

# Dispensable genes and foreign DNA in *Streptococcus mutans*

Janet C. Waterhouse and Roy R. B. Russell

Correspondence  
Roy R. B. Russell  
r.r.russell@ncl.ac.uk

Oral Biology, School of Dental Sciences, University of Newcastle, Newcastle upon Tyne  
NE2 4BW, UK

Received 1 November 2005  
Revised 20 January 2006  
Accepted 8 February 2006

A range of properties, including the ability to utilize various sugars, bind macromolecules and produce mutacins, are known to vary in their occurrence in different strains of *Streptococcus mutans*. In addition, insertion-sequence elements show a limited distribution and sequencing of the genome of *S. mutans* UA159 has revealed the presence of putative genomic islands of atypical base composition indicative of foreign DNA. PCR primers flanking regions suspected of having inserted DNA were designed on the basis of the genome sequence of *S. mutans* UA159 and used to explore variation in a collection of 39 strains isolated in various parts of the world over the last 40 years. Extensive differences between strains were detected, and similar insertion/deletion events appear to be present in the genomes of strains with very different origins. In two instances, insertion of foreign DNA appears to have displaced original *S. mutans* genes. Together with previous results on the occurrence of deletions in genes associated with sugar metabolism, the results indicate that *S. mutans* has a core genome and a dispensable genome, and that dispensable genes have become widely distributed through horizontal transfer.

## INTRODUCTION

*Streptococcus mutans* is regarded as the principal microbial aetiological agent of dental caries so has been proposed as a useful predictor of caries risk and also as the target of novel preventive strategies (Russell, 2003). It has long been recognized that considerable phenotypic variation exists within the species *S. mutans*. The variation includes differences in antigenic polysaccharide to give serotypes c, e, f and k, and variation in patterns of sugar fermentation, bacteriocin type, capacity to produce acid, interaction with salivary components, sucrose-dependent adherence and release of surface proteins. As a consequence of this variety, strains of *S. mutans* differ in their ability to cause caries (de Soet *et al.*, 2000; Kohler & Krasse, 1990; Kohler *et al.*, 1995). Genetic fingerprinting techniques have also demonstrated extensive variation within *S. mutans*. RFLP of *S. mutans* was introduced by Kulkarni *et al.* (1989), who demonstrated extensive variety in RFLP patterns amongst different isolates and were the first to show that an individual could carry multiple genotypes. They also found evidence of intra-family transmission, and this was extended by others using RFLP to follow transmission from mothers

to their infants and to monitor the stability of *S. mutans* in individual mouths (Li & Caufield, 1995). Other techniques based on the variation in DNA sequence, including ribotyping (Alaluusua *et al.*, 1994) and arbitrarily primed PCR (Gronroos & Alaluusua, 2000), have been used to monitor the populations of *S. mutans*. While all these methods indicate the existence of extensive sequence variation, they paradoxically imply a stability of genotype within an individual host. However, fingerprinting techniques give no information on the genetic basis for generation of variety or differences in phenotypic properties (Napimoga *et al.*, 2004). Studies of individual genes have also demonstrated variation within *S. mutans*. This is seen in the sequence variation of individual genes amongst strains (Brady *et al.*, 1991; Ferretti *et al.*, 1989; Fujiwara *et al.*, 1998) and may result in altered function of the encoded proteins. For example, point mutations have been detected in the *pac*, *gfpC* and *srtA* genes that affect surface properties (Lee & McGavin, 2004; Sato *et al.*, 2002). Recombination between the tandem genes *gtfB* and *gtfC* resulting in a partial deletion has been reported, and this can result in altered adherence (Hazlett *et al.*, 1998; Mattos-Graner *et al.*, 2004; Ueda & Kuramitsu, 1988).

The completion of the *S. mutans* UA159 genome sequence (Ajdic *et al.*, 2002) allows its use as a benchmark for studies of other strains of *S. mutans*, which will give new insights into the extent of the diversity within the species. Recently we have shown that the inability of some *S. mutans* isolates to ferment melibiose and raffinose is associated with loss of

Abbreviation: IS, insertion sequence.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are DQ368681 and DQ368682.

Figures showing sequence alignments and a table showing the spectrum of bacitracin inhibition are available as supplementary material with the online version of this paper.

the *msm* and *gal* operons present in strain UA159, and their replacement by *ISSmu3*, a defective insertion element (Robinson *et al.*, 2003). The inability of certain strains to utilize  $\beta$ -glucosides is also associated with deletions in the *bgl* or *cel* genes (Old *et al.*, 2006). Other researchers have reported that genes required for formation of serotype-specific polysaccharide are absent from untypable isolates (Fujiwara *et al.*, 2001). There is also evidence for the existence of genes in some strains of *S. mutans* that are not found in UA159. For example, some strains possess the gene for a collagen-binding adhesin not present in UA159 (Sato *et al.*, 2004a, b). By subtractive hybridization, Saxena *et al.* (2005) identified five genes that were found in one strain of *S. mutans* but not in others, including UA159. A further source of genetic variation is the presence of foreign DNA acquired by horizontal transfer. In contrast to other streptococci (Banks *et al.*, 2004; Ferretti *et al.*, 2004; Herbert *et al.*, 2005), in *S. mutans* there is scant evidence for bacteriophage and plasmids are rare (Caufield *et al.*, 1988). However, a large number of intact or incomplete insertion sequence elements (IS) and transposases were detected in the UA159 genome (Ajdic *et al.*, 2002), which potentially could result in altered RFLP patterns and phenotypic changes due to insertion into genes, effects on gene expression or by mediating deletions and rearrangements. Foreign DNA can often be recognized by its variation from the majority of the genome in base composition or codon preference, and a number of such potential 'genomic islands' have been identified in the UA159 genome sequence (Ajdic *et al.*, 2002) and a genome annotation available at <http://www.stdgen.lanl.gov/oragen/>.

Taken together, these results show that genetic variation within *S. mutans* occurs by a variety of point mutations and chromosomal insertion/deletion events. This paper examines the distribution of a number of insertions and deletions in a collection of strains of *S. mutans*.

## METHODS

**Bacterial strains and growth conditions.** The bacterial strains used in this study are listed in Table 1. *S. mutans* UA159 is the strain for which the genome sequencing has been completed (Ajdic *et al.*, 2002), and is considered a wild-type strain. Other wild-type strains include the common laboratory strains NCTC 10449, GS-5, Ingbritt, 3209, NG8 and V403. Strains At10, B2, MT4653, MT4863, 13.1, 12, 29, AS1, L18, C4 and G8 are independent melibiose-negative isolates used in previous studies (Old *et al.*, 2006; Robinson *et al.*, 2003; Ushiro *et al.*, 1991). Strains V1995 and V1996 were kindly provided by F. Macrina, Virginia Commonwealth University, Richmond, Virginia. Isolates in our local collection were from the saliva of undergraduates or obtained from cases of endocarditis at the Freeman Hospital, Newcastle upon Tyne. *S. mutans* was cultured in Todd-Hewitt broth (Oxoid) supplemented with 0.5% (w/v) yeast extract (THYE) and *Escherichia coli* was cultured in Luria-Bertani (LB) broth, solidified with agar when required.

**Bioinformatic analysis.** The complete 2 032 327 bp *S. mutans* genome sequence (GenBank accession no. AE014133) and its annotation (NC004350) was visualized with the Artemis program

(Rutherford *et al.*, 2000) to locate ORFs. Individual ORFs were analysed with the BLAST suite of programs (Altschul *et al.*, 1990) at <http://www.ncbi.nlm.nih.gov/blast/> to identify known *S. mutans* genes or homologues of other genes in the database, and to allow provisional identification of novel genes. Alignment of nucleotide sequences and amino acid sequences was performed with the CLUSTAL\_X (Thompson *et al.*, 1997) and CLUSTAL W (Thompson *et al.*, 1994) programs, respectively. Complete and incomplete microbial genomes were accessed through GenBank, TIGR (<http://www.tigr.org/tdb/mdb/mdbinprogress.html>) and Virginia Commonwealth University (<http://www.sanguinis.mic.vcu.edu/>). The database of IS elements was accessed at <http://www-is.biotoul.fr/is.html>.

**DNA extraction.** Chromosomal DNA was prepared from cultures of *S. mutans* as described elsewhere (Old *et al.*, 2006). Plasmid DNA was extracted from *Escherichia coli* cultured with shaking at 37 °C for approximately 16 h using the QIAprep spin miniprep kit (Qiagen), according to the manufacturer's instructions.

**PCR.** Primers for PCR were designed using the Primer3 program (Rozen & Skaletsky, 2000), available at [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), using published sequences or data from the genome of *S. mutans* UA159 (GenBank accession no. AE014133). The primers used in this study are listed in Table 2. Amplification was carried out in a 10  $\mu$ l reaction volume comprising 1  $\times$  Extensor Hi-Fidelity PCR Master Mix (ABgene), 1  $\mu$ M forward primer, 1  $\mu$ M reverse primer and 50 ng DNA template using a GeneAmp9700 thermal cycler. Cycling parameters comprised an initial denaturation at 95 °C for 3 min, 30 cycles of denaturation at 95 °C for 1 min, annealing at  $x$  °C for 1 min and extension at 72 °C for  $y$  min (where annealing temperature  $x$  and extension time  $y$  depend on the  $T_m$  of the primers used and the expected amplicon size, respectively), and a final extension of 72 °C for 10 min. PCR products to be used as templates for sequencing reactions were gel-purified using the QIAEX II gel extraction kit (Qiagen).

**Cloning and sequencing.** Sequencing of the TnSmu2 region was performed using a PCR product as the template, while the *gfpA* region from V1996 was cloned first. The *gfpA* region was amplified using primers flanking this gene (Fig. 1b), and the amplicon was gel-purified using the QIAEX II gel extraction kit (Qiagen). The amplicon was treated with T4 DNA polymerase to create blunt ends, and digested with *DraI* or *HaeIII*, then dephosphorylated using Antarctic phosphatase (New England Biolabs). The reaction products were purified using the QIAquick PCR purification kit (Qiagen) and the pooled digestion product cloned into PCR Blunt II TOPO (Invitrogen), using an (undigested) insert:vector molar ratio of approximately 10:1. The ligated plasmids were transformed into chemically competent Top10 cells (Invitrogen), spread onto LB agar containing 50  $\mu$ g kanamycin ml<sup>-1</sup> and incubated overnight at 37 °C. Colonies were cultured in LB broth containing 50  $\mu$ g kanamycin ml<sup>-1</sup>, and the plasmid DNA extracted and digested with *EcoRI* to release the insert. Plasmids containing inserts of the different sizes recovered were used as templates in sequencing reactions with plasmid-based primers. DNA sequencing was carried out by MWG Biotech using dye terminator chemistry. The sequences of the TnSmu2 region from strain LML7 and the *gfpA* region from V1996 have been deposited in GenBank with the accession numbers DQ368681 and DQ368682, respectively. For each sequence, 20 nt at each end identical to the sequence of UA159 have been included to facilitate location of the novel sequences on the reference UA159 genome.

**Detection of glucan-binding proteins.** Glucan-binding proteins in concentrated culture supernatants were detected after SDS-PAGE with specific rabbit antiserum against GbpA or GbpD and biotin-label dextran as described previously (Shah & Russell, 2004).

**Table 1.** Strains of *S. mutans*

UA159 was the strain selected for genome sequencing, other strains are arranged according to the year in which they were isolated or first reported in the literature. The stock of GS-5 was obtained from Howard Kuramitsu, Department of Oral Biology, State University of New York, Buffalo, USA. All were isolated from the oral cavity apart from those marked with an asterisk, which were from cases of endocarditis. + indicates that the strain gave the same result as UA159; deletions are indicated by  $\Delta$  and insertions by  $\Omega$ . Values in columns TnSmu2 and *gtfBC* indicate the size of the bands detected by PCR, in kb. 1367 and 1367H represent the different alleles of SMU.1367 present in UA159 and LML7, respectively. ND, no PCR product detected.

Strain	Isolation	Year	Serotype	<i>msm</i>	<i>bgl</i>	<i>cel</i>	TnSmu2	1367	1367H	<i>gbpA</i>	<i>gtfBC</i>	<i>cnm</i>	<i>smb</i>	<i>mutI</i>	<i>mutII</i>	IS199	IS3
UA159	Birmingham, USA	1982	c	+	+	+	+	+	-	+	10+5	-	-	-	-	-	+
Ingbritt	Malmö, Sweden	1966	c	+	+	+	5.5	-	+	+	5	-	-	-	-	-	-
GS-5 (HK)	Boston, USA	1966	c	+	+	+	+	+	-	+	10+5	-	$\Omega$	-	-	-	+
AT10	Malmö, Sweden	1968	c	$\Delta$	$\Delta$	$\Delta$	5.5	-	+	+	5	-	-	-	-	$\Omega$	-
B2	Malmö, Sweden	1968	c	$\Delta$	$\Delta$	$\Delta$	5.5	-	+	+	5	-	-	-	-	-	-
NG8	New Guinea	1970	c	+	+	+	ND	+	-	+	5	-	-	-	-	-	-
NCTC 10449	London, UK	1972	c	+	+	+	+	+	-	+	5	-	-	-	-	-	-
3209	Birmingham, UK	1979	c	+	+	+	+	+	-	+	10+5	-	-	-	-	-	-
MT4653	Osaka, Japan	1980	e	$\Delta$	$\Delta$	$\Delta$	5.5	-	+	+	5	-	-	-	-	$\Omega$	-
MT4863	Osaka, Japan	1980	c	$\Delta$	+	+	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
V403	Richmond, USA	1986	c	+	+	+	ND	+	-	$\Omega$	10+5	$\Omega$	-	-	-	$\Omega$	-
V1996	Richmond, USA	1991	c	+	+	+	ND	+	-	$\Omega$	10	-	-	-	-	-	-
C4	London, UK	1991	c	$\Delta$	+	$\Delta$	ND	-	+	+	ND	-	-	-	-	-	-
12	London, UK	1991	c	$\Delta$	$\Delta$	$\Delta$	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
29	London, UK	1991	c	$\Delta$	$\Delta$	$\Delta$	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	+
19	London, UK	1991	f	$\Delta$	+	+	ND	ND	ND	+	ND	-	-	-	-	-	-
21	London, UK	1991	k	$\Delta$	+	+	ND	ND	ND	+	ND	-	-	-	-	$\Omega$	-
G8	London, UK	1991	c	$\Delta$	+	$\Delta$	5.5	-	+	+	5	-	-	-	-	$\Omega$	-
LT11	Farmington, USA	1993	c	+	+	+	4.5	+	-	+	10+5	-	-	-	-	-	-
Fsm1	Newcastle, UK	1994	c	+	+	+	ND	+	-	+	5	-	-	-	-	-	-
Fsm2	Newcastle, UK*	1994	c	+	+	+	ND	+	-	+	10+5	-	-	-	-	-	+
Fsm5	Newcastle, UK*	1994	c	+	+	+	ND	+	-	+	10+5	-	$\Omega$	-	-	-	+
AS1	Newcastle, UK	1995	c	$\Delta$	$\Delta$	$\Delta$	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
13.1	Newcastle, UK	1995	e	$\Delta$	$\Delta$	$\Delta$	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	+
28	Newcastle, UK*	1997	c	+	+	+	+	+	-	+	10+5	-	$\Omega$	$\Omega$	-	-	-
34	Newcastle, UK*	1997	c	+	$\Delta$	+	5.5	-	+	$\Omega$	10+5	-	-	-	-	$\Omega$	-
35	Newcastle, UK*	1997	c	+	+	+	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
60	Newcastle, UK*	1997	c	+	+	+	ND	+	-	+	10+5	-	-	-	-	-	+
66	Newcastle, UK*	1997	c	+	+	+	ND	+	-	+	10+5	-	-	-	-	-	+
L13	Helsinki, Finland	2000	e	$\Delta$	+	$\Delta$	5.5	-	+	+	10+5	$\Omega$	-	$\Omega$	-	-	-
L18	Helsinki, Finland	2000	c	$\Delta$	+	+	4.5	-	+	+	10+5	-	-	-	-	-	-
LML1	Newcastle, UK	2001	e	$\Delta$	+	+	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
LML2	Newcastle, UK	2001	c	+	+	+	5.5	-	+	+	ND	-	-	-	-	-	+
LML3	Newcastle, UK	2001	c	$\Delta$	+	+	+	+	-	+	10+5	-	-	-	-	-	+
LML4	Newcastle, UK	2001	e	+	+	+	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
LML5	Newcastle, UK	2001	c	+	$\Delta$	+	5.5	-	+	$\Omega$	10+5	-	-	-	$\Omega$	$\Omega$	-
LML6	Newcastle, UK	2001	c	+	+	+	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
LML7	Newcastle, UK	2001	c	+	+	+	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	-
LML8	Newcastle, UK	2001	c	+	+	+	5.5	ND	ND	+	10+5	$\Omega$	-	-	-	-	-
LML9	Newcastle, UK	2001	c	+	+	+	5.5	-	+	+	10+5	-	-	-	-	$\Omega$	+

Sucrose-dependent agglutination was measured by growth in tubes containing THYE broth with 5% sucrose (Banas & Gilmore, 1991; Shah & Russell, 2004).

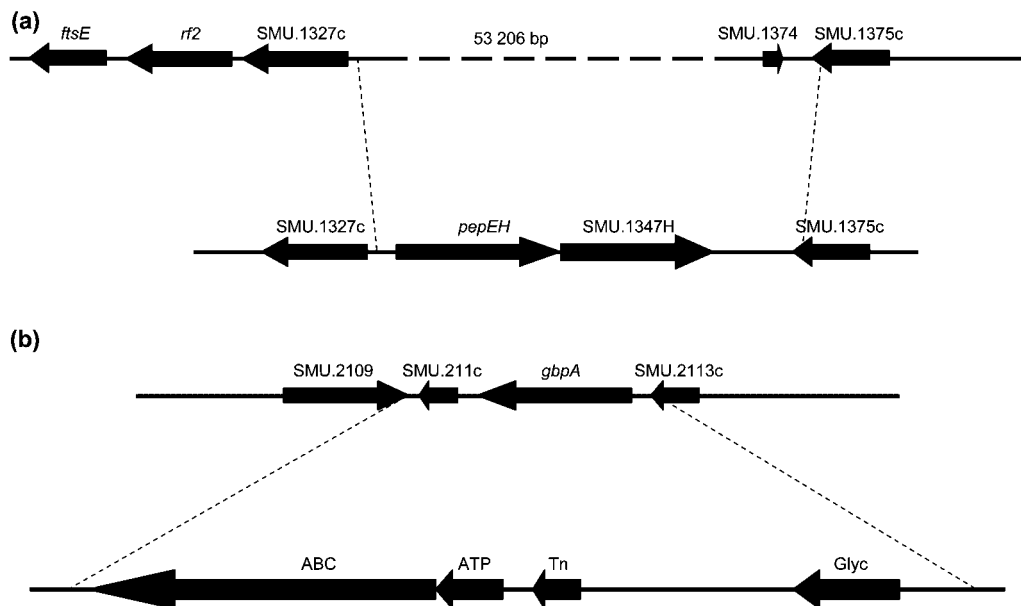
**Mutacin activity.** Production of mutacins and mutacin-like substances was detected by the delayed antagonism test. Overnight broth cultures of producer strains were stabbed into plates containing

THYE broth agar and incubated for 24 h at 37 °C. The plates were then overlaid with THYE soft agar containing 0.1 ml overnight culture of indicator strain. Plates were observed for zones of inhibition after 24 h incubation. All 39 strains were tested against a panel of indicator strains comprising *S. mutans* UA159, GS5 and NCTC1049, and representative strains of the oral streptococci *Streptococcus sobrinus*, *Streptococcus sanguis*, *Streptococcus oralis*, *Streptococcus mitis*,

**Table 2.** Primers used for PCR amplification

Primers for amplifying the *gtfB*, *gtfC* region and serotype-specific genes were as described by others (Hazlett *et al.*, 1998; Nakano *et al.*, 2004; Shibata *et al.*, 2003).

Primer	Sequence	Location	GenBank accession no.
SMU.1325F	5'-TTCAATGGCAACAACACGAT-3'	SMU.1325 ( <i>ftsE</i> )	AE014133
SMU.1326F	5'-TGTCCGATGATCTTTCACCA-3'	SMU.1326	AE014133
SMU.1327F	5'-TTCACCCAAAGCCCAGATAG-3'	SMU.1327	AE014133
SMU.1335F	5'-CTTCATTCCCTCAGCAATAAC-3'	SMU.1335	AE014133
SMU.1375R	5'-GGTTGCTGGATGCTGGTTAT-3'	SMU.1375	AE014133
SMU.1367F	5'-AAAACCTAGCGGTGCTGTTG-3'	Within SMU.1367	AE014133
SMU.1367R	5'-CAATTTTTGAAAACCCAATGC-3'		
SMU.1367HF	5'-TCTTTTTGCTCTGGCAGTGTT-3'	Within SMU.1367H	This paper
SMU.1367HR	5'-TCATCTCATCGCCTTCTTCC-3'		
smbF	5'-GTTTTTGATCCGCAAGCAGT-3'	SMU.1943 ( <i>leuRS</i> )	AE014133
smbR	5'-GCTCTTGCCAACGACAAAAT-3'	SMU.1942	
mutIF	5'-ATGCCGAGCTGAAAGAAAAA-3'	SMU.650 ( <i>ats</i> )	AE014133
mutIR	5'-CCAGACTAGCATGGTGCTCA-3'	SMU.659	
mutIIF	5'-AGCGGTTGAAAACAGTCAGG-3'	SMU.97 ( <i>pyrG</i> )	AE014133
mutIIR	5'-GCTGAAACGATTGCCATTTT-3'	SMU.99 ( <i>fba</i> )	
IS199F	5'-GCCGTCATTTACAGATGAG-3'	Within IS199	L23843
IS199R	5'-TTCCACCAGTGGACATAATCC-3'		
ISSmu3F	5'-TGGAGTTGGAGTGGACACAA-3'	Within ISSmu3	AY205145
ISSmu3R	5'-ACTGGATTGGACTTGCCAAA-3'		
IS3F	5'-TGAAGTGCACCCCAAAAGTT-3'	Within IS3	AE014133
IS3R	5'-TCTGGGCTGTTTCCTTACG-3'		



**Fig. 1.** (a) Gene arrangement of the TnSmu2 region of the *S. mutans* chromosome in strains UA159 (top line) and LML7 (lower line). (b) Gene arrangement of the *gbpA* region of the *S. mutans* chromosome in strains UA159 (top line) and V1996 (lower line). Abbreviations are those used in the GenBank annotation of the genome (Ajdic *et al.*, 2002) and as described in the text.

*Streptococcus salivarius*, *Streptococcus gordonii*, *Streptococcus constellatus*, *Staphylococcus aureus*, *Enterococcus faecium*, *Enterococcus faecalis* and *Micrococcus luteus*.

**Peptidase assay.** Extracts of *S. mutans* were prepared as described by Simpson & Russell (1998) and assayed for activity against *N*-acetyl-L-aspartic acid  $\alpha$ -(*p*-nitroanilide) (Sigma) by measuring *p*-nitroaniline release at  $A_{410}$  (Lassy & Miller, 2000).

**Collagen binding.** Microtitre trays were coated overnight at 4 °C with 10  $\mu$ g collagen type I  $\text{ml}^{-1}$  (Sigma) in 0.1 M acetic acid. The wells were washed three times with PBS and blocked for 1.5 h with 5% BSA in PBS. Then the wells were again washed with PBS, 0.01% Tween 20. Cells from overnight cultures of *S. mutans* grown in brain–heart infusion medium were collected by centrifugation, resuspended in PBS at  $\text{OD}_{500}$  1.0 and added to the wells. After 3 h, adherent cells were washed three times with PBS then fixed with 25% (v/v) formaldehyde at room temperature for 30 min. After a further three washes with PBS, adherent cells were stained with 0.5% crystal violet in water for 1 min, washed three times with PBS and the dye dissolved by adding 7% acetic acid before measuring the  $A_{595}$ .

## RESULTS

### TnSmu2 region

The *S. mutans* UA159 genome contains a large region of sequence of approximately 50 kb, located at around 7 o'clock on the circular genome map (approx. base pairs 1 250 000–1 300 000), that has a significantly lower G + C content than the rest of the genome (Ajdic *et al.*, 2002; <http://www.stdgen.lanl.gov/oragen/>). This region, named TnSmu2, is flanked by multiple remnants of transposases and includes several large ORFs similar to gramicidin and bacitracin synthases (Table 3). There is also an exact duplication of the gene pairs SMU.1347c/SMU.1348c and SMU.1365c/SMU.1367c.

In order to investigate the occurrence of TnSmu2 in our collection of strains, chromosomal DNA from each was used as template in PCR reactions with primers based in ORFs SMU.1325 (*ftsE*), which has a G + C content characteristic of the genome as a whole (36.82 mol% G + C), and SMU.1338, a low G + C region that is predicted to lie within TnSmu2 (Tables 2 and 3). Amplicons of identical size were produced from UA159, GS5, NCTC 10449, 3209, LML3 and 28, indicating that all these strains contain TnSmu2 and that they most probably have a similar map location to that in UA159. Using a forward primer based in SMU.1325 together with a reverse one based in SMU.1375, outside the far end of TnSmu2, no amplicon was observed from the above six strains due to the expected large size. However, 21 strains yielded a product of 5.5 kb and 2 strains a product of 4.5 kb, indicating that SMU.1325 and SMU.1375 are much closer together than in UA159. Eleven strains gave no product, presumably because they contain an insert too large to detect and/or because their primer target sequence differs from that in UA159. However, most of these strains contained SMU.1367 with sequence typical of UA159 (see below) suggesting that they also contain TnSmu2. The results are summarized in Table 2.

To examine more closely the affected region, PCR reactions were performed with forward primers based in SMU.1326 and SMU.1327. Both of these gave products for the 21 strains when paired with a reverse primer in SMU.1375. However, a forward primer based in SMU.1335 paired with the reverse primer in SMU.1375 gave a product only with the strains already shown to carry TnSmu2, suggesting that SMU.1335 is absent from the majority of strains. At the other end of the anomalous region, the strains that do not carry TnSmu2 showed a variety of patterns as some gave PCR products with primers located in SMU.1347 and its duplicate SMU.1365, SMU.1361, SMU.1367 and/or SMU.1369 indicating that there is extensive sequence variation in this IS-rich region (data not shown).

Strains Ingbritt and LML7 were chosen for sequencing, using the PCR product obtained with primers based in SMU.1325 and SMU.1375. The nucleotide sequences were aligned with that of UA159 and showed that all three matched exactly up to a point at base number 1 252 041 between SMU.1327 and SMU.1329, one of the predicted transposases flanking TnSmu2. Thereafter, a stretch of 1849 bp replaces 53 206 bp of UA159 sequence up to position 1 305 128 in SMU.1375c, where alignment with UA159 sequence recommences (Fig. 1a). Ingbritt and LML7 contained an ORF that shows homology to the *pepE*, peptidase E, gene of *Salmonella enterica* that encodes a peptidase specific for aspartyl dipeptides (Larsen *et al.*, 2001; Lassy & Miller, 2000) and has been named *pepEH*. Immediately adjacent to *pepEH*, and probably cotranscribed with it, is an ORF that a BLASTX search of the databases shows to most closely resemble SMU.1367 of UA159, named SMU.1367H. Supplementary Fig. S1 shows the sequence alignment (available with the online version of this paper). A search of the available finished and unfinished bacterial genome databases indicated that a pair of ORFs homologous to those found in LML7 and Ingbritt exists in the other oral streptococci, *S. gordonii* and *S. sobrinus* and *Leuconostoc mesenteroides*, but not in other species of streptococci. The *pepEH* ORF in the oral streptococci is not as long as that in *Salmonella enterica* and lacks several of the residues essential for peptidase activity (see Supplementary Fig. S2 available with the online version of this paper), suggesting that they are not functioning as peptidases. No aspartyl dipeptidase activity could be detected in any of the strains of *S. mutans* in this study. SMU.1367H ORFs in Ingbritt and LML7 differ from each other at a few nucleotide positions, but have identical predicted amino acid sequence and show substantial divergence from that predicted for SMU.1367 in UA159 (see Supplementary Fig. S1). Despite these differences, the ORFs are of almost the same length and have the same predicted function as an *S*-adenosyl-L-methionine methyltransferase. PCR primers that distinguish between SMU.1367 and SMU.1367H were used to examine the distribution of the two sequence variants in the strain collection and it was found that only one form was present in each strain. Those without TnSmu2 all resembled LML7 and Ingbritt (Table 1).

**Table 3.** ORFs in the TnSmu2 region showing their GenBank annotation and G+C content

The overall composition of the UA159 genome is 37.46 mol% G+C. Asterisks indicate ORFs missing in strains Ingbritt and LML7.

ORF	mol% G+C	Annotation
SMU.1324	35.04	<i>ftsX</i>
SMU.1325	38.81	<i>ftsE</i>
SMU.1326	41.92	rf2 peptide chain release factor
SMU.1327	37.16	Conserved hypothetical protein, ferredoxin?
SMU.1329*	42.19	Transposase
SMU.1330*	37.58	Transposase, IS1167
SMU.1331*	40.47	Transposase
SMU.1332*	42.04	Transposase
SMU.1334*	26.40	<i>sfp</i> phosphopantetheinyl transferase
SMU.1335*	32.31	<i>trans</i> -2-Enoyl-ACP reductase
SMU.1336*	27.68	<i>pksD</i> polyketide synthesis
SMU.1337*	25.20	$\alpha/\beta$ Superfamily hydrolases
SMU.1338*	29.09	ABC membrane spanning permease
SMU.1339*	28.37	<i>bacD</i> bacitracin synthase (168 260 kDa)
SMU.1340*	29.56	<i>bacA2</i> bacitracin/gramicidin/tyrocidin synthase (185 505 kDa)
SMU.1341*	29.40	Gramicidin/mycosubtilin synthase (141 661 kDa)
SMU.1342*	28.17	<i>bacA1</i> bacitracin synthase (313 640 kDa)
SMU.1343*	29.07	Polyketide synthase (126 826 kDa)
SMU.1344*	26.60	Malonyl CoA-acyl carrier protein transacylase
SMU.1345*	28.33	Peptide synthase similar to <i>mycA</i>
SMU.1346*	28.05	<i>bacT</i> thioesterase II-like protein
SMU.1347*	29.57	ABC transporter permease protein (= SMU.1365)
SMU.1348*	33.61	ATP-binding protein (= SMU.1366)
SMU.1349*	29.68	TetR/AcrR family
SMU.1351*	38.21	Transposase
SMU.1352*	36.11	Transposase
SMU.1353*	37.17	Transposase
SMU.1354*	40.61	Transposase
SMU.1355*	38.77	Transposase
SMU.1356*	38.77	Transposase
SMU.1357*	33.33	Transposase
SMU.1358*	42.15	Transposase
SMU.1359*	23.42	Hypothetical protein
SMU.1360*	32.55	Hypothetical protein
SMU.1361*	29.03	Transcriptional regulator
SMU.1363*	35.10	Transposase
SMU.1365*	29.58	ABC transporter permease protein (= SMU.1347)
SMU.1366*	33.61	ATP-binding protein (= SMU.1366)
SMU.1367	34.66	Hypothetical protein
SMU.1368	35.00	Hypothetical protein
SMU.1369	40.65	Hypothetical protein
SMU.1370	37.05	Transposase
SMU.1372	40.74	Transposase
SMU.1373	36.61	Hypothetical protein
SMU.1374	34.80	Hypothetical protein
SMU.1375	35.31	Hypothetical protein <i>yqkA</i>
SMU.1377	42.11	Hypothetical protein

### **gbpA region**

The gene *gbpA* encodes a glucan-binding protein that consists mainly of a series of amino acid repeats similar to

those found at the carboxy-terminus of glucosyltransferases (Banas *et al.*, 1990). During attempts to amplify the *gbpA* gene by PCR using primers based on the published

sequence, F. Macrina, Virginia Commonwealth University, (personal communication) found that some strains, including V403, V1995 and V1996, failed to yield an amplicon. The *gfpA* gene was also undetectable in these strains using Southern blotting. We confirmed the observation in the three strains with primers internal to *gfpA*, and screening of our collection revealed that strains 34 and LML5 isolated in Newcastle also failed to yield a product. Examination of culture supernatants of strains V403, V1995, V1996, 34 and LML5 using Western blotting and antiserum specific to GbpA or biotin-labelled dextran showed that these strains failed to produce GbpA (data not shown).

The relationship of *gfpA* and flanking genes is shown in Fig. 1(b). SMU.2113c encodes a hypothetical protein, as does SMU.2111c. SMU.2109c encodes a putative multidrug efflux pump. PCR amplification of the *gfpA* region using primers based on flanking sequence yielded the expected 3.2 kb amplicon for UA159 and also for 35/39 of the other strains screened (Table 1). In strains V403, V1995, V1996, 34 and LML5, an amplicon of approximately 10 kb was obtained with these primers (Table 1). V1996 was selected for more detailed analysis of this region, and the nucleotide sequence of the 10 kb amplicon was determined by sub-cloning fragments into plasmids and by primer walking. The points at which the nucleotide sequence alignment of the novel sequence with the UA159 genome sequence stopped were located at 1 979 630 and 1 982 111 bp, and showed the absence of 2418 bp (comprising all of SMU.2111c and *gfpA*, and 25 bp of the 3' end of SMU.2113c) in V1996 and its replacement by 8930 bp (Fig. 1b). This novel sequence encoded a long ORF predicted to be an ABC transporter and putative exporter of polyketide antibiotics (COG 3559.1, Pfam 02687.11) with immediately upstream a predicted ATP-binding protein. Also in this region is a transposase belonging to the IS3 family of IS elements and an incomplete ORF that shows homology to glycosyltransferases of a type not found elsewhere in the UA159 genome. The G+C content of the 10 kb region was 33.5 mol%, similar to that of the UA159 genome as a whole.

In comparison with its parent strain 3209, an isogenic *gfpA* mutant showed altered sucrose-dependent aggregation (Banas & Gilmore, 1991). However, within our collection of strains there was extensive variation in the appearance of sucrose-containing broth cultures after overnight incubation and there was no correlation with the presence or absence of *gfpA* (not shown). This finding can be explained by the fact that multiple enzymes and glucan-binding proteins are involved in the formation, breakdown and binding of the polymers derived from sucrose (Colby & Russell, 1997; Shah & Russell, 2004) so eight or more gene products may contribute to the aggregation phenomenon. This interaction between multiple genes is supported by the findings of Hazlett *et al.* (1998), who found that when the 3209 *gfpA* mutant was inoculated into gnotobiotic rats, there was an increase in the proportion of the population in which *gtfB* and *gtfC* had recombined to give a chimeric

*gtfBC* gene. The two alternatives can be distinguished by PCR using primers flanking the *gtfB gtfC* region (Hazlett *et al.*, 1998), and within the strain collection we could detect both 10 kb amplicons representing both genes intact and 5 kb amplicons indicating recombination had occurred. Most strains contain both alternatives within a single population, as described by Hazlett *et al.* (1998). However, the absence of *gfpA* does not appear to have resulted in long-term selection of *gtfBC* recombinants in V403, V1995, V1996, 34 or LML5 (Table 1).

### ***cnm* region**

It has recently been reported that 5/14 of a collection of Japanese strains of *S. mutans* carry the *cnm* gene encoding a wall anchored 120 kDa collagen-binding protein (Sato *et al.*, 2004a, b). We therefore examined our collection of strains for the ability to bind collagen, and found that only strains V403, L13 and LML8 bound to collagen immobilized in a microtitre tray. Primers based within the *cnm* gene gave PCR products from only these three strains.

### **Mutacin production**

The collection of 39 strains was tested for the production of mutacins or other inhibitor substances against a panel of indicator strains. A total of 3 strains (L13, 34 and 66) inhibited all 15 indicators and other strains showed a wide variety of inhibition patterns, inhibiting at least 4 indicators. The strains containing TnSmu2 were amongst the least inhibitory strains, and there was no correlation between presence of inserts in the *mutI*, *mutII* or *smb* regions (see below) and the spectrum of inhibition (see the Supplementary Table available with the online version of this paper).

### ***mutI* region**

Mutacins I and III are encoded by a similar set of genes encoding lantionine-containing small peptide antibiotics (lantibiotics), made up of 23 and 24 aa, respectively. Their synthesis is determined by a series of *mut* genes that are located in a 16 kb region next to the *ats* gene, alanyl-tRNA synthase, in strains that have been shown to produce mutacin I or III (Qi *et al.*, 1999a, 2000). UA159 has four *mut* genes in the *ats* region but does not possess the full set required for synthesis and secretion of active mutacin. Primers were designed based in *ats* (SMU.650) and SMU.659, which corresponds to the conserved *orfY* outside the far end of the *mut* region (Qi *et al.*, 2000). UA159 gave the 7.4 kb amplicon predicted from the genome sequence and most of the other strains gave products of this size or smaller. Only strains L13 and 28 gave products of about 16 kb, similar to that found in mutacin-producing strains.

### ***mutII* region**

The gene encoding lantibiotic mutacin II is also absent from UA159 but in strain T8 is flanked by sequence corresponding to *pyrG* and *fbaA* (Chen *et al.*, 1999; Qi *et al.*, 1999b).

Primers based in these genes were, therefore, used to amplify the intervening region from all strains. One strain, LML8, contained an 11 kb insert of the size found in T8. All other strains gave the 1.2 kb product predicted from the UA159 genome.

### ***smb* region**

Smb is a dipeptide lantibiotic not found in UA159, but present in GS5, encoded by a set of genes on a 9.5 kb fragment that includes several transposases and occurs at a position corresponding to a point between SMU.1942c and *syl* on the UA159 genome (Yonezawa & Kuramitsu, 2005). Primers based in SMU.1942c and *syl* were used to screen the strain collection for the presence of the *smb* island and two were found to have a 9.5 kb insert corresponding to that in GS5 (Table 2).

### **IS elements**

The genome of UA159 contains numerous scattered complete and fragmented IS elements, the most frequent being members of the IS3 family with 7 complete elements and 15 fragments (Ajdic *et al.*, 2002). On the basis of the complete elements, primers were designed to amplify IS3. Other IS elements reported in *S. mutans* are IS199 (Macrina *et al.*, 1996) and the defective ISSmu3 associated with loss of the *msm* and *gal* genes (Robinson *et al.*, 2003). ISSmu3 was detected only in those strains previously found to have the *msm/gal* deletion. The occurrence of IS3 and IS199 in the strain collection is presented in Table 2.

## **DISCUSSION**

*S. mutans* UA159 was selected as the strain for genome sequencing because it displays a wide range of phenotypic properties considered 'wild-type' for the species, with the added advantage that it has been shown to be transformable and to be capable of inducing caries in experimental animals. The strain also appears to be a good representative at the genetic level. For example, the map derived from the genome sequence agrees with earlier simple linkage maps of other strains obtained by PFGE (Cappiello *et al.*, 1999; Okahashi *et al.*, 1990), and there are numerous examples where hybridization probes or PCR primers based on UA159 detect corresponding DNA sequence in other strains of *S. mutans* (and vice versa). The UA159 genome sequence thus provides a scaffold for comparative studies with a range of strains and an exploration of the extent of diversity within the species. Recent developments in the study of bacterial genomes have made it abundantly clear that the sequencing of a single genome is unlikely ever to provide full information about the species. In those instances where a number of genomes have been characterized, it has been found that there exists a 'core genome' of universally present genes and a 'dispensable genome' that represents a pool of genetic material that may be found in a variable number of strains within the species. The terms metagenome, supra-genome and pan-genome have been proposed for the

full portfolio of genetic material (Banks *et al.*, 2004; Fux *et al.*, 2005; Tettelin *et al.*, 2005). In a freely recombining species the variety of components of this pan-genome may approach infinity, though bacterial species that have evolved recently are genetically restricted and species that occupy a tightly defined niche may show very little diversity (Medini *et al.*, 2005).

The data presented in this paper combined with other recent reports (Fujiwara *et al.*, 1998; Old *et al.*, 2006; Robinson *et al.*, 2003; Sato *et al.*, 2004a; Saxena *et al.*, 2005) show that *S. mutans*, like other streptococci, has core and dispensable genes. Core genes would be expected to be those essential for cellular function or survival, and for which there is a powerful selective force to ensure they are maintained. Within the dispensable genome, a number of different classes of genes can be discerned. The first are those that are typical of the *S. mutans* genome in terms of G + C content, codon usage and being members of paralogous families. Within this group can be placed *msm*, *gal*, *bgl*, *cel* (Old *et al.*, 2006; Robinson *et al.*, 2003) and *gbpA*, *cmm*, *pepEH* (this paper). Other examples are the *gtfB/gtfC* pair, which sometimes recombine to give a single *gtfBC* gene (Hazlett *et al.*, 1998; Mattos-Graner *et al.*, 2004). All of these genes must give a competitive advantage under certain conditions and the frequency with which they occur may be a reflection of the selective pressure for maintaining them in the pool. An interesting case is SMU.1367, which may have an essential function but occurs in at least two sequence variants (represented by strains UA159 and LML7) that appear to encode the same functional protein from different nucleotide sequences and have diverged from a common ancestor.

Genes encoding mutacins can be included within the group of dispensable genes, since these seem to be distinctive to *S. mutans*. It is well known that there is extensive variation in the possession of mutacin genes (Bekal-Si Ali *et al.*, 2002) but in this paper we have focused on their insertion at specific locations on the genome map derived from UA159. Genes encoding MutI/MutIII, MutII and Smb mutacins all seem to have specific chromosomal locations and the identification of stable flanking genes will facilitate future studies of variation in the number of genes, and their coding function, within the mutacin loci. For example, UA159 has only 4 of the 11 or more ORFs associated with production of mutacin I (Qi *et al.*, 2000) and a variety of insert sizes was observed for the other strains in our collection (data not shown).

The second major class of dispensable genes is composed of fragments of foreign DNA, occurring in recognizable genomic islands with G + C content and maybe codon preference different from the bulk of the *S. mutans* genome. TnSmu2 is the largest such island in UA159 and has a G + C content more typical of members of the *Bacillaceae*. Some of the ORFs in this region (SMU.1339–SMU.1343) are the largest in the genome but even these are well below the size of the complex multi-functional enzymes that

synthesize peptide antibiotics such as bacitracin, gramicidin and tyrocidin (Eppelmann *et al.*, 2001; Kratzschmar *et al.*, 1989; Mootz & Marahiel, 1997). It thus appears that ORFs that were larger in their original host have been fragmented and inactivated at the time of or after their transfer to *S. mutans*. Our examination of the inhibition spectrum of strains carrying or lacking TnSmu2 did not provide any evidence that this foreign DNA is responsible for the production of antimicrobials in *S. mutans*.

The introduction of TnSmu2, or the multiple IS elements associated with it, resulted in displacement of *pepEH*. This gene shows properties that suggest it is a genuine *S. mutans* gene, a fact supported by the observation that it also occurs in other related streptococci, though its function is unknown. The *gfpA* region studies in this paper provided another example of gene displacement by incoming DNA. *gfpA* also appears to be a genuine *S. mutans* gene in terms of G+C composition, codon usage and its close relationship to the glucan-binding domains of GtfB, GtfC and GtfD (Banas *et al.*, 1990). The genes that have displaced it in strains such as V1996 are one encoding a transposase typical of an IS element and a pair of genes encoding an ABC transport system, possibly for export of polyketides. There are several paralogues of these ABC genes in the UA159 genome (SMU.1166c/SMU.1167c, SMU.1927/PsaB and SMU.1347c/SMU.1348c within TnSmu2); their closest match in BLASTP searches is to genes from the sulphate-reducing archaeon *Archaeoglobus fulgidus*. Although their G+C composition is not outside the range found in *S. mutans* genes, it seems highly probable that these too are foreign genomic islands.

We have previously speculated that the occurrence of chromosomal deletions might provide a useful marker for following the evolution of *S. mutans*. Since deletions are irreversible genetic events all clonal descendants of a strain undergoing a deletion should also carry the deletion. As a corollary, deletions might be expected to accumulate within a clonal line. However, if all the features listed in Table 1 are regarded as insertion/deletion events, it is immediately apparent that there is no clear pattern of linkage of different events, with the possible exception of the deletions of *msm* and *bgl* found in a subset of strains originally selected because of their distinct phenotype (Robinson *et al.*, 2003). These deletions were found in strains isolated from sources widely separated by geography and time, and the same is true of the other genetic changes detected. Either the same genetic events have taken place independently in different places at different times, or the ancestral changes have become disseminated through the *S. mutans* population by a process of horizontal transfer. Attempts to analyse the data in Table 1 by a variety of clustering techniques have failed to reveal any pattern of relationships. This suggests that the dispensable genome is in extensive flux, with free assortment of the genetic markers being horizontally exchanged within the worldwide *S. mutans* population as a whole. It should, however, be noted that the collection

of strains studied here was in no way structured, and is restricted to examination of regions of the chromosome known or suspected to be dispensable and subject to variation. It is, therefore, not possible to determine how the regions studied may relate to any clonal structure that might be revealed by examination of the phylogenetic relationships between the strains based on data on housekeeping genes in the core genome, as applied by Balakrishnan *et al.* (2002) in their study of mutacins. The dispensable genome may not reflect the true phylogeny of the core genome or population structure of *S. mutans*, which can better be explored by multilocus sequence typing (Feil, 2004). Nevertheless, the results do suggest that there has been widespread exchange of genetic material both between strains of *S. mutans* and with other species.

*S. mutans* was long considered to be poorly transformable so recombination events would be rare. However, Li *et al.* (2001) have reported that all the *com* competence genes are present, and suggested that transformation may be a frequent occurrence in biofilms. The real extent of gene transfer is as yet unknown but the apparent stability of RFLP patterns suggests that chromosomal rearrangements are not common in *S. mutans* once established in the oral cavity. However, mobile IS elements provide a mechanism for genetic change and in one well-studied example IS199 in *S. mutans* V403 was shown to be functional and to undergo transposition *in vitro* under strong selection pressure (Macrina *et al.*, 1991). Whether other elements, such as IS3 in UA159, are active has yet to be determined. However, it is recognized within the *S. mutans* research community that phenotypic changes have occurred in stocks of long-established laboratory strains. For example, stocks of Ingbritt and GS5 held in different laboratories display variation in properties (Ellwood *et al.*, 1976; Gibbons *et al.*, 1966; Tao *et al.*, 1993) and there are reports of changes during multiple laboratory subcultures (Cvitkovitch & Hamilton, 1994; McBride *et al.*, 1984; Russell & Smith, 1986). This study provides two examples where consequential genetic changes may have been precipitated by laboratory manipulation. Strain LT11, which was derived by chemical mutagenesis of UA159 (Tao *et al.*, 1993), has lost TnSmu2 and the IS3 elements detected in its parent UA159 (Table 1). Similarly, V1996, a mutant of V403 in which *gtf* and *ftf* genes have been knocked out by allelic exchange (Munro *et al.*, 1991), has also undergone further change. A further example of how introduction of one mutation can induce changes in other genes comes from the work of Hazlett *et al.* (1998, 1999) who constructed a *gfpA* knockout mutant and found that the mutation altered biofilm formation and virulence. However, the phenotypic changes could not solely be attributed to lack of a functional GbpA because introduction of the mutation resulted in accumulation within the population of variants in which *gtfB* and *gtfC* had recombined.

This study has confined itself to an investigation of potentially dispensable genes identified either by their anomalous

base composition or because they encode functions that are known to be variable within *S. mutans*. However, base composition and codon bias anomalies will not detect all foreign DNA (Koski *et al.*, 2001), and the phenotype that might be associated with loss of a very large proportion of the genes cannot be predicted or easily detected. Comparison of the gene content of individual strains with a reference microarray representing a sequenced genome overcomes these problems, and we are currently analysing the results of DNA–DNA hybridization experiments with strains from our collection and a microarray of oligonucleotides covering the ORFs of UA159.

## ACKNOWLEDGEMENTS

We are grateful for the assistance of Saskia Crowe, Nicola Griffiths, Lesley Old, Chris Riddell, Susanna Sherwin and Estelle Tarroux in screening the strain collection.

## REFERENCES

- Ajdic, D., McShan, W. M., McLaughlin, R. E. & 16 other authors (2002). Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proc Natl Acad Sci U S A* **99**, 14434–14439.
- Alaluusua, S., Alaluusua, S. J., Karjalainen, J. & 7 other authors (1994). The demonstration by ribotyping of the stability of oral *Streptococcus mutans* infection over 5 to 7 years in children. *Arch Oral Biol* **39**, 467–471.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Balakrishnan, M., Simmonds, R. S., Kilian, M. & Tagg, J. R. (2002). Different bacteriocin activities of *Streptococcus mutans* reflect distinct phylogenetic lineages. *J Med Microbiol* **51**, 941–948.
- Banas, J. A. & Gilmore, K. S. (1991). Analysis of *Streptococcus mutans* and *Streptococcus downei* mutants inactivated in the *gfp* and *gtfA* genes. In *Genetics and Molecular Biology of Streptococci, Lactococci and Enterococci*, pp. 282–283. Edited by G. M. Dunne, P. P. Cleary & L. L. McKay. Washington, DC: American Society for Microbiology.
- Banas, J. A., Russell, R. R. B. & Ferretti, J. J. (1990). Sequence analysis of the gene for the glucan-binding protein of *Streptococcus mutans* Ingbritt. *Infect Immun* **58**, 667–673.
- Banks, D. J., Porcella, S. F., Barbian, K. D. & 7 other authors (2004). Progress toward characterization of the group A *Streptococcus* meta-genome: complete genome sequence of a macrolide-resistant serotype M6 strain. *J Infect Dis* **190**, 727–738.
- Bekal-Si Ali, S., Hurtubise, Y., Lavoie, M. C. & LaPointe, G. (2002). Diversity of *Streptococcus mutans* bacteriocins as confirmed by DNA analysis using specific molecular probes. *Gene* **283**, 125–131.
- Brady, L. J., Crowley, P. J., Ma, J. K., Kelly, C., Lee, S. F., Lehner, T. & Bleiweis, A. S. (1991). Restriction fragment length polymorphisms and sequence variation within the *spaP* gene of *Streptococcus mutans* serotype c isolates. *Infect Immun* **59**, 1803–1810.
- Cappiello, M. G., Hantman, M. J., Zuccon, F. M., Peruzzi, F., Amjad, M., Piggot, P. J. & Daneo-Moore, L. (1999). Physical and genetic map of *Streptococcus mutans* GS-5 and localization of five rRNA operons. *Oral Microbiol Immunol* **14**, 225–232.
- Caufield, P. W., Ratanapridakul, K., Allen, D. N. & Cutter, G. R. (1988). Plasmid-containing strains of *Streptococcus mutans* cluster within family and racial cohorts: implications for natural transmission. *Infect Immun* **56**, 3216–3220.
- Chen, P., Qi, F., Novak, J. & Caufield, P. W. (1999). The specific genes for lantibiotic mutacin II biosynthesis in *Streptococcus mutans* T8 are clustered and can be transferred en bloc. *Appl Environ Microbiol* **65**, 1356–1360.
- Colby, S. M. & Russell, R. R. B. (1997). Sugar metabolism by mutans streptococci. *Soc Appl Bacteriol Symp Ser* **26**, 80S–88S.
- Cvitkovitch, D. G. & Hamilton, I. R. (1994). Biochemical change exhibited by oral streptococci resulting from laboratory subculturing. *Oral Microbiol Immunol* **9**, 209–217.
- de Soet, J. J., Nyvad, B. & Kilian, M. (2000). Strain-related acid production by oral streptococci. *Caries Res* **34**, 486–490.
- Ellwood, D. C., Baird, J. K., Hunter, J. R. & Longyear, V. M. (1976). Variations in surface polymers of *Streptococcus mutans*. *J Dent Res* **55**, C42–C49.
- Eppelmann, K., Doekel, S. & Marahiel, M. A. (2001). Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host *Bacillus subtilis*. *J Biol Chem* **276**, 34824–34831.
- Feil, E. J. (2004). Small change: keeping pace with microevolution. *Nat Rev Microbiol* **2**, 483–495.
- Ferretti, J. J., Russell, R. R. B. & Dao, M. L. (1989). Sequence analysis of the wall-associated protein precursor of *Streptococcus mutans* antigen A. *Mol Microbiol* **3**, 469–478.
- Ferretti, J. J., Ajdic, D. & McShan, W. M. (2004). Comparative genomics of streptococcal species. *Indian J Med Res* **119**, 1–6.
- Fujiwara, T., Terao, Y., Hoshino, T., Kawabata, S., Ooshima, T., Sobue, S., Kimura, S. & Hamada, S. (1998). Molecular analyses of glucosyltransferase genes among strains of *Streptococcus mutans*. *FEMS Microbiol Lett* **161**, 331–336.
- Fujiwara, T., Nakano, K., Kawaguchi, M., Ooshima, T., Sobue, S., Kawabata, S., Nakagawa, I. & Hamada, S. (2001). Biochemical and genetic characterization of serologically untypable *Streptococcus mutans* strains isolated from patients with bacteremia. *Eur J Oral Sci* **109**, 330–334.
- Fux, C. A., Shirtliff, M., Stoodley, P. & Costerton, J. W. (2005). Can laboratory reference strains mirror ‘real-world’ pathogenesis? *Trends Microbiol* **13**, 58–63.
- Gibbons, R. J., Berman, K. S., Knoettner, P. & Kapsimalis, B. (1966). Dental caries and alveolar bone loss in gnotobiotic rats infected with capsule forming streptococci of human origin. *Arch Oral Biol* **11**, 549–560.
- Gronroos, L. & Alaluusua, S. (2000). Site-specific oral colonization of mutans streptococci detected by arbitrarily primed PCR fingerprinting. *Caries Res* **34**, 474–480.
- Hazlett, K. R. O., Michalek, S. M. & Banas, J. A. (1998). Inactivation of the *gfpA* gene of *Streptococcus mutans* increases virulence and promotes *in vivo* accumulation of recombinations between the glucosyltransferase B and C genes. *Infect Immun* **66**, 2180–2185.
- Hazlett, K. R. O., Mazurkiewicz, J. E. & Banas, J. A. (1999). Inactivation of the *gfpA* gene of *Streptococcus mutans* alters structural and functional aspects of plaque biofilm which are compensated by recombination of the *gtfB* and *gtfC* genes. *Infect Immun* **67**, 3909–3914.
- Herbert, M. A., Beveridge, C. J., McCormick, D., Aten, E., Jones, N., Snyder, L. A. & Saunders, N. J. (2005). Genetic islands of *Streptococcus agalactiae* strains NEM316 and 2603VR and their presence in other group B streptococcal strains. *BMC Microbiol* **5**, 31.
- Kohler, B. & Krasse, B. (1990). Human strains of mutans streptococci show different cariogenic potential in the hamster model. *Oral Microbiol Immunol* **5**, 177–180.

- Kohler, B., Birkhed, D. & Olsson, S. (1995). Acid production by human strains of *Streptococcus mutans* and *Streptococcus sobrinus*. *Caries Res* **29**, 402–406.
- Koski, L. B., Morton, R. A. & Golding, G. B. (2001). Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol Biol Evol* **18**, 404–412.
- Kratzschmar, J., Krause, M. & Marahiel, M. A. (1989). Gramicidin S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J Bacteriol* **171**, 5422–5429.
- Kulkarni, G. V., Chan, K. H. & Sandham, H. J. (1989). An investigation into the use of restriction endonuclease analysis for the study of transmission of mutans streptococci. *J Dent Res* **68**, 1155–1161.
- Larsen, R. A., Knox, T. M. & Miller, C. G. (2001). Aspartic peptide hydrolases in *Salmonella enterica* serovar typhimurium. *J Bacteriol* **183**, 3089–3097.
- Lassy, R. A. & Miller, C. G. (2000). Peptidase E, a peptidase specific for N-terminal aspartic dipeptides, is a serine hydrolase. *J Bacteriol* **182**, 2536–2543.
- Lee, S. F. & McGavin, M. K. (2004). Identification of a point mutation resulting in loss of cell wall anchoring activity of SrtA of *Streptococcus mutans* NG5. *Infect Immun* **72**, 4314–4317.
- Li, Y. & Caufield, P. W. (1995). The fidelity of initial acquisition of mutans streptococci by infants from their mothers. *J Dent Res* **74**, 681–685.
- Li, Y. H., Lau, P. C., Lee, J. H., Ellen, R. P. & Cvitkovitch, D. G. (2001). Natural genetic transformation of *Streptococcus mutans* growing in biofilms. *J Bacteriol* **183**, 897–908.
- Macrina, F. L., Jones, K. R., Alpert, C. A., Chassy, B. M. & Michalek, S. M. (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.
- Macrina, F. L., Jones, K. R. & Laloi, P. (1996). Characterization of IS199 from *Streptococcus mutans* V403. *Plasmid* **36**, 9–18.
- Mattos-Graner, R. O., Napimoga, M. H., Fukushima, K., Duncan, M. J. & Smith, D. J. (2004). Comparative analysis of Gtf isozyme production and diversity in isolates of *Streptococcus mutans* with different biofilm growth phenotypes. *J Clin Microbiol* **42**, 4586–4592.
- McBride, B. C., Song, M., Krasse, B. & Olsson, J. (1984). Biochemical and immunological differences between hydrophobic and hydrophilic strains of *Streptococcus mutans*. *Infect Immun* **44**, 68–75.
- Medini, D., Donati, C., Tettelin, H., Massignani, V. & Rappuoli, R. (2005). The microbial pan-genome. *Curr Opin Genet Dev* **15**, 589–594.
- Mootz, H. D. & Marahiel, M. A. (1997). The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol* **179**, 6843–6850.
- Munro, C., Michalek, S. M. & Macrina, F. L. (1991). Cariogenicity of *Streptococcus mutans* V403 glucosyltransferase and fructosyltransferase mutants constructed by allelic exchange. *Infect Immun* **59**, 2316–2323.
- Nakano, K., Nomura, R., Shimizu, N., Nakagawa, I., Hamada, S. & Ooshima, T. (2004). Development of a PCR method for rapid identification of new *Streptococcus mutans* serotype *k* strains. *J Clin Microbiol* **42**, 4925–4930.
- Napimoga, M. H., Kamiya, R. A. U., Rosa, R. T., Rosa, E. A. R., Hoffing, J. F., Mattos-Graner, R. D. & Goncavles, R. B. (2004). Genotypic diversity and virulence traits of *Streptococcus mutans* in caries-free and caries-active individuals. *J Med Microbiol* **53**, 697–703.
- Okahashi, N., Sasakawa, C., Okada, N., Yamada, M., Yoshikawa, M., Tokuda, M., Takahashi, I. & Koga, T. (1990). Construction of *NotI* restriction map of the *Streptococcus mutans* genome. *J Gen Microbiol* **136**, 2217–2223.
- Old, L. A., Lowes, S. & Russell, R. R. B. (2006). Genomic variation in *Streptococcus mutans*: deletions affecting the multiple pathways of  $\beta$ -glucoside metabolism. *Oral Microbiol Immunol* **21**, 21–27.
- Qi, F., Chen, P. & Caufield, P. W. (1999a). Purification of mutacin III from group III *Streptococcus mutans* UA787 and genetic analyses of mutacin III biosynthesis genes. *Appl Environ Microbiol* **65**, 3880–3887.
- Qi, F., Chen, P. & Caufield, P. W. (1999b). Functional analyses of the promoters in the lantibiotic mutacin II biosynthetic locus in *Streptococcus mutans*. *Appl Environ Microbiol* **65**, 652–658.
- Qi, F., Chen, P. & Caufield, P. W. (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.
- Robinson, W. G., Old, L. A., Shah, D. S. H. & Russell, R. R. B. (2003). Chromosomal insertions and deletions in *Streptococcus mutans*. *Caries Res* **37**, 148–156.
- Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365–386.
- Russell, R. R. B. (2003). Microbiological aspects of caries prevention. In *Prevention of Oral Diseases*, pp. 63–75. Edited by J. J. Murray, J. H. Nunn & J. G. Steele. Oxford: Oxford University Press.
- Russell, R. R. B. & Smith, K. (1986). Effect of subculturing on location of *Streptococcus mutans* antigens. *FEMS Microbiol Lett* **35**, 319–323.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945.
- Sato, Y., Okamoto, K. & Kizaki, H. (2002). *gbpC* and *pac* gene mutations detected in *Streptococcus mutans* strain GS-5. *Oral Microbiol Immunol* **17**, 263–266.
- Sato, Y., Okamoto, K., Kagami, A., Yamamoto, Y., Igarashi, T. & Kizaki, H. (2004a). *Streptococcus mutans* strains harboring collagen-binding adhesin. *J Dent Res* **83**, 534–539.
- Sato, Y., Okamoto, K., Kagami, A., Yamamoto, Y., Ohta, K., Igarashi, T. & Kizaki, H. (2004b). Application of *in vitro* mutagenesis to identify the gene responsible for cold agglutination phenotype of *Streptococcus mutans*. *Microbiol Immunol* **48**, 449–456.
- Saxena, D., Li, Y. & Caufield, P. W. (2005). Identification of unique bacterial gene segments from *Streptococcus mutans* with potential relevance to dental caries by subtraction DNA hybridization. *J Clin Microbiol* **43**, 3508–3511.
- Shah, D. S. H. & Russell, R. R. B. (2004). A novel glucan-binding protein with lipase activity from the oral pathogen *Streptococcus mutans*. *Microbiology* **150**, 1947–1956.
- Shibata, Y., Ozaki, K., Seki, M., Kawato, T., Tanaka, H., Nakano, Y. & Yamashita, Y. (2003). Analysis of loci required for determination of serotype antigenicity in *Streptococcus mutans* and its clinical utilization. *J Clin Microbiol* **41**, 4107–4112.
- Simpson, C. L. & Russell, R. R. B. (1998). Identification of a homolog of CcpA catabolite repressor protein in *Streptococcus mutans*. *Infect Immun* **66**, 2085–2092.
- Tao, L., MacAlister, T. J. & Tanzer, J. M. (1993). Transformation efficiency of EMS-induced mutants of *Streptococcus mutans* of altered cell shape. *J Dent Res* **72**, 1032–1039.
- Tettelin, H., Massignani, V., Cieslewicz, M. J. & 43 other authors (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ‘pan-genome’. *Proc Natl Acad Sci U S A* **102**, 13950–13955.

**Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

**Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997).** The CLUSTAL\_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

**Ueda, S. & Kuramitsu, H. K. (1988).** Molecular basis for the spontaneous generation of colonization-defective mutants of *Streptococcus mutans*. *Mol Microbiol* **2**, 135–140.

**Ushiro, I., Lumb, S. M., Aduse-Opoku, J., Ferretti, J. J. & Russell, R. R. B. (1991).** Chromosomal deletions in melibiose-negative isolates of *Streptococcus mutans*. *J Dent Res* **70**, 1422–1426.

**Yonezawa, H. & Kuramitsu, H. K. (2005).** Genetic analysis of a unique bacteriocin, Smb, produced by *Streptococcus mutans* GS5. *Antimicrob Agents Chemother* **49**, 541–548.