiotic resistance regulator (MarR) family (*gbs0330*) and one additional missense SNP in *neuD*. MarR proteins have been described as responsible for controlling the expression of virulence factors and the bacterial response to antibiotic and oxidative stresses (57, 58). The other mutated gene, *neuD*, codes for an acetyltransferase with a role in the biosynthesis of capsular sialic acid and has been implicated in the development of meningitis in neonates (59).

Concerning pair 18 (EOD), two polymorphic mutations were detected, one of which is in a genomic region coding for trehalose-specific PTS enzyme IIABC (*gbs0189* [*trePI*]). Trehalose is known to contribute to the adaptation of bacteria to osmotic stress, while this particular PTS, responsible for trehalose phosphorylation and transport, is an upregulated target of CovR (60). The mutation was identified in approximately 90% of the population obtained specifically from the mother's blood sample (see Table S2 in the supplemental material). Strikingly, in pair 12, the mother strains also from a blood culture were mutated in *ptsG*, another gene involved in carbon source uptake.

Lastly, four mutations were found within the samples of pair 19 (EOD). In the three child populations, a nonsense mutation was detected in the phosphotransacetylase gene (*gbs1159* [*pta*]), which is involved in acetate metabolism and ATP production; a missense mutation in *covR*, altering the same system mutated in pair 12 with essential regulatory functions for virulence (31, 32); and a 1-bp insertion in the homopolymeric A tract affecting the promoter of the alpha-like Rib surface protein-encoding gene (*gbs0147* [*rib*]), a region known to have a significant impact on the expression of this immunogenic protein (61). In regard to the mother sample, a nonsynonymous mutation was found in the *recA* gene (*gbs2047*), coding for a multifunctional protein with a central role in DNA repair by homologous recombination (62).

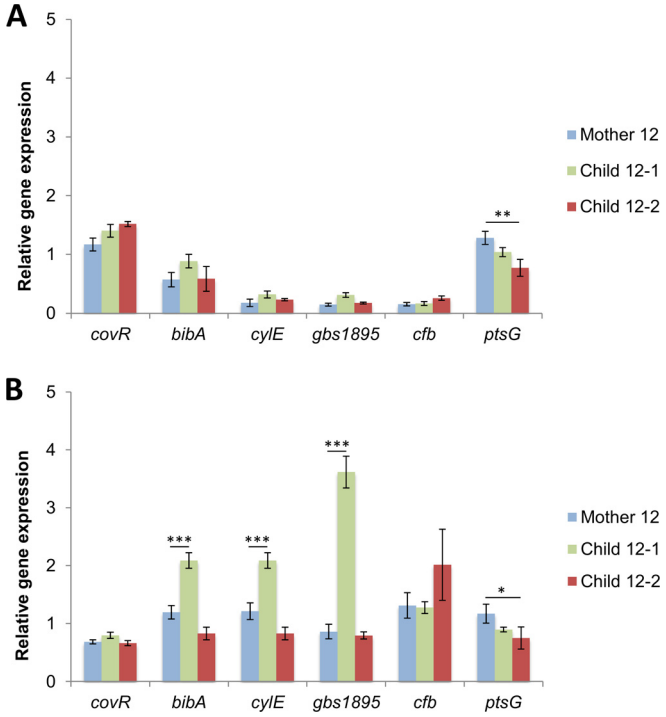


**FIG 3** Relative expression of mutated genes (pair 12). RT-qPCR results obtained with GBS strains from pair 12 in TH medium (A) and after incubation for 1 h in human blood (B). The mother 12 isolate was mutated in the 5′ UTR of *gbs1946* (*ptsG*), and both child strains were mutated in *gbs0668* (D-lactate dehydrogenase) and *gbs1038* (permease). The isolate from child 12-1 tested presented additional mutations affecting *gbs1946* (*ptsG*), *gbs1377* (homocysteine *S*-methyltransferase), and a phage terminase, while the child 12-2 strain specifically was mutated in the 5′ UTR of *gbs1672* (*covR*) and *gbs0231* (putative transporter). Gene expression is represented after normalization to the housekeeping gene *gyrA*. Gene names are indicated below each graph. Experiments were performed in triplicate with three independent cultures. Error bars represent standard deviations. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$.

Almost all of the mutations identified are novel, since most of them were not present in any of the 300 GBS genomes available in the NCBI database as of April 2015, except for two SNPs and two indels. The two SNPs, identified in *neuD* and in the 5′ UTR of *ptsG* from mother 12, were each detected in one genome of human origin. Moreover, the one-nucleotide indel in the 5′ UTR of *gbs0419* from mother 15 was distributed among 122 strains from human, bovine, and fish hosts, while the 18-bp deletion in *gbs1377* was identified in two human GBS strains. Altogether, of the 14 mutations identified in our study in the child strains of five pairs, three fixed variants were associated with pathogenicity-related features in two pairs, comprising two different mutations in *covR* and one in the promoter region of the gene encoding the surface protein Rib.

**Relative gene expression quantification showed distinct transcription levels of virulence-related genes.** In light of the genomic comparison of the 19 pairs of GBS samples, we focused on fixed mutations identified in the child and relating to virulence-associated regions to determine any differences in the expression levels of the affected genes (pairs 12 and 19). In pair 12, with SNPs in the 5′ UTRs of *covRS* and *ptsG*, RT-qPCRs were performed with both mutated genes, as well as four major targets of CovR (Fig. 3; see Table S3 in the supplemental material). Given that CovR is primarily a virulence repressor, we chose as negatively regulated targets the *hvgA* (*bibA*) gene, which encodes a variable surface protein (63) shown to be a hypervirulent adhesin in the ST17 lineage (22), and *cylE*, from the *cyl* operon responsible for the synthesis of β-hemolysin (30). One additional target, coding for a hypothetical secreted protein (*gbs1895*) repressed by CovR, was also tested (32). Contrariwise, the *cfb* gene encoding the CAMP factor toxin (32) was selected as a positively regulated target of CovRS. Seeing that CovR is a major regulator dependent on the host environment, gene expression was determined both in TH medium and after 1 h of incubation in human blood. To compare the expression of the different strains from pair 12, one isolate each from mother 12 and child 12-2 (Table 1) was used, representing the genotype observed within 100% of the population obtained from blood (Table 2; see Table S2 in the supplemental material). Additionally, an isolate collected from the gastric fluid of child 12-1 was also tested and confirmed by sequencing to contain the four SNPs observed at a higher frequency in the original sample (*gbs0668*, *gbs1038*, and *gbs1946*), as well as the 18-bp deletion in *gbs1377* and the 1,132-bp deletion of a phage terminase (Table 2; see Table S2 in the supplemental material). Comparing gene expression levels between growth in TH and after 1 h of incubation in human blood showed that *covR* expression was lower in whole human blood, contrasting with the transcription of the CovR negatively regulated targets *bibA*, *cylE*, and *gbs1895*, which was significantly increased (Fig. 3). However, no difference was observed between expression levels of the mother strain and the child strain with the mutation in the 5′ UTR of *covR* (child 12-2). Surprisingly, there was a significant increase in the expression of *bibA*, *cylE*, and *gbs1895* in the strain isolated from gastric fluid (child 12-1) after incubation in human blood, hinting that it could be the result of an indirect effect of one of the other mutations found in this strain. This difference was confirmed to be significant for *gbs1895* following incubation with blood from a second donor, although a smaller effect was observed for *bibA* and *cylE* (see Fig. S3A in the supplemental material). In addition, to check if these expression differences could be associated with the nonsynonymous mutation in the D-lactate dehydrogenase, we quantified the amounts of D-lactate in the supernatants of all of the pair 12 cultures used for the RT-qPCR but found no significant differences (see Fig. S3B in the supplemental material). Lastly, the SNP in the 5′ UTR of *ptsG* found in mother 12 led to a small but significant expression increase solely in comparison to the strain from the child's blood (child 12-2), which was devoid of any mutations affecting this gene (Fig. 3).

As for pair 19, representative isolates from both the mother and the child were tested, with the child isolate containing a nonsynonymous SNP in the *covR* gene, as well as a 1-bp insertion in the promoter of the gene encoding the surface protein Rib. With this in mind, quantitative experiments were carried out with additional sets of primers for the *rib* gene (Fig. 4; see Table S3 in the supplemental material). Interestingly, only after incubation in human blood was there a significant increase in the expression of *hvgA*, *cylE*, and *gbs1895* in the mutated strain retrieved from the child, together with a significant reduction in the transcription of *covR* itself. This suggests that the amino acid change in the *covR* mutant affected the regulation of this system and caused the overexpression of several virulence-associated genes in the child strain. Finally, the *rib* gene in both the mother and child strains was significantly more strongly expressed in whole human blood than

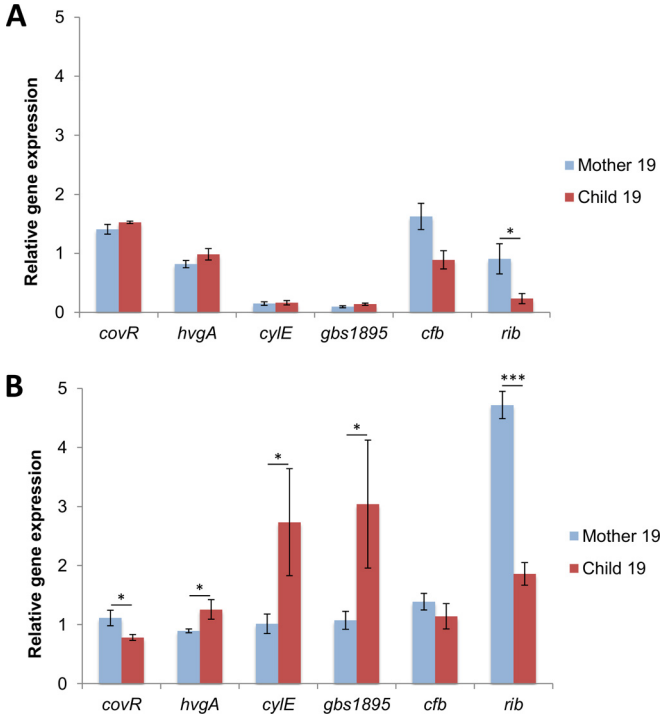


**FIG 4** Relative expression of mutated genes (pair 19). RT-qPCR results obtained with GBS strains from pair 19 in TH medium (A) and after incubation for 1 h in human blood (B). The mother 19 isolate tested was mutated in *gbs2047* (*recA*), and the child strain was mutated in *gbs1159* (*pta*), *gbs1672* (*covR*), and the promoter of *GBSCOH1_0416* (*rib*). Gene expression is represented after normalization to the housekeeping gene *gyrA*. Gene names are indicated below each graph. Experiments were performed in triplicate with three independent cultures. Error bars represent standard deviations. ***, $P < 0.001$; *, $P < 0.05$.

in TH medium, but there was a consistent and significant reduction in expression in the child strain, compared to that in the mother sample, under both of the conditions tested (Fig. 4). The reduced expression of *rib* might have contributed to the mutant strain's escape from the antibodies transmitted by the mother, allowing GBS to replicate and spread within the infected infant.

**Carbohydrate fermentation assays revealed no significant differences in metabolic activity.** Because we identified mutated genes related to sugar transport and metabolism, we performed the API 50 CHL test on GBS pairs harboring genomic differences in genes involved in carbon source metabolism. This assay examines the ability of the bacteria to ferment 49 different carbohydrates. Phenotypic tests were performed in duplicate on strains from pairs 12, 15, 18, and 19, but no significant differences were found under the conditions tested (see Table S4 in the supplemental material).

## DISCUSSION

Neonatal infections caused by GBS remain a significant health issue worldwide (1–4). In this work, we present for the first time a comparative genomic analysis of pairs of GBS samples obtained from infected infants and their mothers. Whole-genome comparison of 10 of 19 mother-child pairs revealed a total of 21 SNPs and seven indels (Table 2). In almost half (42%) of the GBS pairs, no differences were observed. Furthermore, the variants present in mother 15 were absent from the two samples from her child, while strains from pair 19 infecting the blood, cerebrospinal fluid, and urine of the newborn presented the same genomic makeup, distinct from its mother counterpart sample. This suggests that there is a low level of within-host GBS diversity, even if the sampling procedures might not have captured the complete genotypic diversity within each GBS population.

Despite the random nature of genomic mutagenesis, we detected a strong bias for particular types of mutations affecting genes with known functional implications (Fig. 2B). Indeed, only three mutations (11%) altered genes with unknown functions. Furthermore, of the 21 SNPs identified, only 5 appear to be polymorphic, suggesting that most of the mutations were likely selected and fixed in the population. The polymorphic variants could also have arisen from random genetic drift following the transmission of GBS between hosts, although we cannot exclude the possibility that some of these mutations occurred during subculturing in laboratory medium. Notably, most mutations corresponded to nonsynonymous substitutions, with a 13:2 ratio in relation to synonymous changes (Fig. 2A). Moreover, more than one-third of the mutations were located within noncoding sequences (Table 2), although they comprise only 10 to 12% of the GBS genome. Experimentally, their phenotypic impact was supported by the observed change in transcription levels resulting from a mutation in the 5′ UTR of *ptsG* in mother 12, as well as from a 1-bp insertion in the promoter region of the gene encoding the surface protein Rib in child 19 (Fig. 3 and 4). Altogether, these observations reinforce the significance of the mutations found, strongly suggesting that they have undergone a process of positive selection within their host.

Bearing in mind that colonization by GBS is acknowledged to be mostly asymptomatic, we show interesting data on the possible mechanisms further enhancing the virulence of this bacterium. Overall, 14 mutant variants were detected in GBS samples from infected newborns, showing a selective adaptation and evolution of GBS during neonatal disease. Particularly in pair 11, strains contaminating the child's bloodstream underwent two mutations in 14 days. In pair 19, the infant strains were mutated at three different loci after 5 to 6 days following the child's birth (Table 2). On the contrary, strains from the newborn obtained at the time of birth or 1 day later yielded the most similarities to the corresponding mother strains, as exemplified by pairs 1, 3, 6, 8, 9, 16, and 17, in which no differences were detected. Altogether, identical pairs made up more than half of the cases of EOD analyzed and only one of the six instances of LOD that were part of this study. Considering the immaturity of the newborn immune system at the time of birth, colonization by any GBS isolate transferred from the mother might be enough to lead to an infection state. Conversely, the development of the neonate's defenses throughout the first few weeks of life would be prone to elicit selective pressure on GBS strains with any fitness advantage.

From a total of 14 mutations identified in the strains from infected infants, two independently fixed mutations were found related to the two-component response regulator gene of the *covRS* locus, a central regulator of GBS virulence (32). This points toward the presence of selective pressures in favor of the alteration of CovR regulation during neonatal infections. In GAS, a broad range of nonsynonymous substitutions in the regulator gene (*covR*), as well as small indels disrupting the sensor gene (*covS*), have been described in invasive strains (64, 65). These mutations were shown to have a profound transcription effect, increasing the

expression of virulence-associated genes (65). Recently, a systematic screening for hyperpigmented GBS strains identified the presence of molecular alterations of *covR* in four isolates (66), but none of these strains were of neonatal origin. Here, we showed in GBS a transcriptional effect of a nonsynonymous mutation in the *covRS* locus after incubation in human blood (Fig. 4). We conclude that, similarly to GAS, mutations in the *covRS* locus might be selected during the development of GBS infections in neonates, but the effect of the mutations is considerably more subtle.

Another small indel in the child samples (pair 19) was pinpointed in the poly(A) tract upstream of the promoter of the gene encoding the alpha-like Rib protein. Short sequence repeat variations in this homopolymer tract, decreasing the expression of this family of surface proteins, were shown to be selected in a mouse model of immunization (61). Furthermore, smaller amounts of antibodies against alpha and Rib proteins, present in the host, have also been reported to be associated with a higher incidence of disease (67). Therefore, alpha-like proteins have been considered potential antigens for the development of a vaccine against GBS (67–69). Here, we observed a 4-fold increase in the expression of the *rib* gene following incubation in whole human blood, leading to a high level of expression of this immunogenic protein. However, the transcription levels of *rib* were significantly lower in the mutant strain colonizing the infant than in the mother strain, especially after incubation in blood (Fig. 4). Thus, the mutation in the poly(A) tract might have allowed GBS in the child to evade the protection provided by the antibodies transmitted from the mother.

Although the main focus of our work was characterizing GBS's transition from commensal to invasive after mother-to-child transmission, particular cases of infection in the mother were also part of this study, contributing to a better understanding of the pathology of GBS outside the context of neonatal disease. In this regard, two samples obtained from women diagnosed with septicemia (pairs 12 and 18) revealed parallel mutations in sugar transport systems, namely, altering glucose- and trehalose-specific PTS enzyme-encoding genes. Owing to the unique nutritional components in the mother's bloodstream, variations in metabolism-related mechanisms might provide a better adaptation of these clones and promote their progression to disease. Experimental results hinted at a functional effect of the mutation in the gene encoding a glucose-specific PTS enzyme (*ptsG*), showing different levels of expression in the mother (mother 12) and child (child 12-2) strains recovered from blood cultures (Fig. 3).

By combining the allelic profile with the available clinical information (i.e., the sampling date and body site from which the GBS strain was isolated), we are able to hypothesize the most likely direction of GBS transmission between the mother and child within each pair (Fig. 5). In pairs 7 and 11, however, the mother and child strains belong to distantly related lineages (Fig. 1 and Table 1). Genomic analyses showed no indication of the mixed presence of both clones in either the mother or the child samples. In pair 7, both strains were obtained solely from blood cultures, so it is unknown whether the mother's vaginal tract was colonized by the strain that contaminated her child. Conversely, the distinct strains from pair 11 were obtained from the vaginal tract of the mother and a blood culture of her child, suggesting that there was no mother-to-child transmission. Therefore, GBS colonization might have occurred through contamination by another carrier. Apart from these unique situations, in five pairs (4, 12, 15, 18, and

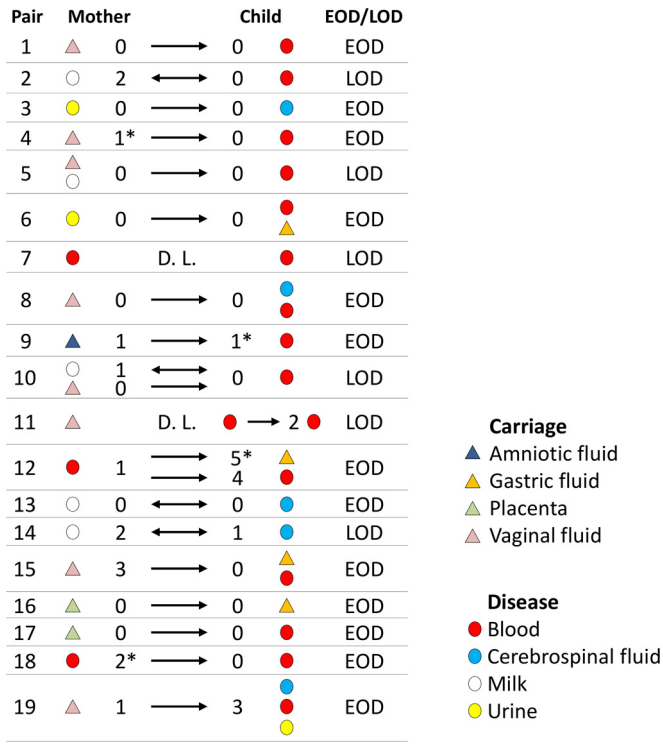


**FIG 5** Transmission pathways between GBS samples. Graphical representation of the most likely transmission route hypothesized for each pair of GBS samples. The colonization site and whether it corresponds to a situation of carriage or disease are represented by different shapes and colors according to the key. The arrows indicate the presumed direction of transfer inferred from the clinical information available together with the number of mutations determined in the mother and child GBS samples. Polymorphic variants, present in <97% of the population sampled, are indicated by asterisks. When the child was diagnosed with EOD and sampled at the time of birth, even if the mutant strain was retrieved from the mother, we assumed mother-to-child transmission. Ambiguous transmission direction is indicated by a double-headed arrow. D. L. indicates that the two strains belong to distinct lineages, so the origin of the newborn GBS clone is unknown.

19), both the collection date (at birth or just after) and the origin of the sample (blood culture or vaginal fluid) are in favor of mother-to-child transmission of GBS (Fig. 5), which is acknowledged as the most frequent form of contamination. Intriguingly, three mutant GBS strains (pairs 2, 10, and 14) were collected from the mother's milk after the child's birth (110, 16, and 33 days, respectively). This indicates that the mutations identified in the GBS sample from the mother occurred after colonization of the infant by the wild-type variants, raising the possibility that the child contaminated the milk through breastfeeding (Fig. 5). In agreement with this hypothesis, previous studies have shown that the microbial composition of human milk may be shaped by multiple sources, including the baby's oral microbiome (70). Alternatively, it is also possible that the milk was colonized independently by a GBS strain from the mother herself. In fact, in pair 10, the vaginal sample from the mother was also positive for the same GBS strain infecting her child, whereas in pairs 2 and 14, no additional body sites of the mother were sampled. If the wild-type variant initially colonized the mother's milk, these strains might have mutated independently in the mammary gland after having been transmitted to the baby. To summarize, we provide new findings regarding

the pathogenesis and etiology of neonatal GBS infections by analyzing the *in vivo* diversity and evolution of clones in carriage and disease. Ultimately, this shows that the transient colonization of neonates by GBS elicits a particular adaptation and evolution of this species that might underlie its switch from commensal to pathogen.

## ACKNOWLEDGMENTS

This work was supported by the French National Research Agency (grant ANR-13-PRTS-0006-04), by the Labex IBEID, and by the Institut de Veille Sanitaire, INSERM, University Paris Descartes. Sequencing was performed at the Pasteur Genopole, a member of France Génomique (ANR10-IBNS-09-08). A.A. is a scholar in the Pasteur-Paris University (PPU) International Ph.D. program and received a stipend from the ANR-LabEx IBEID.

We thank Laurence Ma for her help in performing the Illumina sequencing and Pedro Escoll Guerrero for his help in carrying out the experiments with human blood. We also thank Carmen Buchrieser for critical reading of the manuscript and Dimitri Desvillechabrol, Pierre-Emmanuel Douarre, and Romain Guérillot for fruitful discussions.

## REFERENCES

1. **Edmond KM, Kortsalioudaki C, Scott S, Schrag SJ, Zaidi AK, Cousens S, Heath PT.** 2012. Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. Lancet **379:**547–556. http://dx.doi.org/10.1016/S0140-6736(11)61651-6.
2. **Verani JR, McGee L, Schrag SJ.** 2010. Prevention of perinatal group B streptococcal disease: revised guidelines from CDC, 2010. MMWR Recomm Rep **59**(RR10):1–32. http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5910a1.htm.
3. **Poyart C, Reglier-Poupet H, Tazi A, Billoet A, Dmytruk N, Bidet P, Bingen E, Raymond J, Trieu-Cuot P.** 2008. Invasive group B streptococcal infections in infants, France. Emerg Infect Dis **14:**1647–1649. http://dx.doi.org/10.3201/eid1410.080185.
4. **Edwards MS, Baker CJ.** 2005. Group B streptococcal infections, p 1091–1156. *In* Remington JS, Klein JO Wilson CB, Nizet V, Maldonado Y (ed), Infectious diseases of the fetus and newborn infant. Elsevier Saunders, Philadelphia, PA.
5. **Rosinski-Chupin I, Sauvage E, Mairey B, Mangenot S, Ma L, Da Cunha V, Rusniok C, Bouchier C, Barbe V, Glaser P.** 2013. Reductive evolution in *Streptococcus agalactiae* and the emergence of a host adapted lineage. BMC Genomics **14:**252. http://dx.doi.org/10.1186/1471-2164-14-252.
6. **Rato MG, Bexiga R, Florindo C, Cavaco LM, Vilela CL, Santos-Sanches I.** 2013. Antimicrobial resistance and molecular epidemiology of streptococci from bovine mastitis. Vet Microbiol **161:**286–294. http://dx.doi.org/10.1016/j.vetmic.2012.07.043.
7. **Kalmus P, Aasmae B, Karssin A, Orro T, Kask K.** 2011. Udder pathogens and their resistance to antimicrobial agents in dairy cows in Estonia. Acta Vet Scand **53:**4. http://dx.doi.org/10.1186/1751-0147-53-4.
8. **Wyder AB, Boss R, Naskova J, Kaufmann T, Steiner A, Graber HU.** 2011. *Streptococcus* spp. and related bacteria: their identification and their pathogenic potential for chronic mastitis—a molecular approach. Res Vet Sci **91:**349–357. http://dx.doi.org/10.1016/j.rvsc.2010.09.006.
9. **Keefe GP.** 1997. *Streptococcus agalactiae* mastitis: a review. Can Vet J **38:**429–437.
10. **Mian GF, Godoy DT, Leal CAG, Yuhara TY, Costa GM, Figueiredo HCP.** 2009. Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia. Vet Microbiol **136:**180–183. http://dx.doi.org/10.1016/j.vetmic.2008.10.016.
11. **Fluegge K, Siedler A, Heinrich B, Schulte-Moenting J, Moennig MJ, Bartels DB, Dammann O, von Kries R, Berner R, German Pediatric Surveillance Unit Study.** 2006. Incidence and clinical presentation of invasive neonatal group B streptococcal infections in Germany. Pediatrics **117:**e1139-45. http://dx.doi.org/10.1542/peds.2005-2481.
12. **Kalliola S, Vuopio-Varkila J, Takala AK, Eskola J.** 1999. Neonatal group B streptococcal disease in Finland: a ten-year nationwide study. Pediatr Infect Dis J **18:**806–810. http://dx.doi.org/10.1097/00006454-199909000-00012.
13. **Neto MT.** 2008. Group B streptococcal disease in Portuguese infants younger than 90 days. Arch Dis Child Fetal Neonatal Ed **93:**F90–F93.
14. **Zangwill KM, Schuchat A, Wenger JD.** 1992. Group B streptococcal disease in the United States, 1990: report from a multistate active surveillance system. MMWR CDC Surveill Summ **41:**25–32.
15. **Libster R, Edwards KM, Levent F, Edwards MS, Rench MA, Castagnini LA, Cooper T, Sparks RC, Baker CJ, Shah PE.** 2012. Long-term outcomes of group B streptococcal meningitis. Pediatrics **130:**e8-15. http://dx.doi.org/10.1542/peds.2011-3453.
16. **Bedford H, de Louvois J, Halket S, Peckham C, Hurley R, Harvey D.** 2001. Meningitis in infancy in England and Wales: follow up at age 5 years. BMJ **323:**533–536. http://dx.doi.org/10.1136/bmj.323.7312.533.
17. **Gibson RL, Nizet V, Rubens CE.** 1999. Group B streptococcal beta-hemolysin promotes injury of lung microvascular endothelial cells. Pediatr Res **45:**626–634. http://dx.doi.org/10.1203/00006450-199905010-00003.
18. **Nizet V, Kim KS, Stins M, Jonas M, Chi EY, Nguyen D, Rubens CE.** 1997. Invasion of brain microvascular endothelial cells by group B streptococci. Infect Immun **65:**5074–5081.
19. **Hood M, Janney A, Dameron G.** 1961. Beta hemolytic streptococcus group B associated with problems of the perinatal period. Am J Obstet Gynecol **82:**809–818.
20. **Baker CJ, Barrett FF.** 1974. Group B streptococcal infections in infants—importance of various serotypes. JAMA **230:**1158–1160.
21. **Boswihi SS, Udo EE, Al-Sweih N.** 2012. Serotypes and antibiotic resistance in group B streptococcus isolated from patients at the Maternity Hospital, Kuwait. J Med Microbiol **61:**126–131. http://dx.doi.org/10.1099/jmm.0.035477-0.
22. **Tazi A, Disson O, Bellais S, Bouaboud A, Dmytruk N, Dramsi S, Mistou MY, Khun H, Mechler C, Tardieux I, Trieu-Cuot P, Lecuit M, Poyart C.** 2010. The surface protein HvgA mediates group B streptococcus hypervirulence and meningeal tropism in neonates. J Exp Med **207:**2313–2322. http://dx.doi.org/10.1084/jem.20092594.
23. **Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, Davies HD.** 2009. Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada. J Clin Microbiol **47:**1143–1148. http://dx.doi.org/10.1128/JCM.01424-08.
24. **Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, Kunst F, Glaser P, Rusniok C, Crook DW, Harding RM, Bisharat N, Spratt BG.** 2003. Multilocus sequence typing system for group B *Streptococcus*. J Clin Microbiol **41:**2530–2536. http://dx.doi.org/10.1128/JCM.41.6.2530-2536.2003.
25. **Brochet M, Rusniok C, Couve E, Dramsi S, Poyart C, Trieu-Cuot P, Kunst F, Glaser P.** 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. http://dx.doi.org/10.1073/pnas.0803654105.
26. **Da Cunha V, Davies MR, Douarre PE, Rosinski-Chupin I, Margarit I, Spinali S, Perkins T, Lechat P, Dmytruk N, Sauvage E, Ma L, Romi B, Tichit M, Lopez-Sanchez MJ, Descorps-Declere S, Souche E, Buchrieser C, Trieu-Cuot P, Moszer I, Clermont D, Maione D, Bouchier C, McMillan DJ, Parkhill J, Telford JL, Dougan G, Walker MJ, DEVANI Consortium, Holden MTG, Poyart C, Glaser P.** 2014. *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. Nat Commun **5:**4544 http://dx.doi.org/10.1038/ncomms5544.
27. **Jones AL, Knoll KM, Rubens CE.** 2000. Identification of *Streptococcus agalactiae* virulence genes in the neonatal rat sepsis model using signature-tagged mutagenesis. Mol Microbiol **37:**1444–1455. http://dx.doi.org/10.1046/j.1365-2958.2000.02099.x.
28. **Chmouryguina I, Suvorov A, Ferrieri P, Cleary PP.** 1996. Conservation of the C5a peptidase genes in group A and B streptococci. Infect Immun **64:**2387–2390.
29. **Lindahl G, Stalhammar-Carlemalm M, Areschoug T.** 2005. Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. Clin Microbiol Rev **18:**102–127. http://dx.doi.org/10.1128/CMR.18.1.102-127.2005.
30. **Spellerberg B, Pohl B, Haase G, Martin S, Weber-Heynemann J, Lutticken R.** 1999. Identification of genetic determinants for the hemolytic activity of *Streptococcus agalactiae* by ISS1 transposition. J Bacteriol **181:**3212–3219.
31. **Graham MR, Smoot LM, Migliaccio CAL, Virtaneva K, Sturdevant DE, Porcella SF, Federle MJ, Adams GJ, Scott JR, Musser JM.** 2002. Virulence control in group A streptococcus by a two-component gene regulatory system: global expression profiling and *in vivo* infection modeling.

Proc Natl Acad Sci U S A **99:**13855–13860. http://dx.doi.org/10.1073/pnas.202353699.

32. **Lamy MC, Zouine M, Fert J, Vergassola M, Couve E, Pellegrini E, Glaser P, Kunst F, Msadek T, Trieu-Cuot P, Poyart C.** 2004. CovS/CovR of group B streptococcus: a two-component global regulatory system involved in virulence. Mol Microbiol **54:**1250–1268. http://dx.doi.org/10.1111/j.1365-2958.2004.04365.x.
33. **Santi I, Grifantini R, Jiang SM, Brettoni C, Grandi G, Wessels MR, Soriani M.** 2009. CsrRS regulates group B *Streptococcus* virulence gene expression in response to environmental pH: a new perspective on vaccine development. J Bacteriol **191:**5387–5397. http://dx.doi.org/10.1128/JB.00370-09.
34. **Price EP, Sarovich DS, Mayo M, Tuanyok A, Drees KP, Kaestli M, Beckstrom-Sternberg SM, Babic-Sternberg JS, Kidd TJ, Bell SC, Keim P, Pearson T, Currie BJ.** 2013. Within-host evolution of *Burkholderia pseudomallei* over a twelve-year chronic carriage infection. mBio **4**(4): e00388-3.
35. **Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, Iqbal Z, Rimmer AJ, Cule M, Ip CLC, Didelot X, Harding RM, Donnelly P, Peto TE, Crook DW, Bowden R, Wilson DJ.** 2012. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. Proc Natl Acad Sci U S A **109:**4550–4555. http://dx.doi.org/10.1073/pnas.1113219109.
36. **Marvig RL, Sommer LM, Molin S, Johansen HK.** 2015. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat Genet **47:**57–64. http://dx.doi.org/10.1038/ng.3148.
37. **Paterson GK, Harrison EM, Murray GGR, Welch JJ, Warland JH, Holden MTG, Morgan FJE, Ba X, Koop G, Harris SR, Maskell DJ, Peacock SJ, Herrtage ME, Parkhill J, Holmes MA.** 2015. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. Nat Commun **6:**6560–6560. http://dx.doi.org/10.1038/ncomms7560.
38. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
39. **Poyart C, Tazi A, Reglier-Poupet H, Billoet A, Tavares N, Raymond J, Trieu-Cuot P.** 2007. Multiplex PCR assay for rapid and accurate capsular typing of group B streptococci. J Clin Microbiol **45:**1985–1988. http://dx.doi.org/10.1128/JCM.00159-07.
40. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.
41. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome Analysis ToolKit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:**1297–1303. http://dx.doi.org/10.1101/gr.107524.110.
42. **Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA.** 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics **11**(1110)**:**11.10.1–11.10.33. http://dx.doi.org/10.1002/0471250953.bi1110s43.
43. **DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ.** 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet **43:**491–498. http://dx.doi.org/10.1038/ng.806.
44. **Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, Kishony R.** 2014. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nat Genet **46:**82–87.
45. **Deatherage DE, Barrick JE.** 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using *breseq*. Methods Mol Biol **1151:**165–188. http://dx.doi.org/10.1007/978-1-4939-0554-6_12.
46. **Rosinski-Chupin I, Sauvage E, Sismeiro O, Villain A, Da Cunha V, Caliot M-E, Dillies M-A, Trieu-Cuot P, Bouloc P, Lartigue M-F, Glaser P.** 2015. Single nucleotide resolution RNA-seq uncovers new regulatory mechanisms in the opportunistic pathogen *Streptococcus agalactiae*. BMC Genomics **16:**419. http://dx.doi.org/10.1186/s12864-015-1583-4.
47. **Moszer I, Jones LM, Moreira S, Fabry C, Danchin A.** 2002. SubtiList: the reference database for the *Bacillus subtilis* genome. Nucleic Acids Res **30:**62–65. http://dx.doi.org/10.1093/nar/30.1.62.
48. **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.** 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. Mol Microbiol **45:**1499–1513. http://dx.doi.org/10.1046/j.1365-2958.2002.03126.x.
49. **Leekitcharoenphon P, Kaas RS, Thomsen MCF, Friis C, Rasmussen S, Aarestrup FM.** 2012. snpTree—a web-server to identify and construct SNP trees from whole genome sequence data. BMC Genomics **13**(Suppl 7)**:**S6. http://dx.doi.org/10.1186/1471-2164-13-S7-S6.
50. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
51. **Oggioni MR, Trappetti C, Kadioglu A, Cassone M, Iannelli F, Ricci S, Andrew PW, Pozzi G.** 2006. Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. Mol Microbiol **61:**1196–1210. http://dx.doi.org/10.1111/j.1365-2958.2006.05310.x.
52. **Numata S, Nagata M, Mao H, Sekimizu K, Kaito C.** 2014. CvfA protein and polynucleotide phosphorylase act in an opposing manner to regulate *Staphylococcus aureus* virulence. J Biol Chem **289:**8420–8431. http://dx.doi.org/10.1074/jbc.M114.554329.
53. **Clements MO, Eriksson S, Thompson A, Lucchini S, Hinton JCD, Normark S, Rhen M.** 2002. Polynucleotide phosphorylase is a global regulator of virulence and persistency in *Salmonella enterica*. Proc Natl Acad Sci U S A **99:**8784–8789. http://dx.doi.org/10.1073/pnas.132047099.
54. **Rosenzweig JA, Chromy B, Echeverry A, Yang J, Adkins B, Plano GV, McCutchen-Maloney S, Schesser K.** 2007. Polynucleotide phosphorylase independently controls virulence factor expression levels and export in *Yersinia* spp. FEMS Microbiol Lett **270:**255–264. http://dx.doi.org/10.1111/j.1574-6968.2007.00689.x.
55. **Postma PW, Lengeler JW, Jacobson GR.** 1993. Phosphoenolpyruvate-carbohydrate phosphotransferase systems of bacteria. Microbiol Rev **57:**543–594.
56. **Brown NL, Stoyanov JV, Kidd SP, Hobman JL.** 2003. The MerR family of transcriptional regulators. FEMS Microbiol Rev **27:**145–163. http://dx.doi.org/10.1016/S0168-6445(03)00051-2.
57. **Ellison DW, Miller VL.** 2006. Regulation of virulence by members of the MarR/SlyA family. Curr Opin Microbiol **9:**153–159. http://dx.doi.org/10.1016/j.mib.2006.02.003.
58. **Wilkinson SP, Grove A.** 2006. Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. Curr Issues Mol Biol **8:**51–62.
59. **Pailhories H, Quentin R, Lartigue MF.** 2013. The transcription of the *neuD* gene is stronger in serotype III group B streptococci strains isolated from cerebrospinal fluid than in strains isolated from vagina. FEMS Microbiol Lett **349:**71–75.
60. **Elbein AD, Pan YT, Pastuszak I, Carroll D.** 2003. New insights on trehalose: a multifunctional molecule. Glycobiology **13:**17R-27R. http://dx.doi.org/10.1093/glycob/cwg047.
61. **Puopolo KM, Madoff LC.** 2003. Upstream short sequence repeats regulate expression of the alpha C protein of group B Streptococcus. Mol Microbiol **50:**977–991. http://dx.doi.org/10.1046/j.1365-2958.2003.03745.x.
62. **Roca AI, Cox MM.** 1990. The RecA protein—structure and function. Crit Rev Biochem Mol Biol **25:**415–456. http://dx.doi.org/10.3109/10409239009090617.
63. **Santi I, Scarselli M, Mariani M, Pezzicoli A, Masignani V, Taddei A, Grandi G, Telford JL, Soriani M.** 2007. BibA: a novel immunogenic bacterial adhesin contributing to group B Streptococcus survival in human blood. Mol Microbiol **63:**754–767.
64. **Horstmann N, Sahasrabhojane P, Suber B, Kumaraswami M, Olsen RJ, Flores A, Musser JM, Brennan RG, Shelburne SA, III.** 2011. Distinct single amino acid replacements in the control of virulence regulator protein differentially impact streptococcal pathogenesis. PLoS Pathog **7**(10)**:**e1002311. http://dx.doi.org/10.1371/journal.ppat.1002311.
65. **Cole JN, Barnett TC, Nizet V, Walker MJ.** 2011. Molecular insight into invasive group A streptococcal disease. Nat Rev Microbiol **9:**724–736. http://dx.doi.org/10.1038/nrmicro2648.

66. **Lupo A, Ruppen C, Hemphill A, Spellerberg B, Sendi P.** 2014. Phenotypic and molecular characterization of hyperpigmented group B streptococci. Int J Med Microbiol **304:**717–724. http://dx.doi.org/10.1016/j.ijmm.2014.05.003.
67. **Larsson C, Lindroth M, Nordin P, Stalhammar-Carlemalm M, Lindahl G, Krantz I.** 2006. Association between low concentrations of antibodies to protein alpha and Rib and invasive neonatal group B streptococcal infection. Arch Dis Child Fetal Neonatal Ed **91:**F403–F408. http://dx.doi.org/10.1136/adc.2005.090472.
68. **Kvam AI, Mavenyengwa RT, Radtke A, Maeland JA.** 2011. *Streptococcus agalactiae* alpha-like protein 1 possesses both cross-reacting and alp1-specific epitopes. Clin Vaccine Immunol **18:**1365–1370. http://dx.doi.org/10.1128/CVI.05005-11.
69. **Maeland JA, Bevanger L, Lyng RV.** 2004. Antigenic determinants of alpha-like proteins of *Streptococcus agalactiae*. Clin Diagn Lab Immunol **11:**1035–1039.
70. **Jeurink PV, van Bergenhenegouwen J, Jimenez E, Knippels LMJ, Fernandez L, Garssen J, Knol J, Rodriguez JM, Martin R.** 2013. Human milk: a source of more life than we imagine. Benef Microbes **4:**17–30. http://dx.doi.org/10.3920/BM2012.0040.