Oral Microbiology Immunology 2007: 22: 103–110
Printed in Singapore. All rights reserved

ORAL MICROBIOLOGY
AND IMMUNOLOGY

Comparative genome hybridization of *Streptococcus mutans* strains

J. C. Waterhouse[1], D. C. Swan[2], R. R. B. Russell[1]
[1]Oral Biology, School of Dental Sciences and [2]Bioinformatics Support Unit, Newcastle University, Newcastle upon Tyne, UK

Waterhouse JC, Swan DC, Russell RRB. Comparative genome hybridization of Streptococcus mutans *strains.*
Oral Microbiol Immunol 2007: 22: 103–110. © 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard.

The basis for genotypic and phenotypic variation within *Streptococcus mutans* is poorly understood but the availability of the genome sequence of strain UA159 provides an opportunity for comparative studies. Genomic DNA prepared from nine strains of *S. mutans* was used to probe a microarray consisting of oligonucleotides representing 1948 open reading frames of *S. mutans* UA159. A total of 385 (20%) of the UA159 open reading frames were found to be absent from one or more of the test strains. Absent open reading frames frequently occurred in blocks of adjacent open reading frames and represented regions previously experimentally detected by polymerase chain reaction, predicted genomic islands and insertion sequence elements as well as novel open reading frames. Approximately half appear to involve foreign DNA acquired by horizontal transmission. The results indicate the existence of distinct core and dispensable genomes and may help explain the phenotypic and genotypic variation within *S. mutans*.

Key words: genome; microarray; *Streptococcus mutans*

R.R.B. Russell, Oral Biology, School of Dental Sciences, Newcastle University, Newcastle upon Tyne NE2 4BW, UK
Tel.: +44 191 222 8240;
fax: +44 191 222 6137;
e-mail: r.r.russell@ncl.ac.uk
Accepted for publication July 24, 2006

Different isolates of *Streptococcus mutans* are known to vary in a range of phenotypic properties including serotype polysaccharide, ability to utilize various sugars, produce mutacins, bind macromolecules, synthesize polymers from sucrose, generate acid, and tolerate acid conditions. Variation in such properties is believed to underlie the known variation in cariogenicity in experimental animals, though attempts to correlate specific individual properties with caries in humans have been largely unsuccessful. Paralleling the phenotypic variation, genetic fingerprinting techniques have revealed an enormous degree of genetic variation, with distinct genotypes being carried by each person and (sometimes) their close contacts. However, the basis for this phenotypic and genotypic variation is poorly understood.

The main sources of genetic variation are mutation and recombination. Examples of point mutations in isolates of *S. mutans* include those altering the *pac*, *srtA* and *gbpC* genes, which affect surface properties (15, 25). Mutational deletions of groups of genes associated with carbohydrate metabolism have also been identified (20, 24, 31). These deletions may, in at least some cases, have been mediated by the transposition of insertion sequence elements (24) and insertion sequence transposition can also lead to inactivation of single genes (17). Mutations may also be generated by recombination between closely-linked regions of the chromosome as has been proposed for recombination between *gtfB* and *gtfC* to form a single hybrid GTF (30) and has been proposed for deletions of *bgl* genes (24). All these mechanisms have been known for many years but the advent of whole genome sequencing projects has radically altered our view of the complexity of bacterial populations (2). In particular, mobile genetic elements such as plasmids, bacteriophage, transposons, insertion sequence elements and genomic islands have been recognized as important agents generating diversity. Bacteriophages have not been demonstrated in *S. mutans* and plasmids are rare. However, bioinformatic analysis of the *S. mutans* UA159 genome revealed the presence of several putative genomic islands indicative of foreign DNA obtained by horizontal transfer as well as multiple intact and fragmented insertion sequences (1).

We have recently examined a collection of strains of *S. mutans* isolated in various parts of the world over the last 40 years for genetic insertion and deletion events in defined regions of the genome (33). Extensive differences between strains were detected, and similar insertion/deletion events appear to be present in the genomes of strains with very different origins,

indicating extensive recombination within the worldwide *S. mutans* population. In at least two instances, insertion of foreign DNA appears to have displaced original *S. mutans* genes. That study was limited to specific regions associated with a known phenotype that covered only about 5% of the genome; the present work was undertaken to extend the findings to the entire genome. A subset of strains was therefore chosen for DNA-DNA hybridization analysis with a microarray representing the open reading frames of *S. mutans* UA159 to further explore the extent of genomic variation. The strains were selected because they represent the spectrum of variation observed in the previous study. They were isolated in England, Sweden, Finland, Japan and the USA at widely different times so they are a representative sample of *S. mutans* worldwide. In this study we have made no attempt to correlate their genetic properties with their site of isolation or potential cariogenicity.

Materials and methods

Bacterial strains

Strains of *S. mutans* listed in Table 1 were used in previous studies to examine chromosomal deletions (20, 24, 31, 33) and chosen to represent the range of genetic variation identified by a polymerase chain reaction approach to analyse specific regions.

DNA microarrays

Microarray slides were kindly provided through the NIDCR Oral Microbe Microarray Initiative. The genome array consisted of 1948 70-mer oligonucleotides representing open reading frames from *S. mutans* UA159, with each 70-mer represented in quadruplicate. Additional details of the arrays are available at http://pfgrc.tigr.org/descriptions/S_mutans.shtml.

DNA extraction and probe labelling

Genomic DNA was prepared from stationary-phase cultures grown at 37°C in Todd–Hewitt broth supplied with 0.5% (weight/volume) yeast extract and 20 mM DL-threonine using the GenElute™ Bacterial Genomic DNA Kit (Sigma, Poole, UK). The DNA was sonicated for 2 min in a Decon Ultrasonic FS200 water bath using a random frequency sweep of 35–45 kHz, and fragmentation of DNA to 2–5 kilobases was confirmed by gel electrophoresis.

Random hexamers (9 μg) were annealed to 3 μg fragmented DNA by heating to 99°C then chilling on ice. The DNA was labelled either with 0.5 M Cy3-dCTP or Cy5-dCTP (Amersham Biosciences) using 15 U Klenow (Bioline) and 0.1 mM of dATP, dGTP, dTTP, and 0.05 mM of dCTP (Bioline) by incubation at 37°C for 2 h. Quenching of the fluorescent Cy-dCTP dyes was minimized by performing all incubations under reduced lighting conditions. The *S. mutans* UA159 reference strain DNA fragments labelled with either Cy3-dCTP or Cy5-dCTP was combined with the test *S. mutans* strain bearing the opposite fluorescent label, and the probes were purified using the QIAquick polymerase chain reaction Purification Kit (QIAGEN, Crawley, UK).

Hybridization

Hybridization was carried out following protocols provided by The Institute for Genomic Research using solutions from the Pronto!™ Universal Microarray Reagent System (Corning, Corning, NY) according to the manufacturer's instructions. Briefly, microarray slides were primed in Pre-Soak Solution (42°C for 20 min) and inserted into the microarray chamber of a Tecan HS400 Hybridization Station. The slides were then coated with Pre-Hybridization Solution by washing at 42°C for 15 s, soaking for 15 s, a second wash for 15 s, and a final wash for 30 s. The labelled reference and test strain DNAs were added to 80 μl Hybridization Solution, and the mixture was denatured at 95°C for 3 min, injected into the microarray chamber, and hybridized with the microarray slides at 42°C for 16 h with medium agitation frequency. The slides were then washed with Wash Solution 1 for 1 min and soaked for 1 min at 42°C. This procedure was repeated with Wash Solutions 2 and 3, except at a temperature of 23°C. Slides were dried and scanned immediately.

Microarray scanning and data analysis

Microarray slides were scanned at 5-μm resolution at 532 and 635 nm using a GenePix 4000B scanner (Axon, Union City, CA) with GenePix Pro4 software. Competitive hybridization was carried out twice for each strain; UA159 was labelled with Cy3-dCTP and the test strain was labelled with Cy5-dCTP, and in the second hybridization the dyes were switched (dye swaps).

GenePix results files were loaded into GeneSpring GX (Agilent, Santa Clara, CA) on top of a genome annotation created from the GAL file supplied by the NIDCR Oral Microbe Microarray Initiative (http://www.nidcr.nih.gov/Research/Extramural/NIDCR_TIGR_Facility.htm). Slides representing dye swaps were transformed for analysis (swapping signal for control in one sample) and normalized using intensity-dependent (Lowess) normalization. Dye swaps were then treated as replicates. Data from slides were filtered to remove open reading frames flagged as 'reserved' spots from the analysis, and filtered again to remove open reading frames flagged as 'absent' by GenePix Pro to provide a quality-controlled set of genes for further analysis. To provide a cut-off point for deletion detection, the change in intensity between open reading

Table 1. Strains of *Streptococcus mutans* tested in hybridization against the UA159 microarray

Strain	Isolation	Year	Serotype	*msm*	*bgl*	*cel*	*gbpA*	IS3	TnSmu2
UA159	USA	1982	c	+	+	+	+	+	+
GS-5 (HK)	USA	1966	c	+	+	+	+	+	+
AT10	Sweden	1968	c	−	−	−	+	−	−
Mt4653	Japan	1980	e	−	−	−	+	−	−
MT4863	Japan	1980	c	−	+	+	+	−	−
V1996	USA	1991	c	+	+	+	−	−	−
L13	Finland	2000	e	−	+	−	+	−	−
L18	Finland	2000	c	−	+	+	+	−	−
LML5	England	2001	c	+	−	+	-	−	−
LML7	England	2001	c	+	+	+	+	−	−

The genes and regions listed were previously examined by polymerase chain reaction (20,24,33), using primers based on flanking sequences, and are identified as present (+) or absent (−).

frames known from previous experimental work to be deleted in certain test strains was examined. For each strain, open reading frames exhibiting a similar or greater change in intensity between the reference strain and the experimental strain (indicating a deletion) were separated into a list for further analysis.

Sugar utilization

Fermentation and utilization of sugars as sole carbon source were determined as described by Colby et al. (6).

Results

Comparison of test strains with *S. mutans* UA159

The *S. mutans* UA159 genome has 1963 identified open reading frames, of which 1948 are represented on the microarray. Of the 1948 open reading frames, 385 (19.76%) failed to give a positive hybridization result with one or more of the nine test strains, indicating either that they were absent from the test strain or that their sequence differed substantially from that of UA159. These absent open reading frames were scattered throughout the genome. Seventy-two (18.7%) of the affected open reading frames were absent from only one of the nine test strains so the great majority were altered in more than one strain and in the case of 13 open reading frames, DNA from all nine of the test strains failed to hybridize with the UA159 microarray (indicated by a continuous red line in Fig. 1).

A number of regions of the chromosome of the strains used in this study have previously been examined by polymerase chain reaction, to characterize the presence of insertions and deletions in the nine test strains. The results from these previous studies, which used a different experimental approach, provide a control for the microarray data.

Figure 2 represents a 'zoom-in' to microarray results including a previously studied region of the genome. Genomic DNA from strains known to lack genes in the *msm*, *gal* and *bgl* operons and open reading frames needed for serotype c antigen production all failed to hybridize to the microarray spots representing the known blocks of affected open reading frames (Fig. 2). The genotypic changes detected by microarray correlated exactly with the observable phenotypes of sugar utilization patterns and serotype. Strains L13 and MT4653 are both serotype 'e' and lack open reading frames SMU.831–SMU.834 (27); At10, L13, L18, MT4653 and MT4863, which are all melibiose-negative and unable to grow on galactose as sole carbon source, lack open reading frames SMU.876–SMU.888 (24); At10, LML5 and MT4653 are unable to attach the β-glucoside aesculin in the presence of glucose as a result of the absence of SMU.980–SMU.982 (20). No observable phenotype has been associated with the presence or absence of SMU.891–SMU.897, predicted to encode DNA-modifying enzymes. In a few instances, individual open reading frames within the blocks of previously studied deletions gave a positive hybridization result (Table 2), suggesting either that a duplicate open reading frame exists elsewhere in the chromosome of the test strain or that there has been gene rearrangement with respect to UA159.

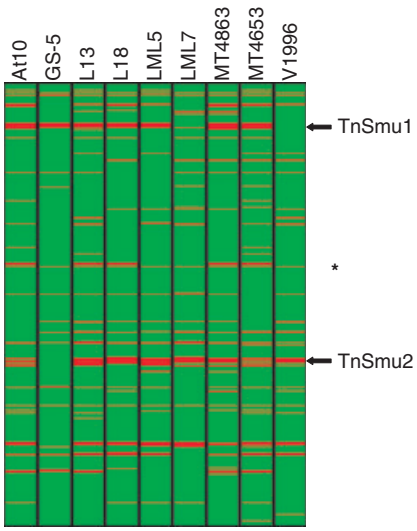

Fig. 1. Streptococcus mutans genetic diversity. Genomic DNA of nine strains of *S. mutans* was hybridized with a microarray representing the open reading frames of *S. mutans* UA159. Green sections indicate where DNA from the test strain hybridized with the UA159 target spots. Red sections indicate where hybridization was more than two-fold reduced compared to control and open reading frames were concluded to be absent or substantially altered in the test strain. The data represent from open reading frame 1 (top) to open reading frame 2165 (bottom). An asterisk indicates the region of the array presented in Fig. 2.

Block deletions and genomic islands

Forty-nine (12.7%) of the absent open reading frames affected individual open reading frames. The others affected a variable number of adjacent open reading frames, the two largest differences from UA159 indicating the absence of 32 or 33 open reading frames from the test strains. A total of 37 such blocks involving more than one open reading frame were identified and are summarized in Table 2. Of the absent open reading frames, just over half are currently only predicted hypothetical proteins with no clues as to their possible function or origin. The next most common class of open reading frames (10%) comprises those identified as intact or fragmented transposons and insertion sequences. Another 10% are associated with carbohydrate metabolism (though this study was biased towards strains known to have defects in sugar fermentation patterns). Other open reading frames commonly affected were those predicted to be concerned with nucleases and restriction enzymes, bacteriocins and ATP binding cassette transport systems.

Each strain gave a unique pattern of hybridization. A spreadsheet detailing all the absent open reading frames for each of the strains is provided as Supplementary Material (Figure S1). Three major groups of affected regions are considered further below.

TnSmu1 and TnSmu2

TnSmu1 was identified from the genome sequence (1) and corresponds to a very large region of around 23 kilobases in UA159 involving SMU.191–SMU.226 (Fig. 1). It lies beside a cluster of tRNA genes and contains multiple predicted transposases and integrases, as well as hypothetical proteins. The only strains in this study that carry TnSmu1, and so show hybridization with the UA159 array, are LML7 and V1996.

TnSmu2 is a large genomic island of about 50 kilobases of suspected foreign DNA with a G + C composition that is atypical for *S. mutans* containing open reading frames related to the antimicrobial peptide synthetases of *Bacillus*. Twenty-nine open reading frames of TnSmu2 (SMU.1329–SMU.1363) are present in UA159 and GS-5, but absent from the other strains in this study, in agreement with previous results obtained by polymerase chain reaction with primers flanking TnSmu2 (33). Although polymerase chain reaction results indicated that this block was missing in its entirety (33), several of the transposases and other open reading frames within TnSmu2 showed hybridization with one or more of the test strains, again indicating the likelihood that transposases may exist in multiple copies located in more than one part of

SMU ID	Description	At10	GS5	L13	L18	LML5	LML7	MTc	Mte	V1996	Notes
SMU.616	hypothetical protein										
SMU.651	putative ABC transporter, substrate-binding protein										
SMU.652	putative ABC transporter, ATP-binding protein										
SMU.653	putative ABC transporter, permease protein										
SMU.654	putative ABC transporter, ATP-binding protein MutF										Mutacin I/III
SMU.655	putative MutE										
SMU.657	putative MutG										
SMU.658	conserved hypothetical protein										
SMU.660	putative histidine kinase SpaK										
SMU.678	putative oxidoreductase, aldo/keto reductase family										
SMU.681	hypothetical protein										
SMU.682	hypothetical protein										
SMU.683	putative ATP-binding protein										
SMU.684	hypothetical protein										
SMU.685	hypothetical protein										
SMU.687	hypothetical protein										
SMU.735	hypothetical protein										
SMU.737	conserved hypothetical protein										
SMU.803	putative ABC transporter, ATP-binding protein										
SMU.804	hypothetical protein										
SMU.811	hypothetical protein										
SMU.829	putative glycosyltransferase										
SMU.831	conserved hypothetical protein (serotype c)										Serotype c
SMU.832	hypothetical protein (serotype c)										
SMU.833	putative glyosyltransferase (serotype c)										
SMU.834	conserved hypothetical protein (serotype c)										
SMU.875	putative transposase, IS150-like										Msm/Gal
SMU.876	regulatory protein MsmR										
SMU.877	alpha-galactosidase Aga										
SMU.878	sugar-binding protein MsmE										
SMU.879	multiple sugar-binding ABC transporter MsmF										
SMU.880	multiple sugar-binding ABC transporter MsmG										
SMU.881	sucrose phosphorylase, GtfA										
SMU.882	multiple sugar-binding ABC transporter MsmK										
SMU.883	dextran glucosidase DexB										
SMU.885	galactose operon repressor GalR										
SMU.886	galactokinase, GalK										
SMU.887	galactose-1-P-uridyl transferase, GalT										
SMU.888	UDP-galactose 4-epimerase, GalE										
SMU.891	type I restriction-modification system DNA methylase										
SMU.892	putative type I restriction-modification system										
SMU.893	putative antiodon nuclease										
SMU.895	possible DNA-damage-inducible protein										
SMU.896	conserved hypothetical protein										
SMU.897	putative type I restriction-modification system										
SMU.980	PTS system, beta-glucoside-specific EII component										Bgl
SMU.981	BglB fragment										
SMU.982	BglB fragment										

Strain abbreviations: MTc = MT4863, Mte = MT4653

	Positive hybridization with UA159 microarray spot
	No hybridization with UA159 microarray spot
	Thick lines separate blocks of adjacent deletions
	Deletions experimentally demonstrated by PCR
	Putative nucleases, restriction enzymes
	Putative transposons, IS elements
	Putative bacteriocins, competence-assoc. ORFs
	Putative ABC systems
	Putative carbohydrate metabolism enzymes

Fig. 2. Representative region of microarray data showing open reading frames that are present in the UA159 genome but absent from one or more of the test strains. open reading frames marked green indicate where DNA from the test strain hybridized with the UA159 microarray; open reading frames marked red indicate that hybridization was not detected. Open reading frames present in all strains are omitted and blocks of absent open reading frames are separated by thick black lines. Data for the complete genome are provided in Figure 3 as Supplementary Material.

the genome and may mediate rearrangements.

Sugar uptake systems

Bioinformatic analysis predicted the occurrence of a genomic island comprised of open reading frames SMU.100–SMU.105 including genes for a sorbose–phosphotransferase uptake system. These open reading frames are all absent from strains At10, L13, L18, LML5, MT4653, MT4863 and V1996 but are present in GS5 and LML7, as well as UA159. None of the strains, whether they carried the sorbose island or not, were able to produce acid from sorbose or grow on semi-defined medium with sorbose as sole carbon source. These results suggest that the genes are not functioning in *S. mutans*. In At10 and MT4863, the deletion extends as far as SMU.116, so that one of the fructose-specific phosphotransferase systems is also lost. In MT4863, SMU.1878 and SMU.1879 (which are predicted to encode components of a mannose-specific phosphotransferase system) are absent, adding to the list of deletions that may modulate the ability to utilize carbohydrates. No differences in the sugar fermentation patterns of At10 or MT4863 could be detected associated with these deletions.

Table 2. Blocks of open reading frames absent in one or more strain of *Streptococcus mutans*

Block	Predicted function	GI[1]	Ref[2]
SMU.42–SMU.47	DNA-modifying enzymes and hypothetical proteins	3	
SMU.52–SMU.58	Hypothetical proteins		
SMU92–SMU.94	Transposase		
SMU.100–SMU.116	Sorbose PTS, transposase, hypothetical proteins, (fructose PTS)	4	
SMU.135–SMU.143	Hypothetical regulators		
SMU.149–SMU.153	Transposase, hypothetical proteins		
SMU.175–SMU.176	Hypothetical proteins		
SMU.191–SMU.226	TnSmu1, ICESt, transposases, hypothetical proteins	6	
SMU.261–SMU.265	Transposases, hypothetical proteins	7	
SMU.276–SMU.277	Hypothetical proteins	7	
SMU.281–SMU.284	Hypothetical proteins	7	
SMU.344–SMU.348	Hypothetical proteins		
SMU.372–SMU.381	Hypothetical proteins		
SMU.504–SMU.505	DNA-modifying enzymes		
SMU.604–SMU.606	Hypothetical proteins	9	
SMU.651–SMU.658	ABC transporter, Mutacin I		(33)
SMU.681–SMU.687	Hypothetical proteins		
SMU.735–SMU.737	Hypothetical proteins		
SMU.803–SMU.804	ABC transporter, hypothetical protein		
SMU.831–SMU.834	Serotype c polysaccharide enzymes		(27)
SMU.876–SMU.888	*msm* and *gal* operons		
SMU.891–SMU.896	DNA-modifying enzymes		
SMU.980–SMU.892	*bgl* operon		(24)
SMU.1026–SMU.1032	Transposases, hypothetical proteins	16	
SMU.1159–SMU.1161	Hypothetical proteins		
SMU.1205–SMU.1209	Hypothetical proteins		
SMU.1254–SMU.1259	DNA-modifying enzymes, hypothetical proteins		
SMU.1262–SMU.1264	Hypothetical protein, histidine metabolism	11	
SMU.1329–SMU.1363	TnSmu2, transposases, bacitracin synthetases etc.	12	(33)
SMU.1484–SMU.1485	DNA-modifying enzymes, hypothetical protein		
SMU.1574–SMU.1577	Hypothetical proteins		
SMU.1584–SMU.1585	Putative cross-reactive antigen, regulator		
SMU.1596–SMU.1605	*cel* operon		(20)
SMU.1750–SMU.1776	Hypothetical proteins		
SMU.1804–SMU.1818	Transposases, *scn* bacteriocin		
SMU.1896–SMU.2054	Transposases, *blp* bacteriocin	15	
SMU.2069–SMU.2070	Hypothetical proteins		

[1]Genomic Islands predicted by the Oral Pathogen Sequence Database http://www.oralgen.lanl.gov/. References are to studies describing sequencing of the affected regions.
[2]ABC, ATP binding cassette; PTS, phosphotransferase.

Bacteriocin-related open reading frames

We previously reported extensive heterogeneity in the region of the genome that encodes production of Mutacin 1 (33). This is confirmed by the microarray data, which revealed that only At10, L18 and MT4863 had the same complement of open reading frames SMU.651–SMU.658 as UA159, the other strains lacked one or more of the open reading frames in this block (Fig. 2). It should be noted that the experiments performed here with the UA159 microarray could not determine whether additional open reading frames were present in any of the strains. Two other blocks are associated with possible production of bacteriocins. SMU.1803–SMU.1818 includes fragments of transposases as well as five open reading frames that are homologous to members of the streptococcal lantibiotic *scn* regulon, which contains nine open reading frames and is found in some strains of *Streptococcus pyogenes* (19). Only LML7 carries the same set of open reading frames found in UA159, the other strains all lacking eight or more of the open reading frames and making it highly improbable that this particular lantibiotic will be produced in *S. mutans*. This may also apply to the block SMU.1888–SMU.1917, which contains a number of open reading frames with a Gly-Gly motif typical of many peptide secretion systems involved in competence, quorum sensing and bacteriocin production but shows extensive variability between the nine strains.

Discussion

We have previously described chromosomal deletions from strains of *S. mutans*, first by DNA–DNA hybridization (31) and later by an approach using polymerase chain reaction primers based on flanking regions to screen for the presence of particular regions of the genome (20, 24, 33). These studies revealed extensive heterogeneity within the species, and supported the idea that there exists a core genome and a flexible genome comprised of dispensable genes and foreign DNA acquired by horizontal transfer. However, those studies were limited to genes associated with known or predicted phenotypes whereas the microarray approach used here examines the genome globally; examining the nine test strains for their possession of DNA capable of hybridizing with individual oligonucleotides representing over 99% of the identified open reading frames in *S. mutans* UA159. The interpretation of the data does not assume that gene order in all strains is the same as that in UA159, though we have argued elsewhere that the core genome in *S. mutans* appears to be well conserved (33).

If the assumption is made that those open reading frames for which the test strains failed to give a positive hybridization result are absent or so altered in sequence as to be non-functional in those strains, it can be deduced that nearly one-fifth of the open reading frames in *S. mutans* UA159 are non-essential for survival. This figure is very close to the size of the 'dispensable genome' in other species as diverse as *Staphylococcus aureus*, *Helicobacter pylori* and *Bordetella pertussis* and other streptococci including *Streptococcus agalactiae* (3) and *Streptococcus pneumoniae* (10). The open reading frames that show variable occurrence include both those that are considered to be normal components of the *S. mutans* core genome, such as *msm*, *gal*, *bgl*, *cel* and *gbpA* genes, that are present in UA159 but are absent from the genomes of other strains (20, 24, 33) and open reading frames that are thought to represent foreign DNA acquired by horizontal transfer from other organisms (33).

Ajdic et al. (1) identified two large transposon-like regions in the genome of UA159 that they proposed might be mobile genetic elements and named them TnSmu1 and TnSmu2. Five of the open reading frames in TnSmu1 correspond to components of the ICE*St1* integrative and conjugative element of *Streptococcus thermophilus* (5). Related elements have been detected in a wide range of bacterial species and have been proposed to be important agents of change within species and of inter-species DNA transfer (4). In many other species, an insertion adjacent to tRNA genes is observed, as in *S. mutans*

UA159. In contrast to TnSmu1, which shows the greatest homology to other streptococci, the DNA composition and sequence homology suggests that TnSmu2 may have originated in a *Bacillus* species (33). TnSmu1 and TnSmu2 were also predicted to consist of foreign DNA in a bioinformatic search of the *S. mutans* UA159 genome for genomic islands carried out for the Oral Pathogen Sequence Database (http://www.oralgen.lanl.gov/). Genomic Islands were predicted on the basis of features including anomalous G + C composition, BLAST analysis, codon usage and direction of transcription. Eleven genomic islands were predicted, nine of which have been confirmed as missing from one or more of the strains studied here (Table 2). In the bioinformatic analysis, Genomic Island 7 (SMU.267–SMU.281) was not predicted with a high degree of confidence and the microarray data showed that only single open reading frames within it are absent in two strains, suggesting that it is probably not an island. In particular, the predicted ribulose–phosphotransferase components in Genomic Island 7 are present in all nine strains examined here and may represent true *S. mutans* genes. It is possible that examination of a wider range of strains would confirm whether the remaining two predicted genomic island regions do indeed occur, or would uncover new ones.

The first genomic deletions identified in *S. mutans* were those that affect the metabolism of melibiose and β-glucosides (20, 24) and it is striking that several variable blocks are associated with carbohydrate metabolism. The bioinformatic analysis of the UA159 sequence indicated that SMU.100–SMU.105, absent from seven of the nine strains, encoded an uptake system that might be specific for sorbose or mannose and that shows properties that indicate it has been acquired as foreign DNA from another species of *Streptococcus* or *Lactobacillus*. However, no phenotypic difference could be detected, indicating that the genes are non-functional. In some strains, a fructose–phosphotransferase is also missing, although this is likely to be compensated for by other fructose uptake systems (29), indicating the complexity of deducing phenotypic properties simply from genome sequence data.

It is well-established that strains of *S. mutans* show a great deal of variability in their production of, and sensitivity to, bacteriocins. A region comparable to SMU.1888–SMU.1917 is found in some strains of *S. pneumoniae* and the genes have been designated bacteriocin-like peptide (*blp*) or pneumococcin (*pnc*) (10, 23). These genes are involved in the complex interplay between bacteriocin production and competence development in *S. pneumoniae* (7, 22) and also in *S. mutans* (12, 16, 21, 32). However, the extensive variation between strains observed within the block indicates that there will be heterogeneity in phenotype. It is interesting to note that this region seems particularly prone to horizontal gene transfer, as components are also found in *S. thermophilus* (14). Another group of open reading frames associated with bacteriocin production is SMU.1803–SMU.1818, which represents an incomplete set of genes for the *scn* lantibiotic regulon described in *S. pyogenes* (19) and is probably not functional. It should be noted that both *blp* and *scn* regions are associated with open reading frames that are components of transposons, an indication that they may originally have been foreign DNA. In contrast to the *scn* and *blp* regions that appear to consist of DNA acquired by horizontal transfer from other streptococci, SMU.286 and SMU.287 (which have recently been identified as encoding a non-lantibiotic mutacin) designated as *nlmT* and *nlmE* are conserved in all nine strains examined so may represent true *S. mutans* genes (11).

The open reading frames associated with *blp (pnc)* and *scn* bacteriocins, as well as ICE*St1*, seem likely to have a streptococcal origin. In contrast, TnSmu2 is most likely to have originated in a *Bacillus* species while the genome contains several copies of ATP binding cassette transporter components for which the closest relative identified by BLAST searches of available databases is *Archaeoglobus fulgidus*. It thus appears that *S. mutans* has acquired foreign DNA from a wide range of sources, including other species of *Streptococcus*, *Bacillus* and the Archeae. It is somewhat surprising that none of the putatively foreign DNA has a recognizable source in other members of the oral microflora, though information on this aspect may well change as more genome sequences enter the public domain. On the basis of available evidence, it appears that much of the foreign DNA is non-functional in *S. mutans*, though it should be noted that half of the identified open reading frames as yet have no proposed function. Evidence as to whether or not any of these open reading frames are expressed may come from transcriptomics and proteomics. For example, Wen et al. (34) found that a number of the open reading frames identified as being absent from some strains showed altered transcription in a *brpA*-deficient mutant of UA159. While this provides evidence that the genes can be active, caution needs to be applied in extrapolating from UA159 to all strains of *S. mutans*. Further delineation of the core genome will help in identifying properties characteristic of the whole species. Indeed, we may still be some way from understanding what genes comprise the core genome, because not all are represented in UA159 (33). Detailed examination of the regions affected by blocks of deletions may be informative in this regard because foreign DNA, particularly transposons and insertion sequences, may displace the authentic *S. mutans* genes (33). One of the displaced genes in about 10% of *S. mutans* strains is *gbpA*, which encodes a glucan-binding protein. GbpA has been proposed as a virulence determinant (18) and it is clearly very important that we understand the extent of such variation, particularly with regard to virulence determinants and potential targets of vaccines or other preventive strategies.

The microarray used in this study is constructed of 70-mer oligonucleotides designed to be unique for each open reading frame. Failure to hybridize has been taken as an indication of the absence of that open reading frame in the test strain, though it might also be the result of extensive sequence divergence that would most probably also indicate altered function. It has been estimated that probe-target identity below c.85% would fail to give a signal (13). The process of designing unique microarray oligomers for each open reading frame means that the hybridizing region lies at a different position in each open reading frame so it cannot immediately be told whether an entire open reading frame is absent, nor can it determine a definitive end-point for a deletion. The same problem arises with arrays that are based on entire open reading frames because hybridization does not have to involve the full length of the coding sequence to yield a signal. An open reading frame might have lost a substantial part of its sequence yet retain its hybridizing region and we have identified several examples of this occurring. Identifying the termini of a deleted region requires sequencing and this can be informative about a possible mechanism and the consequences of deletion. Of the regions studied, the *msm*/*gal* deletion appears to be associated with insertion sequence movement, whereas it has been proposed

that the *bgl* deletion is the result of an intragenomic recombination of homologous regions (24). No explanation could be provided for the *cel* deletion (20). Changes affecting TnSmu2 and *gbpA* represent two instances where blocks present in UA159 are substituted with novel open reading frames that can be considered as part of the entire *S. mutans* gene pool (33). Detailed examination of the sequence around the blocks identified here may well uncover further examples, as well as giving more insight into the mechanisms causing genetic change.

When was the foreign DNA acquired and how has it become disseminated throughout the worldwide pool of *S. mutans* strains? Together with previous results on the occurrence of deletions, the microarray results indicate that *S. mutans* has a core genome and a dispensable genome and that genomic changes as a result of certain insertion/deletion events have become widely distributed, possibly through horizontal transfer between *S. mutans* strains. Analysis of the pattern obtained with a collection of strains has proved to be a powerful tool with which to explore population dynamics in a number of species, and is at least as informative as other approaches such as multi-locus sequence typing (3, 35). The two approaches complement each other but do not necessarily correlate (28). The number of strains of *S. mutans* examined so far is too small and inappropriately structured for useful phylogenetic analysis but the results have considerable implications for our understanding of the properties of *S. mutans* and the evolution of the species. Since there appears to be free and widespread exchange of dispensable genes throughout the species, it seems highly probable that the core genome (including housekeeping genes used for multi-locus sequence typing studies) will also have undergone such exchange. This would suggest that *S. mutans* has a panmictic population structure, in which recombination has played a large part in generating diversity (8).

This study is based entirely upon the UA159 genome, which provides a powerful starting point for comparative studies. Because genomic DNA from the nine strains tested hybridized with 80% of the open reading frames in the UA159 microarray, this strain would appear to be a good representative of the species. However, 12 open reading frame spots hybridized with only UA159, indicating that they are unique to this strain. None of them as yet has predicted functions. It is probable that other strains of *S. mutans* will carry other, novel open reading frames and genomic islands that will only be uncovered by further genome sequencing or subtractive hybridization experiments (9, 26). Indeed, it is interesting that both this study and one using subtractive hybridization identified a predicted malolactic enzyme (SMU.137) as a variable open reading frame (26). A combination of these different approaches will build up a complete picture of the complexity of the *S. mutans* gene pool, and the extent of variation between strains. An awareness of this variation will also be essential in future studies exploring the potential contribution of different genomic regions to the growth, survival and cariogenicity of *S. mutans.*

Acknowledgments

We are grateful to Alan Ward and his colleagues for their assistance with the hybridization experiments and the Pathogen Functional Genomics Resource Centre, Division of Microbiology and Infectious Diseases, NIDCR/NIAID, NIH for provision of the microarray slides.

References

1. Ajdic D, McShan WM, McLaughlin RE et al. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc Natl Acad Sci U S A 2002: **99**: 14434–14439.
2. Binnewies TT, Motro Y, Hallin PF et al. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. Funct Integr Genomics 2006: **6**: 165–185.
3. Brochet M, Couve E, Zouine M et al. Genomic diversity and evolution within the species *Streptococcus agalactiae*. Microbes Infect 2006: **8**: 1227–1243.
4. Burrus V, Pavlovic G, Decaris B, Guedon G. Conjugative transposons: the tip of the iceberg. Mol Microbiol 2002: **46**: 601–610.
5. Burrus V, Pavlovic G, Decaris B, Guedon G. The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. Plasmid 2002: **48**: 77–97.
6. Colby SM, Harrington DJ, Russell RRB. Identification and genetic characterisation of melibiose-negative isolates of *Streptococcus mutans*. Caries Res 1995: **29**: 407–412.
7. de Saizieu A, Gardes C, Flint N et al. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. J Bacteriol 2000: **182**: 4696–4703.
8. Feil EJ, Holmes EC, Bessen DE et al. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc Natl Acad Sci U S A 2001: **98**: 182–187.
9. Guo LH, Shi JN, Zhang Y, Liu XD, Duan J, Wei S. Identification of genetic differences between two clinical isolates of *Streptococcus mutans* by subtractive hybridisation. Oral Microbiol Immunol 2006: **21**: 372–380.
10. Hakenbeck R, Balmelle N, Weber B, Gardes C, Keck W, de Saizieu A. Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. Infect Immun 2001: **69**: 2477–2486.
11. Hale JD, Heng NC, Jack RW, Tagg JR. Identification of *nlmTE*, the locus encoding the ABC transport system required for export of nonlantibiotic mutacins in *Streptococcus mutans*. J Bacteriol 2005: **187**: 5036–5039.
12. Hale JD, Ting YT, Jack RW, Tagg JR, Heng NC. Bacteriocin (mutacin) production by *Streptococcus mutans* genome sequence reference strain UA159: elucidation of the antimicrobial repertoire by genetic dissection. Appl Environ Microbiol 2005: **71**: 7613–7617.
13. He Z, Wu L, Li X, Fields MW, Zhou J. Empirical establishment of oligonucleotide probe design criteria. Appl Environ Microbiol 2005: **71**: 3753–3760.
14. Hols P, Hancy F, Fontaine L et al. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. FEMS Microbiol Rev 2005: **29**: 435–463.
15. Lee SF, McGavin MK. Identification of a point mutation resulting in loss of cell wall anchoring activity of SrtA of *Streptococcus mutans* NG5. Infect Immun 2004: **72**: 4314–4317.
16. Li YH, Tang N, Aspiras MB et al. A quorum-sensing signaling system essential for genetic competence in *Streptococcus mutans* is involved in biofilm formation. J Bacteriol 2002: **184**: 2699–2708.
17. Macrina FL, Jones KR, Alpert CA, Chassy BM, Michalek SM. Repeated DNA-sequence involved in mutations affecting transport of sucrose into *Streptococcus mutans* V403 via the phosphoenolpyruvate phosphotransferase system. Infect Immun 1991: **59**: 1535–1543.
18. Matsumura M, Izumi T, Matsumoto M, Tsuji M, Fujiwara T, Ooshima T. The role of glucan-binding proteins in the cariogenicity of *Streptococcus mutans*. Microbiol Immunol 2003: **47**: 213–215.
19. McLaughlin RE, Ferretti JJ, Hynes WL. Nucleotide sequence of the streptococcin A-FF22 lantibiotic regulon: model for production of the lantibiotic SA-FF22 by strains of *Streptococcus pyogenes*. FEMS Microbiol Lett 1999: **175**: 171–177.
20. Old LA, Lowes S, Russell RR. Genomic variation in *Streptococcus mutans*: deletions affecting the multiple pathways of β-glucoside metabolism. Oral Microbiol Immunol 2006: **21**: 21–27.
21. Petersen FC, Tao L, Scheie AA. DNA binding-uptake system: a link between cell-to-cell communication and biofilm

formation. J Bacteriol 2005: **187**: 4392–4400.
22. Peterson SN, Sung CK, Cline R et al. Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. Mol Microbiol 2004: **51**: 1051–1070.
23. Reichmann P, Hakenbeck R. Allelic variation in a peptide-inducible two-component system of *Streptococcus pneumoniae*. FEMS Microbiol Lett 2000: **190**: 231–236.
24. Robinson WG, Old LA, Shah DSH, Russell RRB. Chromosomal insertions and deletions in *Streptococcus mutans*. Caries Res 2003: **37**: 148–156.
25. Sato Y, Okamoto K, Kizaki H. *gbpC* and *pac* gene mutations detected in *Streptococcus mutans* strain GS-5. Oral Microbiol Immunol 2002: **17**: 263–266.
26. Saxena D, Li Y, Caufield PW. Identification of unique bacterial gene segments from *Streptococcus mutans* with potential relevance to dental caries by subtraction DNA hybridization. J Clin Microbiol 2005: **43**: 3508–3511.
27. Shibata Y, Ozaki K, Seki M et al. Analysis of loci required for determination of serotype antigenicity in *Streptococcus mutans* and its clinical utilization. J Clin Microbiol 2003: **41**: 4107–4112.
28. Silva NA, McCluskey J, Jefferies JM et al. Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. Infect Immun 2006: **74**: 3513–3518.
29. Tanzer JM, Thompson A, Wen ZT, Burne RA. *Streptococcus mutans*: fructose transport, xylitol resistance, and virulence. J Dent Res 2006: **85**: 369–373.
30. Ueda S, Kuramitsu HK. Molecular basis for the spontaneous generation of colonization-defective mutants of *Streptococcus mutans*. Mol Microbiol 1988: **2**: 135–140.
31. Ushiro I, Lumb SM, Aduse-Opoku J, Ferretti JJ, Russell RRB. Chromosomal deletions in melibiose-negative isolates of *Streptococcus mutans*. J Dent Res 1991: **70**: 1422–1426.
32. van der Ploeg JR. Regulation of bacteriocin production in *Streptococcus mutans* by the quorum-sensing system required for development of genetic competence. J Bacteriol 2005: **187**: 3980–3989.
33. Waterhouse JC, Russell RRB. Dispensable genes and foreign DNA in *Streptococcus mutans*. Microbiology 2006: **152**: 1777–1788.
34. Wen ZT, Baker HV, Burne RA. Influence of BrpA on critical virulence attributes of *Streptococcus mutans*. J Bacteriol 2006: **188**: 2983–2992.
35. Zhou D, Han Y, Song Y et al. DNA microarray analysis of genome dynamics in *Yersinia pestis*: insights into bacterial genome microevolution and niche adaptation. J Bacteriol 2004: **186**: 5138–5146.

Supplementary Material

The following supplementary material is available for this article:

Figure S1. Microarray data showing open reading frames that are present in the UA159 genome but that are absent from one or more of the test strains. Open reading frames marked green indicate where DNA from the test strain hybridized with the UA159 microarray; open reading frames marked red indicate that hybridization was not detected. Open reading frames present in all strains are omitted and blocks of absent open reading frames are separated by thick black lines Other colour coding is explained at the bottom of the table. (Excel format)

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1399-302X.2006.00330.x (this link will take you to the article abstract)