Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen

Dragana Ajdić*, William M. McShan*, Robert E. McLaughlin*, Gorana Savić*, Jin Chang*, Matthew B. Carson*, Charles Primeaux*, Runying Tian†, Steve Kenton†, Honggui Jia†, Shaoping Lin†, Yudong Qian†, Shuling Li†, Hua Zhu†, Fares Najar†, Hongshing Lai†, Jim White†, Bruce A. Roe†, and Joseph J. Ferretti*‡

*Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104; and †Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019

Communicated by Roy Curtiss, Washington University, St. Louis, MO, August 19, 2002 (received for review June 25, 2002)

Streptococcus mutans **is the leading cause of dental caries (tooth decay) worldwide and is considered to be the most cariogenic of all of the oral streptococci. The genome of *S. mutans* UA159, a serotype c strain, has been completely sequenced and is composed of 2,030,936 base pairs. It contains 1,963 ORFs, 63% of which have been assigned putative functions. The genome analysis provides further insight into how *S. mutans* has adapted to surviving the oral environment through resource acquisition, defense against host factors, and use of gene products that maintain its niche against microbial competitors. *S. mutans* metabolizes a wide variety of carbohydrates via nonoxidative pathways, and all of these pathways have been identified, along with the associated transport systems whose genes account for almost 15% of the genome. Virulence genes associated with extracellular adherent glucan production, adhesins, acid tolerance, proteases, and putative hemolysins have been identified. Strain UA159 is naturally competent and contains all of the genes essential for competence and quorum sensing. Mobile genetic elements in the form of IS elements and transposons are prominent in the genome and include a previously uncharacterized conjugative transposon and a composite transposon containing genes for the synthesis of antibiotics of the gramicidin/bacitracin family; however, no bacteriophage genomes are present.**

Decades of epidemiological, biochemical, and animal studies have implicated *Streptococcus mutans* as the principal causative agent of human dental caries (tooth decay). Dental caries is one of the most common infectious diseases afflicting humans, and tends to remain untreated in many underdeveloped areas, leading to considerable suffering that is often alleviated only by the loss or extraction of the infected tooth (1). Although 200–300 bacterial species have been found associated with dental plaque, only the presence of *S. mutans* has been consistently linked with the formation of human dental caries (2). Additionally, *S. mutans* is occasionally associated with non-oral infections, principally subacute bacterial endocarditis (3). Additional information about *S. mutans* can be found in several recently published reviews (4–8).

The complete genome sequences of several Gram-positive cocci have been reported recently, including *Streptococcus pyogenes* (9), *Streptococcus pneumoniae* (10, 11), *Staphylococcus aureus* (12), and *Lactococcus lactis* (13). In this report, we present the complete sequence of another Gram-positive coccus, a Bratthall serotype c strain of *S. mutans*.

Materials and Methods

S. mutans UA159, Bratthall serotype c, was obtained from Page Caufield (University of Alabama, Birmingham) and is available through the American Type Culture Collection (ATCC 700610). The *S. mutans* genome sequence was determined using a shotgun high-throughput sequencing approach as described (9). The detailed methods are published as supporting information on the PNAS web site (www.pnas.org), and are also available at the University of Oklahoma Health Sciences Center web site (http://microgen.ouhsc.edu/) and the University of Oklahoma Advanced Center for Genome Technology web site (www.genome.ou.edu/). The *S. mutans* genome sequence is available at www.genome.ou.edu/smutans.html, www.oralgen.lanl.gov, and through GenBank (accession no. AE014133).

Results and Discussion

General Findings. The genome of *S. mutans* strain UA159 is organized in one circular chromosome (Fig. 1) of 2,030,936 base pairs. The average GC content of 36.82% is similar to that of other low GC organisms (see Table 1, which is published as supporting information on the PNAS web site). It possesses 65 tRNAs and five rRNA operons. The protein-coding sequences represent 85.66% of the genome and are organized in 1,963 ORFs (see Fig. 3), with an average gene length of 885 bases. The ORFs are arranged so that transcription is biased in the direction of DNA replication with 78% of the genes encoded on the leading strand. The average GC content of the gene-coding sequences is 37.54%. The distribution of predicted ORFs according to functional categories is presented in Table 2, which is published as supporting information on the PNAS web site. Of the total number of predicted ORFs, 63% have assigned putative function, 21% have homologs from different species, although their function is unknown, and 16% are unique to *S. mutans*. This distribution of ORFs is similar to that found in other bacterial genomes sequenced to date. Comparison of all *S. mutans* UA159 ORFs with the coding regions of several other Gram-positive organisms showed the greatest extent of similarity to *S. pneumoniae* and *S. pyogenes* (see Table 3, which is published as supporting information on the PNAS web site).

The origin of replication (*oriC*) of the *S. mutans* chromosome was identified by its comparison to the *S. pyogenes* (9) and *Bacillus subtilis* (14) *oriC*, as well as its GC nucleotide skew analysis. The location of *oriC* was near the *dnaA* gene, in the region that contains AT-rich sequences representing characteristic *dnaA* boxes. The terminus of replication (*terC*) was also identified using GC skew analysis.

Cell Division and Cell Wall Synthesis. Several clusters of genes involved in cell division were found in the UA159 genome, including those conserved in most bacteria. Additionally, there are more than 60 proteins responsible for cell envelope biogenesis, including 5 penicillin-binding proteins, 3 ABC transporters, 10 glycosyltransferases, and 6 autolysins. Autolysins are responsible for the selective removal of cell wall peptidoglycan and can be classified in different groups according to the type of peptidoglycan bond they hydrolyze (15). *S. mutans* shows enormous variability in chain length, which suggests that it has autolysins

Abbreviation: PTS, phosphotransferase system.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AE014133).

‡To whom correspondence should be addressed. E-mail: Joe-Ferretti@ouhsc.edu.

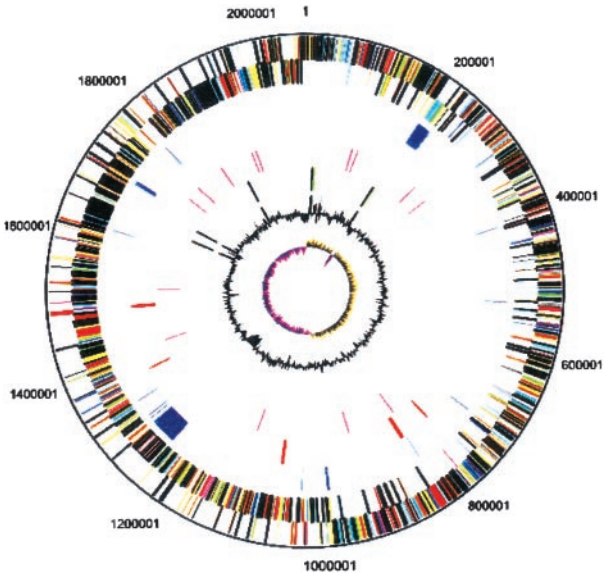


Fig. 1. Circular representation of the *S. mutans*, strain UA159 genome. The outer two circles show the position of the probable ORFs on the complementary DNA strands. The ORFs have been color coded by functional category (the color legend is presented in Fig. 3, which is published as supporting information on the PNAS web site). The remaining circles are, proceeding inward: the location of the mobile genetic elements, the ABC transporters involved in the sugar metabolism, the PTS transporters and PTS Enzyme I, the position of the ribosomal (black) and tRNA (green), the %G+C of the sequence, and the %G+C deviation by strand.

and that they are regulated. In fact, Clark named this organism "mutans" because of the observed changes in cell shape (probably due to wall remodeling) depending on growth conditions (16). Four types of predicted autolysins were found in UA159 including muramidase (amidase 4; SMU.76), amidase 2 (SMU.704) that is similar to the autolysins from *Listeria monocytogenes*, *Listeria innocua*, and *Staphylococcus caprae*, an endolysin (SMU.707) homologous to one associated with a bacteriophage from *S. pyogenes*, and the LrgB family protein (SMU.574) similar to the putative autolysin of *S. aureus*. Two more ORFs that encode a potential lytic transglycosylase (SMU.2147) and an additional LrgB family protein (SMU.1700) were also identified. However, autolysins that belong to the LytC, LytD, endopeptidase II, and lysostaphin families characteristic of bacillus and staphylococcus species were not present in UA159.

Metabolic Pathways. An overview of *S. mutans* basic metabolic pathways is presented in Fig. 2. Carbohydrate metabolism is a key survival strategy for *S. mutans*, and current knowledge about sugar metabolism of this organism combined with genome data suggests that *S. mutans* is capable of metabolizing a wider variety of carbohydrates than any other Gram-positive organism sequenced to date. Genes for transport and metabolism of glucose, fructose, sucrose, lactose, galactose (via Leloir and tagatose 6-phosphate pathways), mannose, cellobiose, β-glucosides, trehalose, maltose/maltodextrin, raffinose, ribulose, melibiose starch, isomaltosaccharides, and possibly sorbose are found in the genome. In addition to sugars, *S. mutans* is able to convert several sugar-alcohols to glycolytic intermediates, and the genes for metabolism of mannitol and sorbitol (glucitol) are also present. Among the newly identified gene clusters is one responsible for the synthesis and degradation of starch (SMU.1535–1541) that includes pullulanase. Further, two key enzymes of the nonoxidative branch of the pentose-phosphate pathway (Fig. 2), ribose-phosphate epimerase (*rpe*), and ribose-phosphate isomerase (*rpiA*) are found in UA159. This finding, in conjunction with the presence of a phosphotransferase system (PTS) transporter for pentose (possible ribulose; Fig. 2), suggests that *S. mutans* is able to synthesize and use certain pentoses. A second copy of sugar-phosphate isomerase with sequence similarity to *rpiB* from *Escherichia coli* was also found (SMU.2142).

The fermentation of carbohydrates by *S. mutans* is the principal source of energy production for the organism. A complete glycolytic pathway was found in UA159, leading to the production of pyruvate that is then reduced to various fermentation products (lactic acid, formate, ethanol, and acetate). These metabolic processes are important for oxidation of NADH, as well as synthesis of an additional ATP due to conversion of pyruvate to acetate. The enzymes responsible for pyruvate metabolism found in *S. mutans* include pyruvate dehydrogenase, pyruvate formate-lyase (the anaerobic counterpart of pyruvate dehydrogenase), phosphotransacetylase, acetate kinase, lactate dehydrogenase, alcohol dehydrogenase, and a newly identified acetoin dehydrogenase. Streptococci are incapable of aerobic respiratory metabolism (17), and the genes required for the aerobic electron transport chain were not present.

S. mutans possesses an incomplete tricarboxylic acid (TCA) cycle and the primary role of the existing TCA enzymes is most likely synthesis of amino acid precursors (18). Because most *S. mutans* strains can grow *in vitro* on minimal medium supplemented with only a few amino acids (19), and because some strains can be cultured anaerobically with ammonia as the sole nitrogen source (20), it was not surprising that all amino acid biosynthetic pathways were identified in the genome. Indeed, UA159 can grow on minimal medium devoid of amino acids if thiosulfate is provided for cysteine biosynthesis (data not shown). Analysis of the histidine operon organization revealed that in addition to eight *his* genes, two other ORFs were present, one that encodes a protein of unknown function, and another that has sequence similarity to phosphoserine phosphatase (SerB) from several organisms. Although the presence of unrelated ORFs in the *his* operon is not unusual for Gram-positive bacteria (21), the presence of a gene for the synthesis of another amino acid was not expected, and might indicate coregulation of *serB* and *his* genes.

The acidification of the local environment by the end products of metabolism inhibits many competing bacterial species, enabling *S. mutans* to maintain its niche, and incidentally causing dental caries in the host. Acid tolerance of *S. mutans* is based primarily on the presence of a membrane-bound, acid-stable, proton-translocating F_0F_1 ATPase that can maintain the intracellular pH at 7.5 (8), and all genes encoding components of this ATPase are present in the genome. In addition to this mechanism, some oral streptococci use the arginine deiminase pathway to survive a decrease in environmental pH (8). Although the enzyme arginine deiminase was not found in the genome, the other enzymes of this pathway, including ornithine carbamoyltransferase and carbamate kinase, were identified. Moreover, an amino acid antiporter different from that described for other arginine deiminase operons is a component of the same gene cluster (SMU.262–265), indicating that UA159 may utilizes an unknown amino acid (possibly agmatine) in this pathway. In addition to acid, other environmental changes can trigger a variety of adaptive responses in this organism, and the expression of numerous genes is affected following exposure to oxidative, osmotic, heat, or starvation stress (22). Genes previously demonstrated to be involved in general stress responses of *S. mutans* were all identified in UA159. In addition, several proteases possibly involved in the stress response and more than 30 other stress-related proteins are present, including cold-shock (SMU.957) and alkaline-shock (SMU.228) protein homologs. *S. mutans* lacks cytochromes, heme-containing proteins, and catalases (23), but is able to grow in the presence of oxygen, and

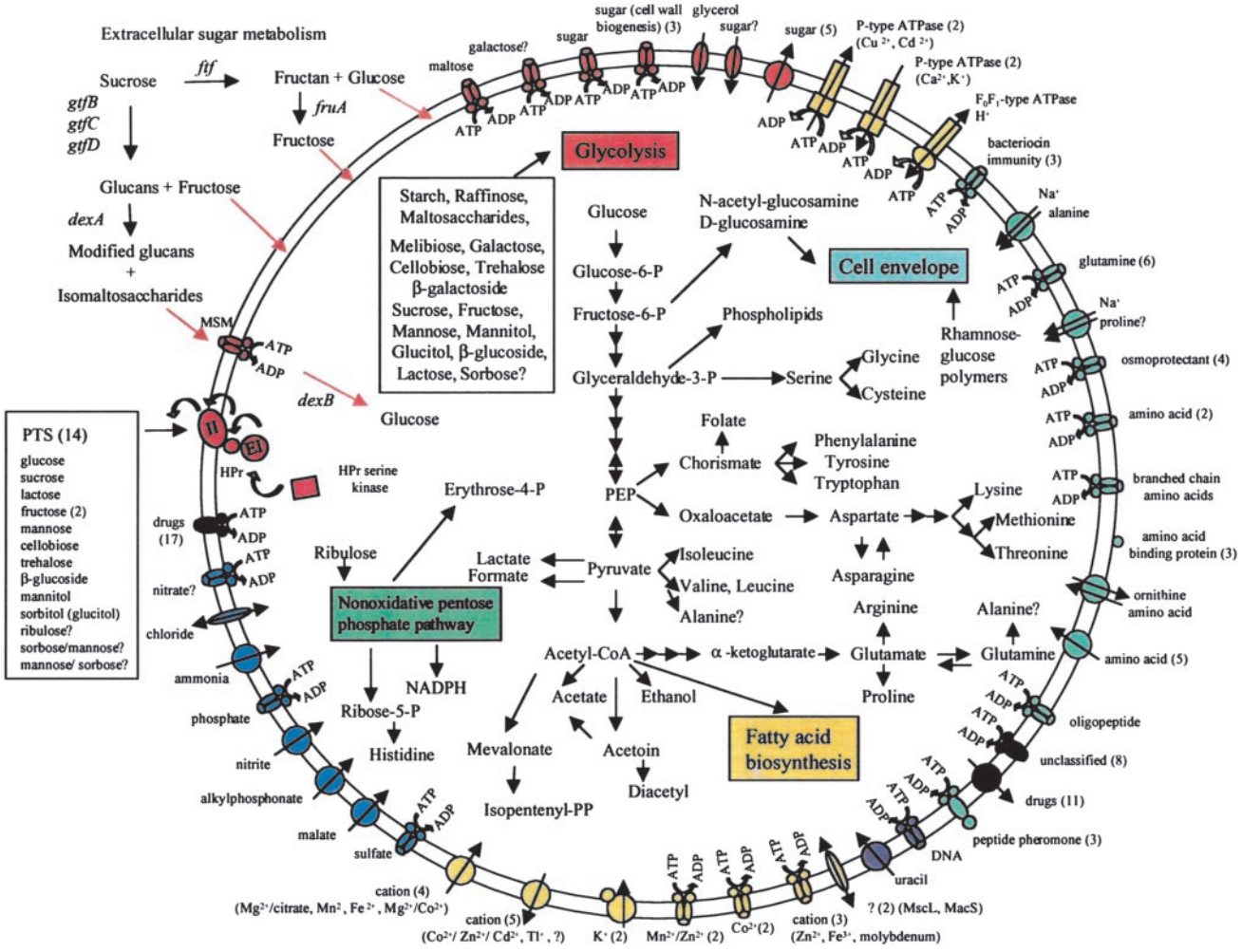
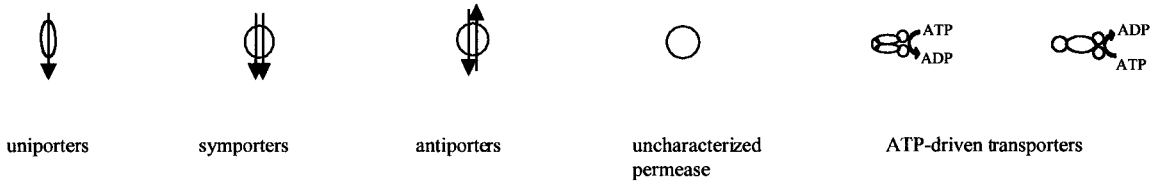


Fig. 2. Reconstruction of specific metabolic pathways and transport mechanisms in *S. mutans*. Based on the annotated genome sequence, extracellular and intracellular sugar metabolism and metabolism of organic compounds are shown. Transporters are grouped by substrate specificity: red, carbohydrates; green, amino acids/peptides; gold, inorganic cations; blue, inorganic anions; purple, DNA/nucleotides; black, drugs and unclassified. Question marks indicate uncertainty of substrate specificity. Arrows represent the direction of solute transport. Each figure shape represents the specific type of transporter:



gtfB, glucosyltransferase-I; *gtfC*, glucosyltransferase-SI; *gtfD*, glucosyltransferase-S; *dexA*, dextranase; *dexB*, dextran glucosidase; *ftf*, fructosyltransferase; *fruA*, fructanase; ATP, adenosine monophosphate; ADP, adenosine diphosphate; P, phosphate.

all genes previously reported to be involved in oxidative stress were present in UA159. In addition, a previously uncharacterized glutathione reductase (SMU.140), two thioredoxin reductases (SMU.463 and SMU.869), a glutaredoxin (SMU.669), and a thiol peroxidase (SMU.924) were also found.

Transport. *S. mutans* UA159 contains more then 280 genes associated with various transport systems, accounting for almost 15% of the total ORFs. Identified UA159 transporters are presented in Fig. 2. Most of the identified transporters are ATP dependent, a characteristic observed in other organisms lacking an electron transport chain. Three types of solute transporting ATPases are present: P-type, F-type, and ABC-type. The P-type ATPases are predicted to be responsible for the transport of calcium and potassium, as well as bacterial resistance to the toxic metals copper and cadmium. F-type ATPases (F_0F_1 ATPases) use an electrochemical gradient of H^+ or Na^+ to synthesize ATP, or they hydrolyze ATP to reverse an electrochemical gradient (24). As mentioned earlier, one H^+-transporting ATP synthase-ATPase was found in the *S. mutans* genome. ABC-type ATPases are the most abundant, with more than 60 ABC transporters in the genome, which accounts for almost 10% of the total number of *S. mutans* ORFs. About one-third of all ABC transporters are categorized as importers, whereas the rest are exporters, sug-

gesting that this organism is capable of actively exporting excess and harmful molecules. ABC transporters exhibit specificity for different substrates: amino acids, carbohydrates, oligopeptides, osmoprotectants (proline/glycine betaine, choline), inorganic ions (Fe^{3+}, Co^{2+}, Mn^{2+}, Zn^{2+}, phosphate, nitrate, sulfate, and molybdenum), bacteriocins, and DNA. Despite the fact that *S. mutans* synthesizes glutamine, it also possesses five predicted glutamine ABC transporters, emphasizing the importance of this amino acid as a principal source of nitrogen and a substrate for amino acid biosynthesis in *S. mutans*.

Three types of transport facilitators are also found in UA159. The first type specifies a pore-type mechanism known to facilitate glycerol uptake in prokaryotes (25). A second type of facilitator is represented by several solute–cation symporters with a common characteristic that Na^+ is used for generation of membrane potential for solute (usually amino acid) transport. A third type involves solute–solute antiporters, present in virtually all living organisms, and the UA159 strain possesses at least one amino acid antiporter of this type (SMU.263).

In addition to these transport mechanisms, putative transport proteins have been identified for the uptake of essential inorganic nutrients (K^+, Fe^{2+}, Mn^{2+}, nitrite, Mg^{2+}, and Co^{2+}), efflux of toxic metal ions (Co^{2+}, Zn^{2+}, Cd^{2+}, and Tl^+), and undefined molecules. In some cases, both uptake systems and efflux transporters are present for ions that are both essential and toxic (Co^{2+} and Zn^{2+}).

The uptake of sugars from the environment is a key component of *S. mutans* carbohydrate metabolism. Although at least five sugar ABC transport systems exist, including the well described multiple sugar metabolism (MSM) system (26), most sugars are transported by phosphoenolpyruvate sugar PTSs, the apparent primary sugar transport systems of Gram-positive bacteria sequenced to date. They consist of two nonspecific energy coupling components, enzyme I (SMU.675) and a heat-stable protein (HPr; SMU.674), and several sugar-specific multiprotein permeases known as enzyme II. The UA159 genome contains 14 PTSs with a variety of enzyme II domains and specificity for different sugars (Fig. 2). This strain also possesses putative carbohydrate efflux proteins (Fig. 2), presumably for actively transporting sugar intermediates that need to be excreted.

Virulence Factors. Virulence factors in *S. mutans* help protect the bacterium against possible host defenses and maintain its ecological niche in the oral cavity, while contributing to its ability to cause host damage. Probable virulence factors include adhesins, glucan-producing and -binding exoenzymes, proteases and cytokine-stimulating molecules. In addition to the factors described below, several other gene products were found in UA159 that may also contribute to the virulence of *S. mutans*, such are hemolysins, the myosin cross-reactive streptococcal antigen and its paralog, a coiled-coil myosin-like protein, and the genes for transport and assimilation of iron and phosphate.

Adhesins. *S. mutans* can adhere to salivary agglutinin, other plaque bacteria, extracellular matrix, and epithelial cell-surface receptors. Two major types of *S. mutans* adhesins mediate this attachment: cell-surface proteins and sucrose-derived glucans. The major *S. mutans* surface receptors SpaP (also known as Pac, antigen I/II, B, P1, SR, MSL-1, and IF; ref. 27) and wall-associated antigen A (WapA; ref. 28) were found in UA159. In addition, a complete operon for synthesis of cell wall polysaccharides (rhamnose-glucose polymers, RGPs; ref. 29), previously suggested to be involved in adhesion (30), was also found.

The SloC protein, an adhesin belonging to the lipoprotein receptor antigen I (LraI) family, was recently characterized (31). Members of the LraI family were found in a range of bacteria, and have a dual role: transport and adhesion (32). The presence of *lraI* operon was confirmed in UA159, and a previously uncharacterized protein was also found possessing the characteristics of a Zn-binding ABC transporter-type lipoprotein that is similar to adhesin AdcA from *S. pneumoniae* (33).

Several genes encoding known and newly characterized proteins similar to streptococcal extracellular matrix-binding proteins were also identified, including the glucan-binding proteins A (34) and C (35) that enable cell aggregation in the presence of dextran (see Table 4, which is published as supporting information on the PNAS web site). One newly characterized protein (SMU.1449) has sequence similarity to the fibronectin/fibrinogen-binding proteins from *S. pneumoniae* (36) and *S. pyogenes* (37) (66% identity and 77% similarity). Fibronectin binding of *S. mutans* has been reported recently (38), and therefore SMU.1449 might indeed be responsible for fibronectin binding in *S. mutans*. Another protein (SMU.360) was found in UA159 that showed 89% identity and 94% similarity to the *S. pyogenes* plasmin receptor (39). Interestingly, this protein is also similar to glyceraldehyde-3-phosphate dehydrogenase, suggesting that this protein might be an enzyme with plasmin-binding activity.

Exoenzymes. *S. mutans* UA159 has the genes to produce many exoenzymes involved in extracellular sucrose metabolism, including well known glucosyltransferases I, SI, and S, fructanase, fructosyltransferase, and dextranase. The *gbpB* (also known as *sagA*) gene that encodes a putative peptidoglycan hydrolase (40) important for cell wall integrity (41) and glucan binding (42) is also present. In addition to these previously described gene products, a previously uncharacterized gene encoding putative exoenzyme was found and named *gbpD*. A motif search of the GbpD protein sequence showed the presence of a putative signal peptide with a predicted cleavage site and three tandem copies of the "A" repeat identified in the glucan-binding domain of glucosyltransferase (43) and other glucan-binding proteins (44, 45). There is evidence that GbpD binds glucan (R. R. B. Russell, personal communication).

Proteases. *S. mutans* proteases contribute to virulence and are involved in the breakdown of host proteins for bacterial nutrition (46) and the direct degradation of host structural proteins (47). These streptococcal proteases might also be involved in the destruction of some components of the host immune system because a previously uncharacterized putative C3-degrading protease homolog (SMU.399) was identified. Overall, strain UA159 possesses numerous proteases, including the ATP-dependent Clp and FtsH proteases, and has at least 15 peptidases, indicating that it is capable of using the peptides generated by these proteases, as well as by those of other bacteria in plaque. Several proteases potentially involved in surface processing were also identified, including the serine protease HtrA, involved in bacterial stress responses (48), and a protease homologous to the HtpX protease from *Streptococcus gordonii* (49). In addition, several previously uncharacterized predicted surface or extracellular proteases were identified, including three Zn-dependent proteases (SMU.160, SMU.1438, and SMU.1784), two proteases related to collagenase (SMU.759 and SMU.761), a serine protease RgpF, and a membrane protease (stomatin/prohibitin homolog; SMU.235). *S. mutans* lacks a homolog of the wall-anchored cell-surface proteases (PrtP of *L. lactis* and C5a peptidase of *S. pyogenes*).

Other surface and extracellular proteins. Analysis of the genome ORFs revealed the presence of a large number of putative surface and extracellular proteins that are possibly involved in attachment and the stimulation of the immune system. Surface and secreted proteins of UA159 were detected using a search for specific motifs. There are 271 membrane, surface, and/or secreted proteins that have a predicted signal peptide, and of these, 135 have a possible cleavage site. In accordance with the large number of these proteins, five signal peptidases (type I and its homologue, type II and IV, and lipoprotein signal peptidase)

were found in the genome. Forty-seven putative lipoproteins were identified with a lipoprotein lipid attachment site motif (properly located within the first 30 amino acids; see Table 4). Further analysis using a revised lipoprotein motif (50) suggests that the actual number of *S. mutans* lipoproteins may be lower (≈27–35). In addition, five proteins were found with the tandem "A" repeats characteristic of glucan-binding domains (see Table 4). Four of these proteins were characterized before (GtfS, GtfSI, GtfI, and GbpA), and the fifth is the previously uncharacterized predicted extracellular protein, GbpD. The sortase target motif (LPxTG motif) was found in ten proteins, but is properly positioned at the C terminus in only six of the proteins: the previously reported FruA, SpaP, WapA, DexA, and GbpC, and a previously uncharacterized secreted protein, WapE, that appears to be wall associated (R. R. B. Russell, personal communication; see Table 4). The single *S. mutans* sortase responsible for anchoring of surface proteins was also identified and is an ortholog of the enzyme from several Gram-positive species (11, 51, 52).

Regulation and Signaling. The UA159 genome contains the major sigma factor (σ^{70}, RpoD) and as found in other sequenced streptococcal and *L. lactis* genomes, sigma B, a general stress transcription factor of *B. subtilis* and *S. aureus*, was not found. Another possible sigma factor is *comX*, a homologue of the alternate sigma factor found in *S. pneumoniae* and *S. pyogenes* (9, 53). However, unlike *S. pneumoniae* or *S. pyogenes*, only one copy of *com*X is present in the UA159 genome. Multiple sigma factors are characteristic of free-living bacteria and 18 were found in *B. subtilis* (54). However, bacterial pathogens that are highly specialized for a particular environmental niche typically encode 1–4 sigma factors (55), a number consistent with that found in *S. mutans*.

More than 100 transcriptional regulators, thirteen putative two-component regulatory systems (TCS) and one independent response regulator, five antiterminators, and homologs of the NusA and NusB proteins that are involved in termination in *B. subtilis* and *E. coli* (56) were identified. As with the other streptococci for which genome information is available, no homolog of the *rho* transcription terminator was found. Additionally, the genes encoding the global regulatory catabolite repressor protein CcpA (57) and a homologue of the stringent response protein RelA from *S. pyogenes* (58) were also found. Interestingly, a previously uncharacterized protein that is 54% similar to Rgg, a positive regulator of *gtfG* in *S. oralis* (59) and *S. gordonii* (60), was also found. Unlike these organisms, *S. mutans rgg* is not located upstream of the *gtf* gene. Whether this gene has a similar regulatory function in *S. mutans* remains to be determined.

Competence-mediated transformation is one of the mechanisms for horizontal gene transfer occurring between bacterial species. The UA159 strain is naturally competent and possesses an operon (*comCDE*) dedicated to quorum sensing and competence, composed of genes that encode a peptide pheromone precursor and an associated TCS responsible for regulation of pheromone production (61). Secretion and processing of the peptide pheromone is accomplished using the ComAB secretion apparatus (62). In addition to these two operons, the UA159 genome also contains the newly identified genes for DNA uptake (*comYA* and *comYB*) that are homologous to the same genes from *S. gordonii* (63). Two other pheromone-like peptides and three characteristic peptide ABC transporters that may play some role in cell signaling were also found (Fig. 2). The presence of a LuxS synthetase suggests that this strain is also able to synthesize autoinducer 2, a hormone-like molecule involved in cell–cell signaling and proposed to be responsible for cross-species communication (64). Therefore, the presence of this gene may be important for plaque formation.

Clinically isolated strains of *S. mutans* often produce bacteriocins that exhibit a wide spectrum of antibacterial activity against competing strains in the oral cavity. Although UA159 is bacteriocin nonproducing strain (65), five bacteriocin-associated genes, similar to the regulatory (*scnK* and *scnR*) and immunity genes (*scnG*, *scnE*, and *scnF*) of the *S. pyogenes* streptococcin A-FF22 gene cluster are present in UA159. However, no genes similar to *scnM* and *scnT*, responsible for maturation and transport, were found, implying either a different processing and transport mechanism for this bacteriocin or an inability to produce it. Interestingly, ORFs homologous to mutacin I and lacticin 481 immunity genes (66) were found. Although these proteins may be remnants of lost systems, they probably provide protection from bacteriocins synthesized by other species.

Mobile Genetic Elements. Mobile genetic elements are prominent features of virtually all bacterial genomes, and in some organisms the associated genes may influence metabolism or pathogenic potential of the organisms. Now that the genomes of several lactic acid bacteria have been determined, the distribution and types of mobile elements found illustrates several important distinctions within this group. In contrast to the closely related low %G+C Gram-positive organisms *S. pyogenes*, *L. monocytogenes*, and *L. lactis* (9, 13, 67), no temperate bacteriophage genomes are detectable in strain UA159. The genomes of *S. pneumoniae* strains TIGR4 and R6 also had no prophage genomes (10, 11), and interestingly, like *S. mutans*, the pneumococci are naturally competent for transformation. Naturally competent bacteria typically have mechanisms in place to prevent the uptake or incorporation of foreign DNA, and these systems may incidentally prevent bacteriophages from becoming resident in these genomes. However, the presence or absence of prophages in a bacterial genome is not invariably linked to competence: *B. subtilis* and *H. influenzae* are both naturally transformable and harbor prophages in their genomes, whereas *Neisseria gonorrhoeae* resembles *S. mutans* in being competent but lacking prophages (D. Dyer, personal communication).

The IS3 family of insertion elements is widely distributed in Eubacteria, being found in over 40 Gram-positive and negative species. Seven complete IS3 family elements, as well as 15 fragments, are found in UA159. Six of the IS3 family members are closely related to IS861 from *Streptococcus agalactiae* (68) and are almost identical at the nucleotide level, suggesting a clonal origin. The seventh family member is a variant of the *Streptococcus thermophilus* IS1193 element (88% identity; GenBank accession no. Y13713). Strain UA159 contains one probable conjugative transposon (TnSmu1) that is related to but distinct from the well known Tn916 from *Enterococcus faecalis*. Prominent in the genome is a large transposon-like region (TnSmu2) containing the genes similar to gramicidin and bacitracin synthetases; these genes include some of the largest ORFs found in the genome (one over 8 kb). This large region, over 40 kb, is flanked by remnants of transposases of the ISL3 family, arranged in inverted orientation relative to each other, whose reading frames are disrupted by frameshifts and contain several gene fragments from other IS element families. The nucleotide composition of this region markedly diverges from the genome average, having a %G+C of 28.9%. Such deviations may point to foreign origins of these elements. Although the transposons, IS elements, and fragments are distributed over the entire genome, there are several regions containing clusters of IS elements or remnants, suggesting that these regions may represent "hot spots" for integration.

Conclusion

The complete genome sequence of *S. mutans* enables a better understanding of the complexity and genetic specificity of this organism. This *Streptococcus*, unlike the related pathogens *S.*

pyogenes and *S. pneumoniae*, is part of the human oral flora, and only incidentally an oral pathogen. As such, it differs from the pathogenic streptococci in several aspects of its basic physiology and in its adaptations to maintain an ecological niche. As the genome analysis shows, *S. mutans* is able to metabolize a wider variety of carbohydrates than any other Gram-positive organism sequenced to date and can synthesize all of its required amino acids. To complement the number of carbohydrates it can use, *S. mutans* devotes a large portion of its coding potential (≈15%) to various transport mechanisms. The number of proteases, peptidases, and other exoenzymes produced by *S. mutans* clearly suggests that it derives resources from host tissues. The analysis of the genome sequence showed that around 16% of the predicted ORFs specified unique genes and revealed the presence of many previously uncharacterized genes predicted to be involved in virulence, transport, and gene regulation. These discoveries provide a basis for future drug development and new approaches in prevention and treatment of dental caries.

We thank Dr. R. R. B. Russell and Dr. D. Savić for helpful discussions; Rose Morales-Diaz, Mounir Elharam, Yonas Tesfai, Sara Downard, and Clayton Powell for their help in fluorescent DNA sequencer data collection; and Phoebe Loh, Sulan Qi, and Bart Ford for performing oligonucleotide synthesis. We are appreciative of the assistance given with some motif searches by Gerry Myers, Thomas Brettin, and Nina Thayer of the Los Alamos National Laboratories. This work was supported by Grant DE12489 from the National Institute of Dental and Craniofacial Research, National Institutes of Health.

1. World Health Organization (1994) WHO/28 (W.H.O., Geneva).
2. Loesche, W. J. (1986) *Microbiol. Rev.* **50,** 353–380.
3. Herzberg, M. C. (2000) in *Streptococcal Infections*, eds. Stevens, D. L. & Kaplan, E. L. (Oxford Univ. Press, New York), pp. 333–370.
4. Burne, R. A. (1998) *J. Dent. Res.* **77,** 443–452.
5. Colby, S. M. & Russell, R. R. B. (1997) *J. Appl. Microbiol. Symp. Suppl.* **83,** 80s–88s.
6. Kuramitsu, H. K. (2001) *Curr. Issues Mol. Biol.* **3,** 35–36.
7. Michalek, S. M., Katz, J. & Childers, N. K. (2001) *BioDrugs* **15,** 501–508.
8. Quivey, R. G., Kuhnert, W. L. & Hahn, K. (2001) *Crit. Rev. Oral Biol. Med.* **12,** 301–314.
9. Ferretti, J. J., McShan, W. M., Ajdić, D., Savic, D. J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A. N., McLaughlin, R., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4658–4663.
10. Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., *et al.* (2001) *Science* **293,** 498–506.
11. Hoskins, J., Alborn, W. E. J., Arnold, J., Blaszczak, L. C., Burgett, S., DeHoff, B. S., Estrem, S. T., Fritz, L., Fu, D. J., Fuller, W., Geringer, C., *et al.* (2001) *J. Bacteriol.* **183,** 5709–5717.
12. Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., *et al.* (2001) *Lancet* **357,** 1225–1240.
13. Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D. & Sorokin, A. (2001) *Genome Res.* **11,** 731–753.
14. Moriya, S. & Ogasawara, N. (1996) *Gene* **176,** 81–84.
15. Smith, T. J., Blackman, S. A. & Foster, S. J. (2000) *Microbiology* **146,** 249–262.
16. Clark, J. K. (1924) *Br. J. Exp. Pathol.* **5,** 141–147.
17. Poolman, B. (1993) *FEMS Microbiol. Rev.* **12,** 125–147.
18. Cvitkovitch, D. G., Gutierrez, J. A. & Bleiweis, A. S. (1997) *J. Bacteriol.* **179,** 650–655.
19. Terleckyj, B. & Shockman, G. D. (1975) *Infect. Immun.* **11,** 656–664.
20. Martin, E. J. S. & Wittenberger, C. L. (1980) *Infect. Immun.* **28,** 220–224.
21. Alifano, P., Fani, R., Lio, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M. S. & Bruni, C. B. (1996) *Microbiol. Rev.* **60,** 44–69.
22. Svensater, G., Sjogreen, B. & Hamilton, I. R. (2000) *Microbiology* **146,** 107–117.
23. Yamamoto, Y., Higuchi, M., Poole, L. B. & Kamio, Y. (2000) *J. Bacteriol.* **182,** 3740–3747.
24. Saier, M. H., Fagan, M. J., Hoischen, C. & Reizer, J. (1993) in *Bacillus subtilis*, eds. Sonenchein, A. L., Hoch, J. A. & Losick, R. (Am. Soc. Microbiol., Washington, DC), pp. 133–156.
25. Lin, E. C. C. (1976) *Annu. Rev. Microbiol.* **30,** 535–578.
26. Russell, R. R. B., Aduse-Opoku, J., Sutcliffe, I. C., Tao, L. & Ferretti, J. J. (1992) *J. Biol. Chem.* **267,** 4631–4637.
27. Jenkinson, H. F. & Demuth, D. R. (1997) *Mol. Microbiol.* **23,** 183–190.
28. Russell, M. W., Harrington, D. J. & Russell, R. R. B. (1995) *Infect. Immun.* **63,** 733–735.
29. Yamashita, Y., Shibata, Y., Nakano, Y., Tsuda, H., Kido, N., Ohta, M. & Koga, T. (1999) *J. Bacteriol.* **181,** 6556–6559.
30. Van Houte, J. (1980) in *Dental Plaque and Surface Interaction in the Oral Cavity*, ed. Leach, S. A. (Information Retrieval, London), pp. 69–100.
31. Kitten, T., Munro, C. L., Michalek, S. M. & Macrina, F. L. (2000) *Infect. Immun.* **68,** 4441–4451.
32. Jenkinson, H. F. (1994) *FEMS Microbiol. Lett.* **121,** 133–140.
33. Dintilhac, A. & Claverys, J. P. (1997) *Res. Microbiol.* **148,** 119–131.
34. Haas, W. & Banas, J. A. (2000) *J. Bacteriol.* **182,** 728–733.
35. Sato, Y. Y. & Kizaki, H. (1997) *Infect. Immun.* **65,** 668–675.
36. Holmes, A. R., McNab, R., Millsap, K. W., Rohde, M., Hammerschmidt, S., Mawdsley, J. L. & Jenkinson, H. F. (2001) *Mol. Microbiol.* **41,** 1395–1408.
37. Courtney, H. S., Li, Y., Dale, J. B. & Hasty, D. L. (1994) *Infect. Immun.* **62,** 3937–3946.
38. Chia, J. S., Yeh, C. Y. & Chen, J. Y. (2000) *Infect. Immun.* **68,** 1864–1870.
39. Lottenberg, R., Broder, C. C., Boyle, M. D., Kain, S. J., Schroeder, B. L. & Curtiss, R. (1992) *J. Bacteriol.* **174,** 5204–5210.
40. Mattos-Graner, R. O., Jin, S., King, W. F., Chen, T., Smith, D. J. & Duncan, M. J. (2001) *Infect. Immun.* **69,** 6931–6941.
41. Chia, J. S., Lee, Y. Y., Huang, P. T. & Chen, J. Y. (2001) *Infect. Immun.* **69,** 2493–2501.
42. Smith, D. J., Akita, H., King, W. F. & Taubman, M. A. (1994) *Infect. Immun.* **65,** 668–675.
43. Ferretti, J. J., Gilpin, M. L. & Russell, R. R. B. (1987) *J. Bacteriol.* **169,** 4271–4278.
44. Banas, J. A., Russell, R. R. B. & Ferretti, J. J. (1990) *Infect. Immun.* **58,** 667–673.
45. Janecek, S., Svensson, B. & Russell, R. R. B. (2000) *FEMS Microbiol. Lett.* **192,** 53–57.
46. Cowman, R. A., Schaefer, S. J., Fitzgerald, R. J., Rosner, D., Shklair, I. L. & Walter, R. G. (1979) *J. Dent. Res.* **58,** 2019–2027.
47. Rosengren, L. & Winblad, B. (1976) *Oral Surg.* **42,** 801–809.
48. Diaz-Torres, M. L. & Russell, R. R. B. (2001) *FEMS Microbiol. Lett.* **204,** 23–28.
49. Vickerman, M. M., Mather, N. M., Minick, P. E. & Edwards, C. A. (2002) *Oral Microbiol. Immunol.* **17,** 22–31.
50. Sutcliffe, I. C. & Harrington, D. J. (2002) *Microbiology* **148,** 2065–2077.
51. Osaki, M., Takamatsu, D., Shimoji, Y. & Sekizaki, T. (2002) *J. Bacteriol.* **184,** 971–982.
52. Bolken, T. C., Franke, C. A., Jones, K. F., Zeller, G. O., Jones, C. H., Dutton, E. K. & Hruby, D. E. (2001) *Infect. Immun.* **69,** 75–80.
53. Lee, M. S. & Morrison, D. A. (1999) *J. Bacteriol.* **181,** 5004–5016.
54. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A., Alloni, G., Azevedo, V., Bertero, M., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997) *Nature* **390,** 249–256.
55. Koonin, E. V., Aravind, L. & Galperin, M. Y. (2000) in *Bacterial Stress Responses*, eds. Storz, G. & Hengge-Aronis, R. (Am. Soc. Microbiol., Washington, DC), pp. 417–444.
56. Shazand, K., Tucker, J., Grunberg-Manago, M., Rabinowitz, J. C. & Leighton, T. (1993) *J. Bacteriol.* **175,** 2880–2887.
57. Simpson, C. L. & Russell, R. R. B. (1998) *Infect. Immun.* **66,** 2085–2092.
58. Mechold, U., Cashel, M., Steiner, K., Gentry, D. & Malke, H. (1996) *J. Bacteriol.* **178,** 1401–1411.
59. Fujiwara, T., Hoshino, T., Ooshima, T., Sobue, S. & Hamada, S. (2000) *Infect. Immun.* **68,** 2475–2483.
60. Sulavik, M. C., Tardif, G. & Clewell, D. B. (1992) *J. Bacteriol.* **174,** 3577–3586.
61. Li, Y.-H., Lau, P. C. Y., Lee, J. H., Ellen, R. P. & Cvitkovich, D. G. (2001) *J. Bacteriol.* **183,** 897–908.
62. Petersen, F. C. & Scheie, A. A. (2000) *Oral Microbiol. Immunol.* **15,** 329–334.
63. Lunsford, R. D. & Roble, A. G. (1997) *J. Bacteriol.* **179,** 3122–3126.
64. Winans, S. C. & Zhu, J. (2000) in *Bacterial Stress Responses*, eds. Storz, G. & Hengge-Aronis, R. (Am. Soc. Microbiol., Washington, DC), pp. 261–272.
65. Chen, P., Qi, F., Novak, J. & Caufield, P. W. (1999) *Appl. Environ. Microbiol.* **65,** 1356–1360.
66. Dufour, A., Rincé, A., Uguen, P. & Pennec, J.-P. L. (2000) *J. Bacteriol.* **182,** 5600–5605.
67. Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., *et al.* (2001) *Science* **294,** 849–852.
68. Rubens, C. E., Heggen, L. M. & Kuypers, J. M. (1989) *J. Bacteriol.* **171,** 5531–5355.