JOURNAL OF BACTERIOLOGY, Oct. 2001, p. 5709–5717 Vol. 183, No. 19
0021-9193/01/$04.00+0 DOI: 10.1128/JB.183.19.5709–5717.2001

Genome of the Bacterium *Streptococcus pneumoniae* Strain R6

JOANN HOSKINS,[1] WILLIAM E. ALBORN, JR.,[1] JEFFREY ARNOLD,[2] LARRY C. BLASZCZAK,[1] STANLEY BURGETT,[2] BRADLEY S. DEHOFF,[2] SHAWN T. ESTREM,[1] LORI FRITZ,[2] DONG-JING FU,[2] WENDY FULLER,[2] CHAD GERINGER,[2] RAYMOND GILMOUR,[1] JENNIFER S. GLASS,[1] HAMID KHOJA,[1] ANGELIKA R. KRAFT,[1] ROBERT E. LAGACE,[3] DONALD J. LEBLANC,[1] LINDA N. LEE,[1] ELLIOT J. LEFKOWITZ,[4] JIN LU,[1] PATTI MATSUSHIMA,[1] SCOTT M. MCAHREN,[2] MARGARET MCHENNEY,[2] KEVIN MCLEASTER,[2] CHRISTOPHER W. MUNDY,[2] THALIA I. NICAS,[1] FRANKLIN H. NORRIS,[2] MARYJEANNE O'GARA,[1] ROBERT B. PEERY,[1] GREGORY T. ROBERTSON,[1] PAMELA ROCKEY,[2] PEI-MING SUN,[1] MALCOLM E. WINKLER,[1] YONG YANG,[1] MICHELLE YOUNG-BELLIDO,[1] GENSHI ZHAO,[1] CHRISTOPHER A. ZOOK,[1] RICHARD H. BALTZ,[1] S. RICHARD JASKUNAS,[1] PAUL R. ROSTECK, JR.,[2] PAUL L. SKATRUD,[1] AND JOHN I. GLASS[1]*

Infectious Diseases Research and Clinical Investigation[1] and Biological Research Technologies and Proteins,[2] Eli Lilly and Company, Indianapolis, Indiana 46285; Incyte Genomics, Palo Alto, California, 94304[3]; and Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294[4]

Received 20 June 2001/Accepted 13 July 2001

***Streptococcus pneumoniae* is among the most significant causes of bacterial disease in humans. Here we report the 2,038,615-bp genomic sequence of the gram-positive bacterium *S. pneumoniae* R6. Because the R6 strain is avirulent and, more importantly, because it is readily transformed with DNA from homologous species and many heterologous species, it is the principal platform for investigation of the biology of this important pathogen. It is also used as a primary vehicle for genomics-based development of antibiotics for gram-positive bacteria. In our analysis of the genome, we identified a large number of new uncharacterized genes predicted to encode proteins that either reside on the surface of the cell or are secreted. Among those proteins there may be new targets for vaccine and antibiotic development.**

Worldwide, approximately 1.1 million deaths annually are attributed to *Streptococcus pneumoniae* infection (22), accounting for 9% of all deaths in underdeveloped countries (37). *S. pneumoniae* disease is not limited to the developing world. Despite the availability of a broad arsenal of antibiotics and a vaccine, *S. pneumoniae* remains one of the top 10 causes of death in the United States (22). Furthermore, nearly one-third of the *S. pneumoniae* isolates obtained from patients in the United States are resistant to penicillin (11, 45) and the incidence of strains resistant to multiple antibiotics is increasing, making infections caused by this organism more difficult to treat.

S. pneumoniae is a gram-positive coccus and a member of the lactic acid bacteria, so named for their primary metabolic byproduct. The lactic acid bacteria include the lactococci, a group important in food and dairy industries, and the genera *Enterococcus* and *Streptococcus*. Bacteria belonging to the genus *Streptococcus* live in association with animal hosts, as either pathogenic or commensal organisms. Human pathogens include the beta-hemolytic species, such as *Streptococcus pyogenes* (Lancefield group A) and *Streptococcus agalactiae* (group B), as well as the human cariogenic species *Streptococcus mutans*. A number of commensal species of streptococci can occasionally cause opportunistic infections. *S. pneumoniae* (also known as pneumococcus or *Diplococcus pneumoniae*) is the major cause of acute bacterial pneumonia and otitis media. *S. pneumoniae* is also a transient commensal, colonizing the throat and upper respiratory tract of 40% of humans. *S. pneumoniae* isolates vary in their polysaccharide capsule, and at least 90 different capsule types have been identified. Specific capsule types are associated with the capacity to cause severe disease.

To aid the search for new therapies, we determined the entire genomic DNA sequence of *S. pneumoniae* strain R6. *S. pneumoniae* R6 is a descendant of the type 2 capsule S (smooth or encapsulated) clinical isolate used by Avery and coworkers to demonstrate the genetic function of DNA (2), and it is used worldwide as a standard laboratory strain. The lack of a polysaccharide capsule in R6 renders it avirulent and a safe strain with which to work. The essential utility of the strain is its genetic malleability.

MATERIALS AND METHODS

Bacterial strain. The *S. pneumoniae* R6 isolate was obtained from Alexander Tomasz (Rockefeller Institute, New York, N.Y.). The strain is *hex*$^+$, not a *hex* mutant as had been reported previously (3). The parental *S. pneumoniae* strain for R6 is R36A, which is a nonencapsulated strain derived from the capsular type 2 clinical isolate strain D39. R36A has multiple interruptions in the type 2 capsular locus inherited from D39 (21). Rollin Hotchkiss assayed single R36A colonies for competence in transformation. *S. pneumoniae* R6 was selected based on a high capacity to be transformed to penicillin resistant by using DNA from a laboratory-constructed isolate of penicillin-resistant *S. pneumoniae*. The sequenced isolate of *S. pneumoniae* R6 is available from the American Type Culture Collection (ATCC BAA-255).

Genome sequencing. Genomic DNA was isolated from bacteria grown in brain heart infusion medium (Becton Dickinson, Franklin Lakes, N.J.). The purification process included multiple phenol extractions, ethanol precipitations, and spoolings. DNA was sheared, size fractionated, and used to create plasmid and

* Corresponding author. Mailing address: Infectious Diseases Research, Lilly Research Labs, Eli Lilly and Company, Indianapolis, IN 46285-0438. Phone: (317) 277-0143. Fax: (317) 276-1743. E-mail: glass_john_i@lilly.com.

fosmid libraries. Clones from those libraries were end sequenced using both dye-primer and dye-terminator DNA sequencing methods. In the random shotgun phase of the project, ≈44,000 sequences were obtained. Gaps were closed either by sequencing spanning PCR products or by directly sequencing from the ends of contigs using custom primers and genomic DNA as a template (18). DNA sequences were analyzed and assembled using PHRED, PHRAP, and CONSED (http://www.phrap.org/) (13, 17). Insertion sequences and rRNA operons present in multiple copies created sequence assembly problems because no single DNA sequence covered the entire repetitive element. We developed a high-scoring-pairs algorithm to correctly assemble contigs flanked by these large repetitive elements (unpublished data). The sequence assembly was confirmed by a combination of Southern blotting, PCR, and comparison of the electrophoretically measured insert sizes to map locations of the end sequences from fosmid and plasmid inserts.

One open reading frame (ORF), encoding a hypothetical surface protein (spr0075), was predicted by this assembly to contain five copies of a 456-bp nearly perfect repeat. Southern blot analysis of this region of the genome suggested that there were seven copies of this repeat present within this gene (data not shown). We predicted that this gene should be approximately 912 bp larger than indicated; however, the reported sequence for the complete genome did not include the additional predicted but unsequenced 912 bp.

Sequence analysis and annotation. Annotation of the *S. pneumoniae* genome was performed by utilizing a combination of programs for gene prediction, similarity searching, and functional assignment. The information from these analyses was imported into a relational database based upon Microsoft SQL Server. The user interface for this database was a series of web pages that accessed the SQL Server database and allowed us to query available analysis data. Additional pages allowed us to directly hand-annotate the individual gene records, thus allowing us to refine start sites and add functional descriptions and notes. This web-based client interface was developed using Microsoft Active Server Pages technology to directly query and update the database records. Basic sequence analysis tools were provided by the Genetics Computer Group (GCG) package of programs (Wisconsin Package version 10.0; GCG, Madison, Wis.).

ORF identification. Determination of potential protein-encoding sequences utilized Glimmer (http://www.tigr.org/softlab/glimmer/glimmer.html) (9) to create organism-specific ORF models that could then be used to search the entire genome for ORFs matching the predictive models. The genome was first arranged such that the initial base of the ATG start codon of the putatively identified *dnaA* gene was base number 1 of the forward strand. Each predicted ORF was then assigned an *S. pneumoniae* identification number. ORF spr0001 was assigned to the putative *dnaA* gene, and each subsequent ORF was then numbered consecutively according to its left-most base (the start codon for ORFs on the forward strand, and the stop codon for genes on the reverse strand).

Gene assignments. BLAST searches were performed on all predicted ORFs using a blastp search of amino acid similarities to sequences in the GenBank nonredundant protein database. The BLAST data were parsed using the blast modules of the BioPerl tool kit (http://www.bioperl.org) and then imported into SQL server tables for analysis. In addition to BLAST similarity searching, we also tentatively identified functional domains within the *S. pneumoniae* ORFs by searching for similarities to the Prosite motif library (20) and the Blocks database of protein families (19). Programs from the GCG package provided composition and hydrophobicity analyses along with scanning for potential signal peptide domains. The results of these additional analyses allowed us to refine the gene assignments initially made with BLAST. Furthermore, alignments with known proteins provided assistance with start-codon prediction.

The results of all of these searches were used to provide putative identification of each *S. pneumoniae* ORF when a significant hit between an *S. pneumoniae* sequence and GenBank sequence was found. A combination of computer-aided gene prediction along with human inspection of each gene record was then used to finalize gene assignments for each *S. pneumoniae* ORF.

RNA identification. To identify genomic sequences that code for tRNAs, the set of programs that encompass the software package tRNAscan-SE (24) was used. rRNAs were identified by their similarity to the corresponding genes in the Ribosome Database Project sequence database (25). The sequences for tmRNA (51), the 4.5S signal recognition particle (53), and RNase P (28) were also identified based upon sequence similarities with known representatives of these RNA genes.

Nucleotide sequence accession number. The genomic sequence was assigned accession number AE007317 in the GenBank data base. The annotated genome and supplementary data are available on the World Wide Web at http://www.lilly.com/s.pneumoniae.

RESULTS AND DISCUSSION

The *S. pneumoniae* single circular chromosome of 2,038,615 bp (40% G+C content) contains 2,043 predicted protein coding regions and 73 noncoding RNA genes, which include four rRNA operons. The genomic origin of replication has not been experimentally identified in *S. pneumoniae*; however, based on the presence of clusters of DnaA boxes and other genomic features, we hypothesize that the *S. pneumoniae* origin of DNA replication is upstream of *dnaA,* which is gene spr0001 in our nomenclature system (16).

As anticipated from the work of Iannelli and colleagues (21), relative to strain D39, the encapsulated strain from which R6 was ultimately derived, we noted a 7,504-bp deletion within the ≈18-kbp region that encodes the capsule biosynthesis genes. This deletion results in the absence of seven complete genes as well as the 3′ end of *cps*2A and the 5′ end of *cps*2H.

Other than genes associated with capsule synthesis, the genes encoding several putative virulence functions are present in the R6 genome (Table 1). These include the genes for previously described *S. pneumoniae* surface proteins, secreted proteins and bacteriocins, and all previously reported two-component response regulator systems (13 potential histidine protein kinases and response regulator pairs plus an unpaired 14th response regulator) (46).

Drug resistance via efflux pumps is an important contributor to virulence in *Staphylococcus aureus* and other gram-positive pathogens. Although *S. pneumoniae* contains 14 genes that are possible antibiotic efflux pumps (Table 2), these efflux pump genes may not be significant contributors to *S. pneumoniae* virulence. Antibiotic extrusion is not as common a source of resistance in *S. pneumoniae* as it is in *S. aureus* (e.g., quinolones). *S. pneumoniae* is not intrinsically resistant to most classes of agents, and among the exceptions (e.g., aminoglycosides and quinolones) resistance is not the result of drug efflux pumps.

Surface proteins. Surface proteins are of special interest because of their potential role in virulence and their possible utility in vaccine development and also because of their potential accessibility to antimicrobial agents. The R6 genome includes single copies of the previously described virulence-associated genes, including four that encode proteins that are under study as vaccine candidates (PspA, PsaA, CbpA, and pneumolysin) (5).

Based on sequence analyses, we predict that a large number of proteins either reside on the *S. pneumoniae* cell surface or are secreted from the cell. These proteins include 471 with predicted signal peptide sequences, 109 possessing lipoprotein lipid attachment sites, and 10 that are recognized by choline-binding domains, an unusual means of surface attachment found in *S. pneumoniae* (Fig. 1). We could predict no function for ≈23% of these potential surface-located or secreted proteins, which likely play roles in pneumococcal cell surface biology.

In *S. aureus*, a transpeptidase called sortase anchors exported proteins containing an LPXTG motif followed by a C-terminal hydrophobic domain and a charged tail to the cell wall peptidoglycan (29, 41). Although sortases are predicted to be present in all gram-positive bacteria, previously no sortase ortholog had been identified in *S. pneumoniae,* nor could we

TABLE 1. Genes found in R6-encoded proteins that have been studied for a role in *S. pneumoniae* virulence or as protective antigens to *S. pneumoniae*[a]

Gene no.	Gene name	Description
spr0121	*pspA*	Surface protein (choline binding), functions in inhibition of complement activation and may involve lactoferrin binding (30, 48)
spr0286	*hysA*	Hyaluronidase (30)
spr0674	*sodA*	Manganese cofactored superoxide dismutase (30)
spr0867	*lytB*	Endo-beta-*N*-acetylglucosaminidase (15)
spr0884	*ppmA*	Surface-located proteinase maturation protein (33)
spr1042	*iga*	Immunoglobulin A1 protease (49)
spr1323	*nox (nadH)*	NADH oxidase (1)
spr1492, spr1493	*psaB, psaC*	With *psaA*, part of manganese ABC transporter operon (30)
spr1494	*psaA*	Adhesin, manganese ABC transporter substrate-binding protein (30)
spr1536, spr1531	*nanA, nanB*	Neuraminidase, sialidase (30)
spr1707	*amiA*	Oligopeptide ABC transporter substrate-binding protein (33)
spr1739	*ply (tacY)*	Pneumolysin (thiol-activated cytolysin) (30)
spr1754	*lytA*	Autolysin (*N*-acetylmuramoyl-L-alanine amidase) (30)
spr1995	*cbpA (pcpA)*	Surface protein (choline binding), functions in adherence and immunoglobulin A inactivation (30)
spr0040[b]	*thmA*	Bacteriocin-like amphipathic pore-forming peptide precursor (26)
spr0059[b]	*bgaC*	Beta-galactosidase 3 (52)
spr0351,[b] spr0337[b]	*pcpC, cbpF (pcpC1)*	Surface protein (choline binding), unknown function
spr0462[b]	*blpS (iR1)*	Regulator of bacteriocin production (39)
spr0465[b]	*iP*	Bacteriocin-like peptide, regulator (39)

[a] Other genes with a role in virulence or pathogenesis include two-component regulator systems (13 paired histidine-kinase and response regulators, one orphan response regulator) (46); genes for peptidoglycan biosynthesis; genes for synthesis of teichoic acid; and genes for choline synthesis, export, and attachment (*pkc*, choline kinase; *proWX* and *proV*, choline transport; *licD1* and *licD2*, choline metabolism).

[b] Published gene(s) for which function or motif suggests roles in pathogenesis but which lack experimental verification.

identify one using BLAST searches. Using a Smith and Waterman algorithm (43), we determined that a single *S. pneumoniae* R6 gene, spr1098, likely coded for a sortase. Furthermore, we identified 13 genes that coded for proteins containing the LPXTG motif and other sortase substrate features. Six of these proteins are known to be present on the cell surface, while seven are novel and currently categorized as hypothetical with no known function (Fig. 1). Our observations about sortase in strain R6 conflict with a report by Pallen and colleagues, who identified four sortase-like protein genes and "many" potential sortase substrates in the genome of a virulent *S. pneumoniae* strain with a type 4 capsule (34).

Lipoteichoic acid (LTA) is another cell surface component that contributes to the bacterium's interaction with the human host. Biochemical studies have not detected the presence of D-alanine, a key component of LTA in many organisms, in *S.*

TABLE 2. Possible/probable drug efflux pumps encoded by *S. pneumoniae* R6

Gene no.	Gene name[a]	Description
spr0137	ABC-NBD&MSP	This is similar to the *E. coli msbA* gene, which is proposed by Milton Saier, Jr., and Ian Paulsen to function in drug efflux (http://www.biology.ucsd.edu/~ipaulsen/transport/)
spr0144	Conserved hypothetical	Possible macrolide efflux permease of the major facilitator superfamily
spr0610	ABC-NBD/truncation	Probably part of a drug resistance ABC transporter pump
spr0875	*pmrA*	Major facilitator superfamily multidrug resistance efflux pump
spr0880	*tehB*	Multidrug resistance; K^+-tellurite ethidium and proflavin transport (there is no *tehA* gene in *S. pneumoniae* R6)
spr0971	*mefE*	ABC transporter membrane-spanning permease for macrolide efflux
spr1023	ABC-MSP	Possible macrolide efflux pump
spr1052	MATE	Hypothetical multi-antimicrobial extrusion (MATE) family transporter
spr1183	ABC-NBD/truncation	Possible multidrug efflux pump, *clyB*/*comA* paralog, missing NH_2 half
spr1203	ABC-NBD&MSP	Hypothetical ABC transporter possibly involved in multidrug resistance or toxin secretion
spr1352	*bta*	Bacterocin transport accessory protein
spr1378–81	ABC-MSP/truncation	This is a fragmented ABC transporter gene similar to the *S. aureus pepT* efflux pump
spr1734–6	ABC-NBD/truncation	This is a fragmented ABC transporter gene similar to the *Mycobacterium tuberculosis* MTb1273 efflux pump
spr1877	MATE	Hypothetical multi-antimicrobial extrusion (MATE) family transporter

[a] Abbreviations: ABC-NDB, ABC transporter nucleotide-binding domain; ABC-NBD&MSP, ABC transporter nucleotide-binding domain and membrane-spanning permease; ABC-MSP, ABC transporter membrane-spanning permease.

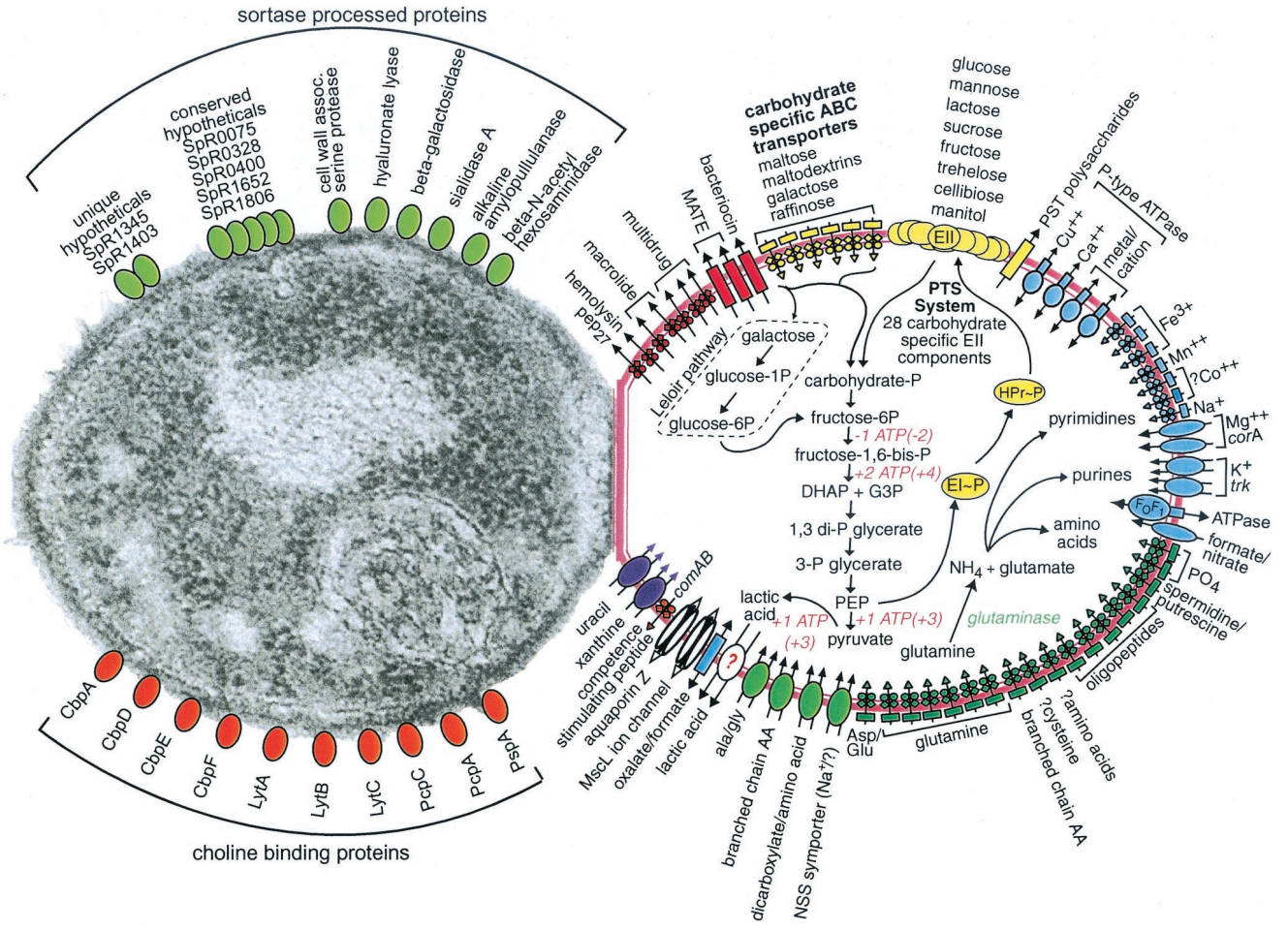

FIG. 1. *S. pneumoniae* substrate transport, carbohydrate and glutamine metabolism, and selected categories of cell surface proteins. Transporters are shown in the cell on the right and are grouped by substrate specificity: multidrug and peptide exporters (red), carbohydrates (yellow), cations (blue), anions and amino acids (green), nucleosides, purines, and pyrimidines (purple), and other substrates (white). Question marks indicate uncertainty in the identity of the genes, direction of transport, or substrate. Permeases are drawn as rectangles; porins are drawn as cylinders; ATPases are drawn as ovals overlapping rectangles; other transporters are drawn as ovals; and ABC transporter elements are depicted as circles for nucleotide-binding proteins, diamonds for membrane-spanning permeases, and rectangles for the substrate-binding proteins. The 21 hypothetical ABC transporters for which no substrate could be predicted are not shown in the figure. Glutaminase, an enzyme we expected to find but could not identify, is listed in green italics. The glycolytic pathway leading to lactate is shown along with the pathway for glutamine to nitrogen-containing compounds. (Additional material on carbohydrate metabolism is at http://www.lilly.com/s.pneumoniae.) The ATP consumption and production values listed are for monosaccharides; ATP values for disaccharide catabolism are in parentheses. The choline-binding proteins are autolysin (LytA), endo-beta-*N*-acetylglucosaminidase (LytB), a surface protein involved in adherence and immunoglobulin A inactivation (CbpA), a putative lactoferrin-binding protein (PspA), and several surface proteins of unknown function (CbpD, PcpA, PcpC, PcpC1). Micrograph reprinted from reference 47 with permission of the publisher.

pneumoniae LTA (14). In conflict with the apparent absence of D-alanine in *S. pneumoniae* LTA, we found an apparently complete *dlt*ABCD operon that is homologous to those responsible for the addition of D-alanine to LTA in *Bacillus subtilis* and in *Lactobacillus casei*. Previously we suggested that this operon may be silent or defective or that these genes may be active under specific physiological conditions (3). Gene expression studies (data not shown) revealed expression of mRNA from each of these genes under normal laboratory growth conditions. The precise role of the *dlt*ABCD operon in the biology of *S. pneumoniae* remains unknown, although inactivation of this operon in *S. mutans* confers increased acid sensitivity (4).

Competence. *S. pneumoniae* competence, i.e., its natural capacity to take up DNA, has been studied in detail and a number of competence-specific operons have been identified (23). *S. pneumoniae* R6 contains all of the genes induced during competence as noted by Lee and Morrison (23), including two identical copies of *comX* (each adjacent to a ribosomal operon). Additional genes reported to be induced during competence, but whose role in this process remains unknown, are also present (36, 40). These 49 putative competence genes are grouped in 30 apparent operons and are found mostly on the leading strands extending away from the putative origin of replication (as are ≈80% of all *S. pneumoniae* genes).

As might be predicted as a consequence of the capacity of *S. pneumoniae* to take up DNA, its genome is littered with genes that are apparently derived from other bacteria. Horizontal gene transfer is clearest for those genes that have been found

only in gram-negative bacterial genomes. There are 40 ORFs that are similar to genes in gram-negative bacteria and that have not been found in other gram-positive genome sequences. This is not surprising, because *S. pneumoniae* occupies the same niche in the human respiratory system as several gram-negative species. Additionally, at least 2% of *S. pneumoniae* genes are significantly truncated relative to orthologous genes characterized in other bacteria (Table 3). Many of the deletions are at the 5′ ends of the ORFs, which suggests that the ORFs may be nonfunctional remnants of their parental genes. This incidence may also be a consequence of competence. Coding regions may be missing from these genes because only part of the ORFs were acquired during the assimilation of foreign DNA, or because the genes were not essential to the pneumococcus and mutations are of no consequence. Transporters are the most frequently truncated genes. Among that set are five ORFs that are similar to genes encoding drug efflux pumps.

Comparative genomics and metabolism. In most respects, the *S. pneumoniae* gene complement is very similar to that of the prototypic gram-positive bacterium *B. subtilis*. More than 53% of the *S. pneumoniae* genes have highly similar counterparts in the *B. subtilis* genome (Fig. 2). Systems for cell division, DNA replication and repair, translation, cell wall biosynthesis, and some central catabolic and biosynthetic pathways are basically the same as in *B. subtilis*. Major cellular systems and features that are notably different include energy metabolism, transport, amino acid biosynthesis, transcription termination, intracellular proteases, and the presence of three large sets of *S. pneumoniae*-specific repetitive elements in the genome.

As is characteristic of the lactic acid bacteria, *S. pneumoniae* is a nutritionally fastidious facultative anaerobe requiring a complex medium for growth. This bacterium obtains energy strictly via fermentation and is incapable of respiratory metabolism, either aerobically or anaerobically, as is true of all streptococcal species (38). The only nutrients from which the streptococci can obtain sufficient energy to support growth and cell division are carbohydrates, which are oxidized to pyruvate via glycolysis (with the exception of a few species that can ferment arginine). We identified a large set of genes that encode enzymes necessary for transport of at least 12 different carbohydrates into the cell and for their subsequent conversion to an intermediate in glycolysis.

S. pneumoniae R6, as expected, encodes all genes necessary for the oxidation of carbohydrates to pyruvate via glycolysis and would be expected to reoxidize most, if not all, of the NADH produced by the reduction of pyruvic acid to lactic acid. R6 contains genes for the synthesis of phosphotransacetylase, acetokinase, and NADH oxidase, which would allow it to convert pyruvate to acetate with concomitant production of an additional ATP, and the reoxidation of NADH (38). Although fermentation is the least energy efficient of oxidative processes, *S. pneumoniae* did not maximize this energy production by exclusively using the phosphoenolpyruvate-dependent phosphotransferase system to import carbohydrates. Five sugar species are imported by using energetically less-efficient ABC transporters. We found no genes that might encode cation antiporters of sugars, although we identified several amino acid/cation symport systems (Fig. 1). All genes necessary for synthesis of the major ATPase of lactic acid bacteria, the F_0F_1-ATPase, are present. This proton pump works at the expense of ATP, but it can also serve as an ATP synthase, as well as serving as the major regulator of intracellular pH among lactic acid bacteria. We did not find the genes required for a complete electron transport chain that might be associated with either aerobic or anaerobic respiration.

No lactic acid bacterial species encodes a complete tricarboxylic acid (TCA) cycle, and the *S. pneumoniae* R6 genome contains none of the 18 genes comprising this aerobic oxidative pathway. In other organisms, including those without complete TCA cycles, some of the TCA enzymes also have roles in the synthesis of certain amino acid precursors. As a result, *S. pneumoniae* R6 is incapable of synthesizing aspartate (and hence lysine, methionine, threonine, and isoleucine) from oxaloacetate, nor can it synthesize glutamate (and hence arginine) via α-ketoglutarate. A defined medium developed specifically for *S. pneumoniae* contains those amino acids, so the incomplete biosynthetic pathways were expected. We were unable to identify complete pathways for the synthesis of glycine, histidine, and leucine, all of which are included in the *S. pneumoniae* defined medium. Valine is also included in the *S. pneumoniae* defined medium, so identification of an apparently complete pathway for valine biosynthesis was unexpected (42).

The presence in *S. pneumoniae* of an ortholog to the *Cercospora nicotianae pdx1* gene suggests *S. pneumoniae* may have a pyridoxal biosynthetic pathway (12). The biosynthesis pathways for other required cofactors (biotin, choline, pantothenate) are either incomplete or absent. Presence of partial pathways for the synthesis of many of these amino acids and cofactors in *S. pneumoniae* is not surprising. In many cases these enzymes make possible the conversion of molecules imported into the cell into other necessary metabolic components. Glutamine is an example of this type of metabolic conversion. Although *S. pneumoniae* cannot make the starting material for glutamate, α-ketoglutarate, it does encode the enzymes needed to utilize glutamine as a nitrogen source. In *S. mutans*, glutamine has been shown to be a principal source of nitrogen (8). There are 22 genes encoding the elements of 7 different ABC transporters predicted to transport glutamine. That represents 10% of the transport genes in *S. pneumoniae*. The allocation of the *S. pneumoniae* genome capacity to glutamine transport suggests that glutamine is also needed for more than its role as a component of proteins.

Hydrogen peroxide is produced by *S. pneumoniae* through the action of pyruvate oxidase (SpxB) under conditions of aerobic growth. This may be a mechanism by which the pneumococcus inhibits the growth of other common pathogens of the human upper respiratory tract such as *Haemophilus influenzae, Moraxella catarrhalis,* and *Neisseria meningitidis*, which are infrequently cocultured with *S. pneumoniae* from patient samples (35). Unlike those gram-negative species, *S. pneumoniae* has the capacity to resist oxidative stress caused by H_2O_2. In *Escherichia coli*, oxidative stress induces the expression of a set of ≈30 proteins under the transcriptional control of OxyR (6). Based on their similarity to OxyR, either spr0593 or spr0828, which are bacterial regulatory proteins of the LysR family, might possibly regulate the *S. pneumoniae* enzymes synthesized in response to H_2O_2. These include superoxide dismutase, glutathione reductase, glutaredoxin, DNA-binding

TABLE 3. Truncated and/or fragmented genes in *S. pneumoniae* R6

Gene no.	Gene name[a]	Description
spr0105	ABC-NBP&MSP-truncation	ABC transporter similar to *comA*
spr0106	ABC-NBP&MSP-truncation	ABC transporter similar to *comA*
spr0223	ABC-SBP-truncation	ABC transporter substrate-binding protein (iron transport)
spr0224	ABC-SBP-truncation	ABC transporter substrate-binding protein (iron transport)
spr0299	Conserved hypothetical	
spr0300	Conserved hypothetical	
spr0324	Transposase G-truncation	Uncharacterized transposase
spr0346	*alkD*-truncation	DNA alkylation repair enzyme
spr0347	*alkD*-truncation	DNA alkylation repair enzyme
spr0349	*cbpG*-truncation	Choline-binding protein G
spr0466	Conserved hypothetical	
spr0467	Conserved hypothetical	
spr0468	Conserved hypothetical	
spr0469	Conserved hypothetical	
spr0523	Transposase H-truncation	Uncharacterized transposase
spr0535	*glnQ*-truncation	ABC transporter ATP-binding protein (glutamine)
spr0536	*glnQ*-truncation	ABC transporter ATP-binding protein (glutamine)
spr0610	ABC-NBD-truncation	ABC transporter ATP-binding protein (unknown substrate)
spr0620	ABC-SBP-truncation	ABC transporter substrate-binding protein (unknown substrate)
spr0621	ABC-SBP-truncation	ABC transporter substrate-binding protein (unknown substrate)
spr0646	*bgl*-truncation	Phospho-beta-gluco or galactosidase
spr0737	Conserved hypothetical	
spr0852	*ald*-truncation	Alanine dehydrogenase
spr0853	*ald*-truncation	Alanine dehydrogenase
spr0854	*ald*-truncation	Alanine dehydrogenase
spr0910	*phtE*-truncation	Pneumococcal histidine triad protein E precursor
spr0957	Relaxase-truncation	Tn*5252*, relaxase
spr0958	Relaxase-truncation	Tn*5252*, relaxase
spr1033	*serB*-truncation	Phosphoserine phosphatase
spr1077	*lacX*-truncation	Function unknown
spr1095	*metY*-truncation	*O*-Acetylhomoserine sulfhydrylase
spr1096	*metY*-truncation	*O*-Acetylhomoserine sulfhydrylase
spr1133	*leuD*-truncation	3-Isopropylmalate dehydratase small subunit
spr1183	ABC-NBD-truncation	ABC transporter ATP-binding protein (possibly multidrug efflux)
spr1186	*npl*-truncation	*N*-Acetylneuraminate lyase subunit
spr1187	*npl*-truncation	*N*-Acetylneuraminate lyase subunit
spr1285	*spnIR*-truncation	Type II restriction endonuclease, uncharacterized
spr1286	*spnIR*-truncation	Type II restriction endonuclease, uncharacterized
spr1378	ABC-MSP-truncation	
spr1379	ABC-MSP-truncation	
spr1380	ABC-MSP-truncation	
spr1381	ABC-MSP-truncation	
spr1520	*npl*-truncation	*N*-Acetylneuraminate lyase subunit
spr1521	*npl*-truncation	*N*-Acetylneuraminate lyase subunit
spr1683	*galE*-truncation	UDP-galactose 4-epimerase
spr1734	ABC-NBD-truncation	ABC transporter ATP-binding protein (unknown substrate)
spr1736	ABC-NBD-truncation	ABC transporter ATP-binding protein (unknown substrate)
spr1772	Conserved hypothetical	
spr1871	*pcp*-truncation	Pyrrolidone carboxyl peptidase
spr1872	*pcp*-truncation	Pyrrolidone carboxyl peptidase
spr1900	Conserved hypothetical	
spr1955	*arcA*-truncation	Arginine deiminase
spr1956	*arcA*-truncation	Arginine deiminase
spr1984	Transposase H	Uncharacterized transposase
spr1985	Transposase H	Uncharacterized transposase
spr1989	*glpD*-truncation	Glycerol 3-phosphate dehydrogenase
spr1990	*glpD*-truncation	Glycerol 3-phosphate dehydrogenase
spr1999	*clpC*-truncation	Class III stress response-related ATPase
spr2000	*clpC*-truncation	Class III stress response-related ATPase
spr2036	Conserved hypothetical	
spr2037	Conserved hypothetical	
spr2038	Conserved hypothetical	
spr2039	Conserved hypothetical	

[a] Abbreviations: ABC-NBD, ABC transporter nucleotide binding domain; ABC-SBP, ABC transporter substrate-binding protein; ABC-MSP, ABC transporter membrane-spanning permease; ABC-NBD&MSP, ABC transporter nucleotide-binding domain and membrane-spanning permease.

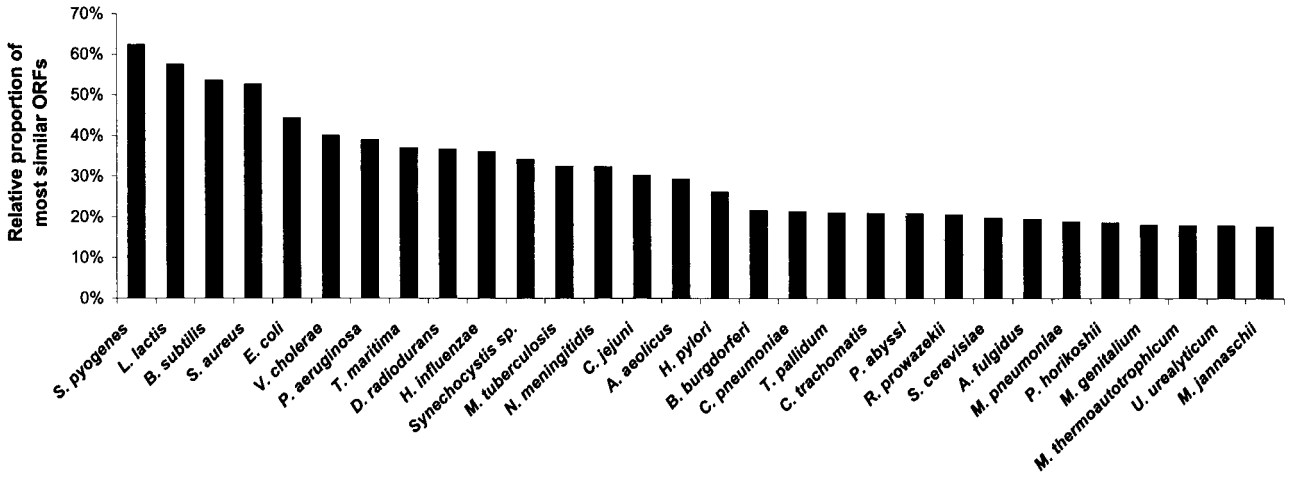

FIG. 2. Comparison of the predicted *S. pneumoniae* R6 ORFs with those of other completely sequenced genomes. All *S. pneumoniae* R6 ORFs were searched against other genomes with BLAST2. The percentage of *S. pneumoniae* R6 ORFs with significant similarity ($E \leq 10^{-12}$) to predicted ORFs in the other genomes is shown.

stress protein, and two thioredoxin reductases. Although there are also genes for several peroxidases that can be used to ameliorate oxidative stress, *S. pneumoniae* does not encode catalase.

Bacterial energy-dependent intracellular proteases perform a variety of tasks, possibly including that of the proteasome, which degrades aberrant and nonfunctional proteins in eukaryotes and archaea (10). *S. pneumoniae* R6 possesses single copies of the genes encoding the ClpP and FtsH proteases, but it is notably deficient of the genes encoding HslV and the ubiquitous Lon protease. While some bacteria with relatively large genomes, such as *E. coli* and *B. subtilis*, encode all four of these energy-dependent proteases, most eubacteria encode only a subset (10). The *S. pneumoniae* energy-dependent protease gene set appears to be characteristic of the gram-positive genera *Enterococcus, Streptococcus*, and *Staphylococcus*, but not of the mycoplasmas.

Repetitive elements. DNA sequences from three classes of repetitive elements, BOX, RUP, and IS, comprise >3% of the *S. pneumoniae* genome. These kinds of repetitive elements make up more of the *S. pneumoniae* genome than of any other bacterial genome sequenced to date. Functions for some of these sequences are controversial. The BOX elements are predicted to form stable secondary structures that may serve as the binding site for a protein responsible for modulating the expression of downstream genes (27). Insertion of heterologous DNA into the BOX element upstream of the *comA* gene produces *S. pneumoniae* incapable of competence (27). Additionally, Weiser showed that insertion of a BOX element upstream of a locus apparently involved in phase variation increases expression of downstream genes encoding the opacity phenotype (50). The 107-bp RUP elements, predicted to form stable secondary structures, are proposed to be active insertion elements transactivated by the transposase of IS*630*-Spn*1* (32).

Almost all BOX and RUP elements are entirely in intergenic spaces. We hypothesized that analysis of the locations of the BOX and RUP elements relative to the transcriptional orientation of the genes surrounding them might offer clues about their potential regulatory roles. In *S. pneumoniae* between adjacent genes in the same transcriptional orientation, the boundaries of transcriptional units are often unclear; accordingly, promoters and transcription termination signals are difficult to identify. Between pairs of adjacent genes oriented 5′ end to 5′ end on opposite strands of the chromosome, or at least in the vicinity of those genes, there are likely to be pairs of transcriptional promoters. Likewise, there must be a transcription termination signal or signals between pairs of adjacent genes oriented 3′ end to 3′ end on opposite strands of the chromosome (factor-independent transcription termination signals have been identified in streptococci that can function bidirectionally [44]). Almost 3 times as many BOX elements and 1.5-fold more RUP elements are located between the genes oriented 3′ to 3′ than between genes oriented 5′ to 5′, and the insertion of IS elements flanking some of these elements may artificially deflate those ratios (Table 4). This suggests a role for the RUP and BOX elements in transcription termination. Another possible function is suggested by the fact that secondary structures that RUP and BOX elements would assume at the 3′ ends of mRNAs are more complex than those observed for other factor-independent transcription termina-

TABLE 4. Transcriptional orientations of genes flanking BOX and RUP repetitive elements

Element	No. (%) of elements[a]		
	→BOX or RUP→	→BOX or RUP←	←BOX or RUP→
BOX	80 (4)	26 (13)	9 (4)[b]
RUP	49 (3)	21 (11)	14 (7)[b]

[a] The arrows indicate the transcriptional orientation of the genes flanking the BOX and RUP elements. The numbers of BOX and RUP elements in each orientation are listed. The parenthetic percentage values are the fractions of all intergenic regions in each transcriptional orientation that contain BOX or RUP elements.

[b] These numbers may be artificially high. Three of these 9 BOX elements and 6 of these 14 RUP elements are adjacent either to small hypothetical genes predicted with low confidence or to insertion elements. Were those genes falsely predicted or the IS elements not present, these BOX and RUP elements would be in the →BOX/RUP→ orientation relative to their flanking genes instead of the ←BOX/RUP→ orientation.

tion signals (27, 32), and *S. pneumoniae* does not encode rho factor. These elements might enhance gene expression by either stabilizing mRNAs or serving as binding sites for regulatory proteins.

Previous analyses indicate numerous IS elements were present in the DNA of various strains of *S. pneumoniae*. The genome of R6 contains at least 60 complete or partial copies of 10 different IS elements, representing the families IS*L3*, IS*5*, IS*630*, IS*3*, IS*30*, and IS*605*. We identified three novel IS elements. We did not find an IS*1202*, which was previously identified in a progenitor of the R6 strain, *S. pneumoniae* D39 (31). Most of the copies of the IS elements appear to be only remnants, as only seven possess the expected full-length sequences of putative transposase genes. The remaining copies all contain frameshifts, stop codons, or both within the ORF, and many have substantial amino acid substitutions, suggesting that they are no longer active. It is possible that these inactive insertion elements still play an important role in the evolution of this genome. For example, they may provide regions of homology that are sites for homologous recombination in the acquisition of genes from related organisms carrying these same insertion sequences but different flanking genes (7).

The capacity to identify all potential genes within this pathogen should greatly facilitate the identification of novel targets for antibiotic discovery as well as new candidates for vaccine development. This process will be significantly enhanced by the comparison of the *S. pneumoniae* R6 sequence to that of pathogenic strains of *S. pneumoniae* (www.tigr.org and http://genome.microbio.uab.edu/strep/). These comparisons, in concert with genetic and gene expression studies, should catalyze expansion of *S. pneumoniae* biology.

ACKNOWLEDGMENTS

We thank Amy Hahn, Travis Bennett, Lara Braverman, Joanne Dyer, Bruce Glover, Ken Holstein, Dennis Howell, Ivan Jenkins, Tammy Jones, Rebecca Leonard, Melud Nabavi, Regine Porter, Patricia Solenberg, and Angie Wu for their technical assistance, Ron Swanson for advice on the sequencing project, and Janet Yother and Dalai Yan for their advice on the manuscript.

Principal authors JoAnn Hoskins and John Glass contributed equally to this project.

This work was supported by Eli Lilly and Company and the Incyte Genomics, Inc. Pathoseq database program.

REFERENCES

1. **Auzat, I., S. Capuy-Regaud, G. LeBras, D. DosSantos, A. D. Ogunniyi, I. LeThonas, J.-R. Garel, G. C. Paton, and M. C. Tombe.** 1999. The NADH oxidase of *Streptococcus pneumoniae*: its involvement in competence and virulence. Mol. Microbiol. **34:**1018–1028.
2. **Avery, O. T., C. MacLeod, and M. McCarty.** 1944. Studies on the chemical nature of the substance inducing transformation of the pneumococcal types. J. Exp. Med. **79:**137–158.
3. **Baltz, R. H., F. H. Norris, P. Matsushima, B. S. DeHoff, P. Rockey, G. Porter, S. Burgett, R. Peery, J. Hoskins, L. Braverman, I. Jenkins, P. Solenberg, M. Young, M. A. McHenney, P. L. Skatrud, and P. R. Rosteck, Jr.** 1998. DNA sequence sampling of the *Streptococcus pneumoniae* genome to identify novel targets for antibiotic development. Microb. Drug Resist. **4:**1–9.
4. **Boyd, D. A., D. G. Cvitikovitch, A. S. Bleiweis, M. Y. Kiriukhin, D. V. Debabov, F. C. Neuhaus, and I. R. Hamilton.** 2000. Defects in D-alanyl-lipoteichoic acid synthesis in *Streptococcus mutans* results in acid sensitivity. J. Bacteriol. **182:**6055–6065.
5. **Briles, D. E., S. Hollingshead, A. Brooks-Walter, G. S. Nabors, L. Ferguson, M. Schilling, S. Gravenstein, P. Braun, J. King, and A. Swift.** 2000. The potential to use PspA and other pneumococcal proteins to elicit protection against pneumococcal infection. Vaccine **18:**1707–1711.
6. **Christman, M. F., G. Storz, and B. N. Ames.** 1989. OxyR, a positive regulator of hydrogen peroxide-inducible genes in *Escherichia coli* and *Salmonella typhimurium*, is homologous to a family of bacterial regulatory proteins. Proc. Natl. Acad. Sci. USA **86:**3484–3488.
7. **Claverys, J. P., M. Prudhomme, I. Mortier-Barriere, and B. Martin.** 2000. Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? Mol. Microbiol. **35:**251–259.
8. **Dashper, S. G., P. F. Riley, and E. C. Reynolds.** 1995. Characterization of glutamine transport in *Streptococcus mutans*. Oral Microbiol. Immunol. **10:**183–187.
9. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
10. **De Mot, R., I. Nagy, J. Walz, and W. Baumeister.** 1999. Proteasomes and other self-compartmentalizing proteases in prokaryotes. Trends Microbiol. **7:**89–92.
11. **Doern, G. V., A. B. Brueggemann, H. Huynh, and E. Wingert.** 1999. Antimicrobial resistance with *Streptococcus pneumoniae* in the United States, 1997–98. Emerg. Infect. Dis. **5:**757–765.
12. **Ehrenshaft, M., P. Bilski, M. Y. Li, C. F. Chignell, and M. E. Daub.** 1999. A highly conserved sequence is a novel gene involved in *de novo* vitamin B6 biosynthesis. Proc. Natl. Acad. Sci. USA **96:**374–378.
13. **Ewing, B., L. Hillier, M. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185.
14. **Fischer, W.** 1988. Physiology of lipoteichoic acids in bacteria, p. 233–296. *In* A. H. Rose and D. W. Tempest (ed.), Advances in microbial physiology, vol. 29. Academic Press, London, England.
15. **Garcia, P., A. C. Martin, and R. Lopez.** 2000. Bacteriophages of *Streptococcus pneumoniae*: a molecular approach, p. 211–222. *In* A. Tomasz (ed.), *Streptococcus pneumoniae*: molecular biology and mechanisms of disease. Mary Ann Liebert, Inc., Larchmont, N.Y.
16. **Gasc, A. M., P. Giammarinaro, S. Richter, and M. Sicard.** 1998. Organization around the *dnaA* gene of *Streptococcus pneumoniae*. Microbiology **144:**433–439.
17. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
18. **Heiner, C. R., K. L. Hunkapiller, S. M. Chen, J. I. Glass, and E. Y. Chen.** 1998. Sequencing multimegabase-template DNA with BigDye terminator chemistry. Genome Res. **8:**557–561.
19. **Henikoff, S., J. G. Henikoff, and S. Pietrokovski.** 1999. Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. Bioinformatics **15:**471–479.
20. **Hofmann, K., P. Bucher, L. Falquet, and A. Bairoch.** 1999. The PROSITE database, its status in 1999. Nucleic Acids Res. **27:**215–219.
21. **Iannelli, F., B. J. Pearce, and G. Pozzi.** 1999. The type 2 capsule locus of *Streptococcus pneumoniae*. J. Bacteriol. **181:**2652–2654.
22. **Klein, D. L.** 1999. Pneumococcal disease and the role of conjugate vaccines. Microb. Drug Resist. **5:**147–157.
23. **Lee, M. S., and D. A. Morrison.** 1999. Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. J. Bacteriol. **181:**5004–5016.
24. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
25. **Maidak, B. L., J. R. Cole, C. T. Parker, Jr., G. M. Garrity, N. Larsen, B. Li, T. G. Lilburn, M. J. McCaughey, G. J. Olsen, R. Overbeek, S. Pramanik, T. M. Schmidt, J. M. Tiedje, and C. R. Woese.** 1999. A new version of the RDP (Ribosomal Database Project). Nucleic Acids Res. **27:**171–173.
26. **Marciset, O., M. C. Jeronimus-Stratingh, B. Mollet, and B. Poolman.** 1997. Thermophilin 13, a nontypical antilisterial poration complex bacteriocin, that functions without a receptor. J. Biol. Chem. **272:**14277–14282.
27. **Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, et al.** 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res. **20:**3479–3483.
28. **Massire, C., L. Jaeger, and E. Westhof.** 1998. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. J. Mol. Biol. **279:**773–793.
29. **Mazmanian, S. K., G. Liu, H. Ton-That, and O. Schneewind.** 1999. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. Science **285:**760–763.
30. **Mitchell, T. J.** 2000. Virulence factors and the pathogenesis of disease caused by *Streptococcus pneumoniae*. Res. Microbiol. **151:**413–419.
31. **Morona, J. K., A. Guidolin, R. Morona, D. Hansman, and J. C. Paton.** 1994. Isolation, characterization, and nucleotide sequence of IS*1202*, an insertion sequence of *Streptococcus pneumoniae*. J. Bacteriol. **176:**4437–4443.
32. **Oggioni, M. R., and J. P. Claverys.** 1999. Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. Microbiology **145:**2647–2653.
33. **Overweg, K., A. Kerr, M. Sluijter, M. H. Jackson, T. J. Mitchell, A. P. de Jong, R. de Groot, and P. W. Hermans.** 2000. The putative proteinase maturation protein A of *Streptococcus pneumoniae* is a conserved surface

protein with potential to elicit protective immune responses. Infect. Immun. **68:**4180–4188.
34. **Pallen, M. J., A. C. Lam, M. Antonio, and K. Dunbar.** 2001. An embarrassment of sortases—a richness of substrates? Trends Microbiol. **9:**97–101.
35. **Pericone, C. D., K. Overweg, P. W. Hermans, and J. N. Weiser.** 2000. Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. Infect. Immun. **68:**3990–3997.
36. **Peterson, S., R. T. Cline, H. Tettelin, V. Sharov, and D. A. Morrison.** 2000. Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. J. Bacteriol. **182:**6192–6202.
37. **Poland, G. A.** 1999. The burden of pneumococcal disease: the role of conjugate vaccines. Vaccine **17:**1674–1679.
38. **Poolman, B.** 1993. Energy transduction in lactic acid bacteria. FEMS Microbiol. Rev. **12:**125–148.
39. **Reichmann, P., and R. Hakenbeck.** 2000. Allelic variation in a peptide-inducible two-component system of *Streptococcus pneumoniae*. FEMS Microbiol. Lett. **190:**231–236.
40. **Rimini, R., J. Birger, G. Feeger, T. Roberts, M. de Francesco, A. Gozzi, F. Faggioni, E. Domenici, D. Wallace, N. Frandssen, and A. Polissi.** 2000. Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. Mol. Microbiol. **36:**1279–1292.
41. **Schneewind, O., D. Mihaylova-Petkov, and P. Model.** 1993. Cell wall sorting signals in surface proteins of gram-positive bacteria. EMBO J. **12:**4803–4811.
42. **Sicard, A. M.** 1964. A new synthetic medium for *Diplococcus pneumoniae* and its use for the study of reciprocal transformations at the *amiA* locus. Genetics **50:**31–44.
43. **Smith, T. F., and M. S. Waterman.** 1981. Identification of common molecular subsequences. J. Mol. Biol. **147:**195–197.
44. **Steiner, K., and H. Malke.** 1997. Primary structure requirements for in vivo activity and bidirectional function of the transcription terminator shared by the oppositely oriented *skc*/*rel-orf1* genes of *Streptococcus equisimilis* H46A. Mol. Gen. Genet. **255:**611–618.
45. **Thornsberry, C., P. T. Ogilvie, H. P. Holley, Jr., and D. F. Sahm.** 1999. Survey of susceptibilities of *Streptococcus pneumoniae, Haemophilus influenzae*, and *Moraxella catarrhalis* isolates to 26 antimicrobial agents: a prospective U.S. study. Antimicrob. Agents Chemother. **43:**2612–2623.
46. **Throup, J. P., K. K. Koretke, A. P. Bryant, K. A. Ingraham, A. F. Chalker, Y. Ge, A. Marra, N. G. Wallis, J. R. Brown, D. J. Holmes, M. Rosenberg, and M. K. Burnham.** 2000. A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. Mol. Microbiol. **35:**566–576.
47. **Tomasz, A.** 2000. *Streptococcus pneumoniae*: molecular biology and mechanisms of disease, p. 11. Mary Ann Liebert, Inc., Larchmont, N.Y.
48. **Tu, A.-H. T., R. L. Fulgham, M. A. McCrory, D. E. Briles, and A. J. Szalai.** 1999. Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. Infect. Immun. **67:**4720–4724.
49. **Wani, J. H., J. V. Gilbert, A. G. Plaut, and J. N. Weiser.** 1996. Identification, cloning, and sequencing of the immunoglobulin A1 protease gene of *Streptococcus pneumoniae*. Infect. Immun. **64:**3967–3974.
50. **Weiser, J.** 2000. Phase variation of *Streptococcus pneumoniae*, p. 225–231. *In* V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, and J. I. Rood (ed.), Gram-positive pathogens. ASM Press, Washington, D.C.
51. **Williams, K. P.** 2000. The tmRNA website. Nucleic Acids Res. **28:**158–161.
52. **Zahner, D., and R. Hakenbeck.** 2000. The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. J. Bacteriol. **182:**5919–5921.
53. **Zwieb, C., and T. Samuelsson.** 2000. SRPDB (Signal Recognition Particle Database). Nucleic Acids Res. **28:**171–172.