Identification of Unique Bacterial Gene Segments from *Streptococcus mutans* with Potential Relevance to Dental Caries by Subtraction DNA Hybridization

Deepak Saxena,[1]* Yihong Li,[2] and Page W. Caufield[1]

*Division of Diagnostics, Infectious Disease, and Health Promotion[1] and Department of Basic Science and Craniofacial Biology,[2]
New York University College of Dentistry, New York, New York 10010*

Using DNA subtractive hybridization, 49 unique gene segments were identified from a strain of *Streptococcus mutans* that was isolated from a patient with severe early childhood caries (S-ECC). Further hybridization with DNA from other *S. mutans* strains isolated from both caries-active and caries-free subjects yielded five unique sequences of DNA common to strains associated with S-ECC.

Dental caries is one of the most common chronic childhood diseases (19). The prevalence of early childhood caries ranges from 0.8% to 72% (11). The major bacterial component associated with dental caries has been attributed to mutans streptococci, notably, *Streptococcus mutans* and *S. sobrinus* (4, 7, 13). Not all individuals colonized by mutans streptococci, however, manifest the disease (5). *S. mutans* strains may differ in the propensity to cause disease (12) or virulence as a function of the acquisition and expression of pathogenicity islands (PAIs) or other genetic elements (6, 9). For example, some strains of *S. mutans* express altered metabolic properties via the presence of an insertion sequence upstream of the sucrose hydrolase operon (14). The investigation of DNA loci associated with strains identified as virulent may enhance our understanding of why some children have the disease and others do not (3, 4, 10, 14, 15).

Our approach to identify unique segments involves PCR-based DNA suppressive subtractive hybridization (SSH). This technique provides preferential enrichment of sequences that differ between two or more strains of bacteria isolated from patients with and without a particular disease (3, 18, 20). The application of PCR-based SSH as an approach in identifying genetic elements unique to *S. mutans* strains associated with severe early childhood caries (S-ECC) is described in this study.

The study included five *S. mutans* strains isolated from children with or without S-ECC having a single genotype of *S. mutans*. A putatively virulent strain of *S. mutans*, AF199 (a tester strain), was isolated from the dental plaque of a 7-year-old child from Central Africa with S-ECC (decayed, missing, or filled teeth [dmft] = >14). Strain AF194 (a driver strain) was isolated from a caries-free child who lived in the same village. Two strains, BZ12 and CH19, were isolated from a 2.5-year-old Brazilian boy (dmft = 20) and a 5.5-year-old Chinese boy (dmft = 15), respectively. BZ18 was isolated from a 3-year-old caries-free Brazilian boy. All *S. mutans* strains were isolated

using mitis salivarius-bacitracin selective medium (8). Genomic DNA was extracted using a genomic DNA extraction kit (QIAGEN, Valencia, CA). The final concentration of DNA was adjusted spectrophotometrically to 1 μg/μl.

DNA fragments present in AF199 (the target DNA of the tester strain) but not in AF194 (the nontarget DNA of the driver strain) were isolated using the PCR-Select bacterial genome subtraction kit (Clontech Laboratories, Inc., Palo Alto, CA). An overview of the study design is shown in Fig. 1. Briefly, 2 μg of genomic DNA from each strain was digested with RsaI. Tester DNA was subdivided into two portions and ligated separately with adaptor I or II (provided by the manufacturer). Two hybridizations were performed. Initially, an excess of driver DNA was added to each adaptor-ligated tester sample, heat denatured, and allowed to anneal. In a second hybridization, the two primary hybridization samples were mixed together without denaturing with an excess of freshly denatured driver DNA. After the second hybridization, the entire sample was subjected to a primary PCR (primer 1, 5′-CTAATACGACTCACTATAGGGC-3′) and a secondary PCR (nested primer 1, 5′-TCGAGCGGCCGCCCGGGCAG GT-3′; nested primer 2R, 5′-AGCGTGGTCGCGGCCGAG GT-3′) to amplify the tester-specific sequences using the Advantage 2 polymerase mix (Clontech) and these PCR-amplified fragments were cloned using a TOPO TA cloning kit (Invitrogen, San Diego, CA).

Purified plasmid DNA containing subtracted fragments were sequenced by using the M13 universal primer. Sequencing reactions were conducted on an ABI model 377 DNA sequencer. Sequencing was preformed in both directions, and the resultant sequences were aligned. Both nucleic acid and predicted protein compositions were compared to those archived in GenBank using BLAST (National Center for Biotechnology Information [NCBI]). Sequences were also analyzed for protein coding regions via ORF finder (NCBI) (16).

Confirmatory hybridizations within and between strains were performed using dot blotting to identify clones that contained unique DNA. Selected clones containing subtracted fragments (approximately 300 to 1,500 bp) were amplified by PCR with nested primers 1 and 2R. The PCR products were purified, arrayed (100 ng per spot) on multiple BrightStar-Plus

* Corresponding author. Mailing address: Division of Diagnostics, Infectious Disease, and Heath Promotion, New York University College of Dentistry, 345 E. 24th Street, New York, NY 10010. Phone: (212) 998-9203. Fax: (212) 995-3994. E-mail: ds100@nyu.edu.
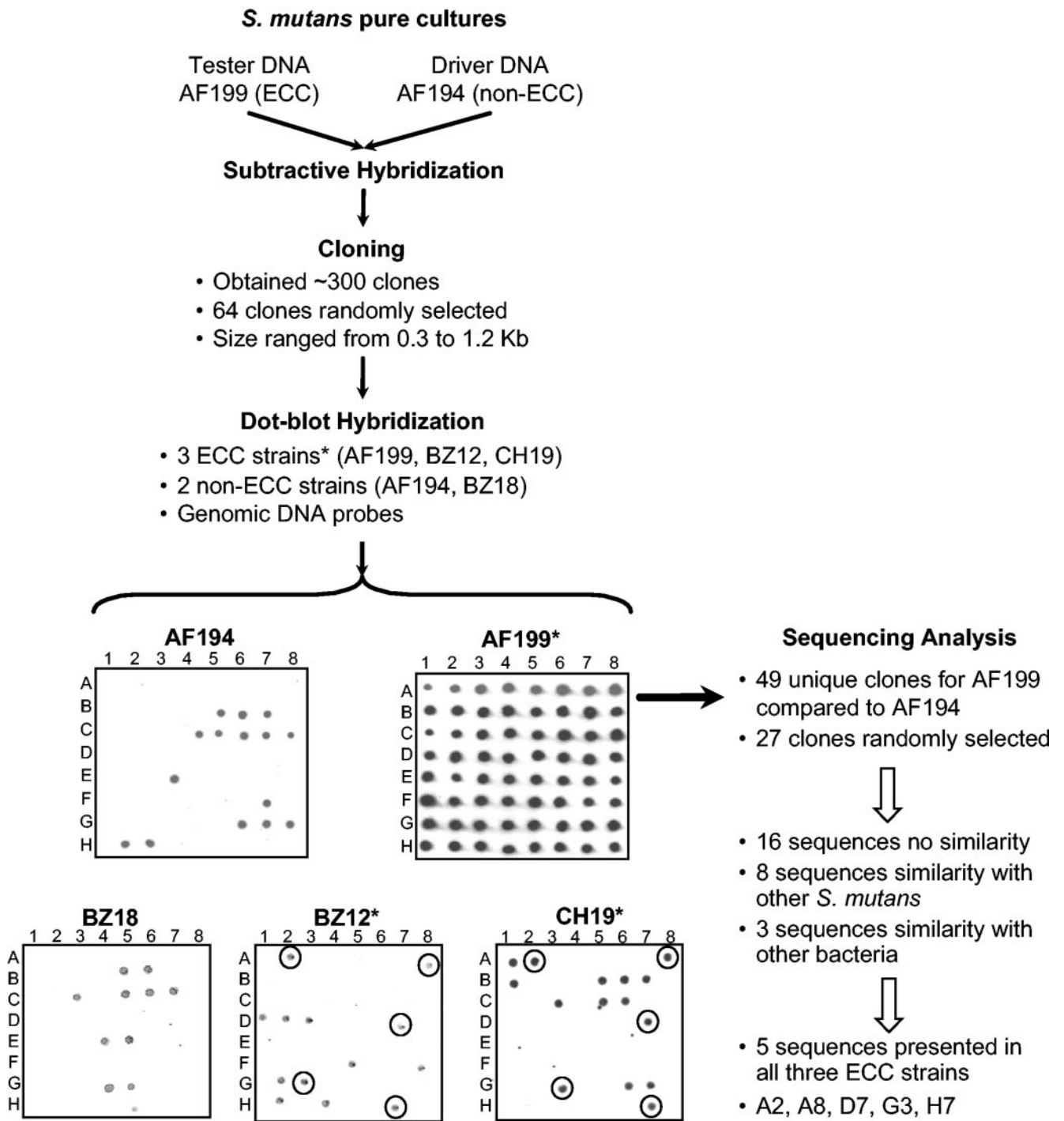
FIG. 1. Overall study design for subtractive DNA hybridization. After the initial subtraction, 64 colonies were randomly selected for dot blot hybridization against genomic DNA samples of S-ECC strains AF199, BZ12, and CH19, as well as caries-free strains AF194 and BZ15. Five individual blots show different hybridization results. The five unique clones (A2, A8, D7, G3, and H7) were present in all S-ECC strains but absent in caries-free strains. The GC content differed significantly from the species designated GC content of 36%.

nylon membranes (Ambion, Inc., Austin, TX), denatured by adding NaOH to a final concentration of 100 mM, and then neutralized with equal amount of 20× SSC buffer (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). The samples were hybridized with RsaI-digested, biotin-labeled (Ambion, Inc.)

genomic DNA using the AF199, AF194, BZ12, BZ18, or CH19 strain. The hybridization was performed in ULTRAhyb hybridization buffer (Ambion, Inc.) at 42°C. Hybridized double-stranded DNA was detected using a streptavidin-alkaline phosphatase conjugate (BrightStar detection kit; Ambion).

TABLE 1. Sequence analysis of gene fragments detected in *S. mutans* AF199

Subtracted clone	Size (bp)	BLASTX hit & accession no.	Organism	E value[a]	Predicted function or property
A2	493	Not detected			
A3	506	AF207710	*S. mutans* CH43	−48	Structural gene for mutacin I
A5	350	None detected			
A8	473	None detected			
B3	509	None detected			
C1	480	None detected			
C5	380	NC_004350	*S. mutans* UA159	0	Putative transposase fragment
C6	1,486	AE015011	*S. mutans* UA159	0	Putative translation elongation factor P
C7	467	NC_004350	*S. mutans* UA159	0	Malolactic enzymes
C8	650	AE014944	*S. mutans* UA159	0	Putative transcriptional regulator
D2	706	None detected			
D3	504	AF077024	*S. mutans* AF199	−145	Hypervariable region
D6	1,165	AE008400	*S. pneumoniae* R6	−44	Excinuclease ABC subunit A
D7	603	None detected			
D8	525	None detected			
E8	354	AJ550464	*Bacillus silvestris*	−166	Partial 16S rRNA gene
F2	417	None detected			
F3	570	AE014878	*S. mutans* UA159	0	Conserved hypothetical protein
F5	710	None detected			
F8	450	None detected			
G1	391	None detected			
G3	610	None detected			
G5	575	None detected			
G7	475	NC_004350	*S. mutans* UA159	−166	Putative hemolysin
H5	590	None detected			
H7	690	None detected			
H8	523	NC_002945	*Mycobacterium bovis* AF2122/97	0	dTDP-rhamnose modification protein

[a] E value represents the number of times this match or a better one would be expected to occur purely by chance in a search of the entire database.

The suppressive subtractive hybridization results are shown in Fig. 1. More than 300 clones were obtained from the initial subtractive hybridization of AF199. The fragment sizes ranged from 0.3 kb to 1.2 kb. Sixty-four of the clones were randomly selected for further analysis. Of these 64 clones, 49 were unique for AF199 and 27 were sequenced. Sixteen clones had no significant matches with known sequences in GenBank (Table 1). Six clones resembled genes found in other strains of *S. mutans*, similar to the following functional genes: a putative transposase fragment (C5), a putative translation elongation factor (C6), a malolactic enzyme (C7), a putative transcriptional regulator (C8), a hypothetical protein (F3), and a putative hemolysin (G7). The A3 clone contained a portion of the structural gene for mutacin I (15). Another clone (D3) matched the hypervariable region found in plasmid-containing strains of *S. mutans*, including AF199 (21). These two loci (plasmid and mutacin I) served as controls because AF199

contains both loci while the driver strain does not. Clones D6, E8, and H8 showed significant sequence similarity to excinuclease ABC subunit A of *S. pneumoniae*, the partial 16S rRNA gene of the soil bacterium *Bacillus silvestris*, and the *rmlD* gene from *Mycobacterium bovis*, respectively.

Next, the 64 tester-specific DNA clones were hybridized against the three *S. mutans* strains, two from S-ECC children (BZ12 and CH19) and one from a caries-free child (BZ18). The percentage of commonality of fragments with AF199 was 23% for CH19, 20% for BZ12, and 16% for BZ18. Five of the 64 clones were present in all three caries-active *S. mutans* strains, whereas they were absent in caries-free strains (Table 2). The GC contents of individual fragments ranged from 38% to 56% and had an average difference of 8.98% from the GC content of UA159 sequences, which was 36.8% (Table 2). Four fragments had GC contents of greater than 45%. Of these, two had no significant hits to any GenBank submission based on

TABLE 2. Subtracted sequences present in AF199, BZ12, and CH19

Subtracted clone	Size (bp)	%GC	Difference (%GC)[a]	Predicted no. of ORFs[b]	Similar protein or conserved domain
A2	493	38	1.2	3	*Pseudomonas* avirulence D protein
A8	423	42.3	5.5	4	Polyketide synthase COG3321
D7	493	43.6	6.8	4	None detected[c]
G3	460	56	19.2	3	TraQ protein (transposon, *Bacteroides fragilis* YCH46; *P. gingivalis* W83)
H7	653	49.1	12.3	5	None detected[c]

[a] Differences were based on the GC content of UA159, which was 36.8% (2).
[b] NCBI ORF finder (16).
[c] No significant similarities to archived GenBank sequences.

BLAST searches. The remaining three fragments had homologies to the protein sequences of organisms other than *S. mutans*.

Two putative conserved domains of 107 amino acids for *Pseudomonas* avirulence D protein (AvrD) and conserved domains of 118 amino acids for polyketide synthase were detected in clones A2 and A8, respectively (Table 2). Clone G3 exhibited similarity to the TraQ protein (transposon, *Bacteroides fragilis* YCH46; *Porphyromonas gingivalis* W83). Clones D7 and H7 exhibited no significant similarities to other GenBank database entries. Sequence analysis of clones D7 and H7 indicated that there are four or five possible open reading frames, respectively, but, no protein matches were found in the existing database. It is possible that these unique gene fragments may constitute a new class of proteins (10); however, their direct role in caries processes is unknown.

The GC composition of fragments identified from AF199 indicates that these fragments may be part of PAIs or horizontally transferred genes, as observed with known PAIs from other organisms, which differ from the core genome in GC composition, on average, by 8.7% (17).

Agron et al. (1) showed a 96% recovery rate of coding genes using SSH and different restriction enzymes in two *Helicobacter pylori* strains whose complete genomic sequences are known. SSH examination of other putative pathogenic strains should yield additional genetic loci associated with caries if, indeed, strains differ in virulence by virtue of specifically acquired genetic loci. Extension of the SSH approach should also be able to identify the unique genetic loci within a total plaque biofilm associated with caries, regardless of which organisms harbor the segment. These genetic elements could be applied in a microarray format, allowing a patient's plaque sample to be analyzed for the presence or absence of these indicators. This, in turn, may provide valuable diagnostic tools in risk assessment for this severe form of dental caries.

Nucleotide sequence accession numbers. The nucleotide sequences unique to AF199 and other cariogenic strains of *S. mutans* were deposited in GenBank under accession numbers DQ072007 and DQ072008.

REFERENCES

1. **Agron, P. G., M. Macht, L. Radnedge, E. W. Skowronski, W. Miller, and G. L. Andersen.** 2002. Use of subtractive hybridization for comprehensive surveys of prokaryotic genome differences. FEMS Microbiol. Lett. **211:**175–182.
2. **Ajdic, D., W. M. McShan, R. E. McLaughlin, G. Savic, J. Chang, M. B. Carson, C. Primeaux, R. Tian, S. Kenton, H. Jia, S. Lin, Y. Qian, S. Li, H. Zhu, F. Najar, H. Lai, J. White, B. A. Roe, and J. J. Ferretti.** 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc. Natl. Acad. Sci. USA **99:**14434–14439.
3. **Akopyants, N. S., A. Fradkov, L. Diatchenko, J. E. Hill, P. D. Siebert, S. A. Lukyanov, E. D. Sverdlov, and D. E. Berg.** 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95:**13108–13113.
4. **Becker, M. R., B. J. Paster, E. J. Leys, M. L. Moeschberger, S. G. Kenyon, J. L. Galvin, S. K. Boches, F. E. Dewhirst, and A. L. Griffen.** 2002. Molecular analysis of bacterial species associated with childhood caries. J. Clin. Microbiol. **40:**1001–1009.
5. **Burt, B. A., W. J. Loesche, and S. A. Eklund.** 1985. Stability of selected plaque species and their relationship to caries in a child population over 2 years. Caries Res. **19:**193–200.
6. **Finlay, B. B., and S. Falkow.** 1997. Common themes in microbial pathogenicity revisited. Microbiol. Mol. Biol. Rev. **61:**136–169.
7. **Gibbons, R. J., and J. V. Houte.** 1975. Bacterial adherence in oral microbial ecology. Annu. Rev. Microbiol. **29:**19–44.
8. **Gold, O. G., H. V. Jordan, and J. V. Houte.** 1973. A selective medium for *Streptococcus mutans*. Arch. Oral Biol. **18:**1357–1364.
9. **Groisman, E. A., and C. Mouslim.** 2000. Molecular mechanisms of *Salmonella* pathogenesis. Curr. Opin. Infect. Dis. **13:**519–522.
10. **Hinton, J. C.** 1997. The *Escherichia coli* genome sequence: the end of an era or the start of the FUN? Mol. Microbiol. **26:**417–422.
11. **Ismail, A. I., and W. Sohn.** 1999. A systematic review of clinical diagnostic criteria of early childhood caries. J. Public Health Dent. **59:**171–191.
12. **Kohler, B., and B. Krasse.** 1990. Human strains of mutans streptococci show different cariogenic potential in the hamster model. Oral Microbiol. Immunol. **5:**177–180.
13. **Loesche, W. J., J. Rowan, L. H. Straffon, and P. J. Loos.** 1975. Association of *Streptococcus mutants* with human dental decay. Infect. Immun. **11:**1252–1260.
14. **Macrina, F. L., K. R. Jones, C. A. Alpert, B. M. Chassy, and S. M. Michalek.** 1991. Repeated DNA sequence involved in mutations affecting transport of sucrose into *Streptococcus mutans* V403 via the phosphoenolpyruvate phosphotransferase system. Infect. Immun. **59:**1535–1543.
15. **Qi, F., P. Chen, and P. W. Caufield.** 2000. Purification and biochemical characterization of mutacin I from the group I strain of *Streptococcus mutans*, CH43, and genetic analysis of mutacin I biosynthesis genes. Appl. Environ. Microbiol. **66:**3221–3229.
16. **Rombel, I. T., K. F. Sykes, S. Rayner, and S. A. Johnston.** 2002. ORF-FINDER: a vector for high-throughput gene identification. Gene **282:**33–41.
17. **Schmidt, H., and M. Hensel.** 2004. Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. **17:**14–56.
18. **Tinsley, C. R., and X. Nassif.** 1996. Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. Proc. Natl. Acad. Sci. USA **93:**11109–11114.
19. **U.S. Department of Health and Human Services.** 1991. Healthy people 2000; national health promotion and disease prevention objective. Full report, with commentary. DHHS publication 91-50212. U.S. Department of Health and Human Services, Washington, D.C.
20. **Winstanley, C.** 2002. Spot the difference: applications of subtractive hybridisation to the study of bacterial pathogens. J. Med. Microbiol. **51:**459–467.
21. **Zhou, X., P. W. Caufield, Y. Li, and F. Qi.** 2001. Complete nucleotide sequence and characterization of pUA140, a cryptic plasmid from *Streptococcus mutans*. Plasmid **46:**77–85.