molecular oral microbiology

molecular oral microbiology

Clonal structure of *Streptococcus sanguinis* strains isolated from endocarditis cases and the oral cavity

T. Do[1], S.C. Gilbert[1], J. Klein[2], S. Warren[2], W.G. Wade[1] and D. Beighton[1]

1 Department of Microbiology, King's College London, London, UK
2 Directorate of Infection, Guy's and St Thomas' NHS Trust, London, UK

Correspondence: David Beighton, The Henry Wellcome Laboratories for Microbiology and Salivary Research, KCL Dental Institute Floor 17, Guys Tower, London Bridge, SE1 9RT, England. Tel: +44-(0)-2071887465; fax: +44-(0)-2071887466; E-mail: david.beighton@kcl.ac.uk

Keywords: CLONALFRAME; endocarditis; multi-locus sequence typing; *Streptococcus sanguinis*
Accepted 3 June 2011
DOI: 10.1111/j.2041-1014.2011.00618.x

SUMMARY

A collection of *Streptococcus sanguinis* strains from patients with endocarditis (*n* = 21) and from the oral cavity (*n* = 34) was subjected to a multi-locus sequence typing analysis using seven housekeeping genes, carbamoyl-phosphate synthetase (*carB*), Co/Zn/Cd efflux system component (*czcD*), D-alanyl-D-alanine ligase (*ddl*), DNA polymerase III (*dnaX*), glucose-6-phosphate dehydrogenase (*gdh*), DNA-directed RNA polymerase, beta subunit (*rpoB*) and superoxide dismutase (*sodA*). The scheme was expanded by the inclusion of two the putative virulence genes, bacitracin-resistance protein (*bacA*) and saliva-binding protein (*ssaB*), to increase strain discrimination. Extensive intra-species recombination was apparent in all genes but inter-species recombination was also apparent with strains apparently harbouring *gdh* and *ddl* from unidentified sources and one isolate harboured a *sodA* allele apparently derived from *Streptococcus oralis*. The recombination/mutation ratio for the concatenated housekeeping gene sequences was 1.67 (95% confidence limits 1.25–2.72) and for the two virulence genes the r/m ratio was 3.99 (95% confidence limits 1.61–8.72); recombination was the major driver for genetic variation. All isolates were distinct and the endocarditis strains did not form distinct sub-clusters when the data were analysed using CLONALFRAME. These data support the widely held opinion that infecting *S. sanguinis* strains are opportunistic human pathogens.

INTRODUCTION

Streptococcus sanguinis, formerly *Streptococcus sanguis*, was originally proposed to describe a group of 36 streptococcal strains isolated from subacute bacterial endocarditis (White & Niven, 1946). In addition to its role in cardiovascular disease, *S. sanguinis* is a member of the normal oral microbiota (Aas *et al.*, 2005; Dewhirst *et al.*, 2010). Although *S. sanguinis* is apparently ubiquitous in the oral biofilm the identification of the species using conventional biochemical tests is difficult. Kilian *et al.* (1989) reported the use of a battery of tests to identify oral streptococci and divided strains identified as *S. sanguinis* into four biovars and in a later study a similar battery of tests was used to divide *S. sanguinis* strains into three biotypes (Beighton *et al.*, 1991). The reliability of these subdivisions within the species may be questioned because they were not confirmed by DNA-based analyses. Commercial test kits used to identify viridans streptococci are reported to be unreliable for

the differentiation of *S. sanguinis* and all other oral streptococcal species (Hoshino *et al.*, 2005).

Given the pathogenic role of some oral streptococci, it is important to be able to correctly identify members of this group. 16S ribosomal RNA (rRNA) gene sequence analysis is of value for the identification of some species, but *Streptococcus oralis* and *Streptococcus mitis* cannot be reliably discriminated using this technique, nor can they be distinguished from *Streptococcus pneumoniae* or *Streptococcus pseudopneumoniae* (Arbique *et al.*, 2004). Therefore alternative approaches have been suggested based on single or multiple gene sequence analysis. Individual gene sequences used to identify species of viridans streptococci include the genes encoding D-alanine:D-alanine ligase (*ddl*) (Garnier *et al.*, 1997; Kawamura *et al.*, 1999), glutamate dehydrogenase (*gdh*) (Okwumabua *et al.*, 2003), the β subunit of RNA polymerase (*rpoB*) (Drancourt *et al.*, 2004), manganese-dependent superoxide dismutase (*sodA*) (Poyart *et al.*, 1998, 2002; Kawamura *et al.*, 1999), the spacer region between the 16S and 23S rRNA genes and the *groESL* genes (Teng *et al.*, 2002; Chen *et al.*, 2005). Hoshino *et al.* (2005) used concatenated partial sequence (1630 nucleotides) derived from four streptococcal genes (*ddl*, *gdh*, *rpoB* and *sodA*) and demonstrated that 115 clinical oral streptococcal isolates could all be assigned to clusters that contained type strains, indicating the usefulness of this approach to identify oral streptococci. The *S. sanguinis* strains ($n = 11$, including type, reference and clinical isolates) formed a distinct cluster but with apparent diversity within the cluster. A more comprehensive approach to the identification of primarily viridans streptococci used concatenated partial sequences (3630 nucleotides) of seven housekeeping genes (*map*, *pfl*, *ppa*, *pyk*, *rpoB*, *sodA* and *tuf*) (Bishop *et al.*, 2009). A total of 420 viridans streptococcal isolates were investigated and sequence analysis determined that the majority of isolates clustered with type strains, again supporting a multiple gene sequencing approach to the identification of oral streptococci. There was no evidence that *S. sanguinis* isolates from infections were distinct from those isolated from the healthy oral cavities.

The multi-locus sequencing approach employed in these previous studies (Hoshino *et al.*, 2005; Bishop *et al.*, 2009) to identify viridans streptococci was originally devised as multi-locus sequence typing (MLST) for the identification of virulent strains within a single species (Maiden *et al.*, 1998). Internal fragments of individual housekeeping genes are sequenced in both directions and the different allele sequences are allocated an allele number so that a sequence type (ST) is a unique set of seven allele numbers. These data can be used to identify epidemic clones and also to investigate the phylogeny of species. There has been no previous report of the use of an MLST approach to investigate the diversity of *S. sanguinis* isolates and to determine if a specific ST, or STs, identify clones associated with endocarditis. In this study we have studied a collection of *S. sanguinis* strains including some isolated from patients with endocarditis and others recovered from supra-gingival and sub-gingival dental plaque of medically fit subjects. We have included seven housekeeping genes in the analysis as well as two putative virulence genes, *bacA* and *ssaB* to potentially increase the discrimination of the sequence analysis (Edwards *et al.*, 2008). The *bacA* gene product is undecaprenol kinase, a membrane-bound isoprenol kinase that phosphorylates cytoplasmic undecaprenol monophosphate, overcoming the inhibition caused by bacitracin. In an animal model of endocarditis the disruption of this gene resulted in a 55-fold reduction in virulence. *ssaB* encodes a cell-wall-bound lipoprotein and is an orthologue of *fimA* and *sloC*, genes required for endocarditis virulence in *Streptococcus mutans* and *Streptococcus parasanguis* (Burnette-Curley *et al.*, 1995; Paik *et al.*, 2003). In *S. sanguinis ssaB* codes for an adhesion-mediating inter-species binding factor with saliva-binding activity (Ganeshkumar *et al.*, 1991) and its inactivation of dramatically reduced endocarditis virulence (Das *et al.*, 2009).

METHODS

Bacterial strains

The isolates used in this study are listed in Table 1. The isolates were cultivated anaerobically on Fastidious Anaerobe Agar supplemented with 5% (volume/volume) horse blood (FAA; LabM, Bury, UK) for 24 h, anaerobically, at 37°C and were identified on the basis of partial 16S rRNA sequences analysed searching the Ribosomal Database Project (Cole *et al.*, 2009). The collection included 21 blood culture isolates, including the type strain of the species

Table 1 The collection of clinical and oral *Streptococcus sanguinis* isolates

		Alleles[2]									
Isolate	ST[1]	*carB*[2]	*czcD*	*ddl*	*dnaX*	*gdh*	*rpoB*	*sodA*	*bacA*	*ssaB*	Origin
SK36	1	1	1	1	1	1	1	1	1	1	Oral
S53H	2	1	1	1	24	1	1	6	22	17	Oral
SW48	3	1	1	1	1	1	1	6	16	17	Endocarditis
SW77	4	1	1	1	33	1	35	16	16	17	Endocarditis
SW78	5	1	1	27	34	1	1	16	30	17	Endocarditis
W1990	6	1	6	3	18	5	5	14	34	4	Oral
W589	7	1	6	1	1	1	1	28	1	17	Oral
AC6966	8	2	7	6	2	6	6	3	3	3	Endocarditis
AC7123	9	3	8	7	3	7	7	4	4	4	Endocarditis
SW59	10	3	36	13	31	26	34	23	28	15	Endocarditis
AC7134	11	4	9	8	4	8	8	5	5	5	Endocarditis
S65E	12	4	28	1	25	21	26	5	23	4	Oral
AC7156	13	5	10	9	5	1	9	6	6	6	Endocarditis
SW37	14	5	34	1	10	1	32	13	26	25	Endocarditis
F10D1	15	6	11	10	6	3	11	7	7	7	Oral
F16C2	16	7	12	3	7	1	12	8	2	8	Oral
G122E	17	7	17	3	13	1	5	13	2	4	Oral
G123D	18	7	18	3	14	14	16	14	11	12	Oral
SW10	19	7	17	3	20	1	30	6	2	25	Endocarditis
NCTC7863	20	7	38	18	14	5	21	14	2	2	Endocarditis
W11-2A1	21	7	39	28	31	29	36	25	29	28	Oral
F6C7	22	8	13	11	9	10	13	9	9	10	Oral
F7A14	23	8	14	12	10	11	13	10	9	10	Oral
F9C2	24	8	15	12	11	12	14	11	9	10	Oral
G121E	25	9	16	13	12	13	15	12	10	11	Oral
G124B	26	10	19	3	15	5	16	5	12	12	Oral
S69G	27	10	43	31	26	34	39	5	24	20	Oral
G41E	28	11	20	14	16	15	17	12	13	14	Oral
G42E	29	12	21	13	17	13	18	2	14	15	Oral
G52C	30	13	22	15	18	16	19	15	15	16	Oral
G62A	31	14	1	16	1	17	1	16	16	17	Oral
S43J	32	14	44	1	1	33	1	6	19	1	Oral
G69G	33	15	23	17	19	18	20	8	17	12	Oral
S36C	34	16	24	19	20	1	22	17	2	12	Oral
S40C	35	17	25	20	21	14	23	5	18	1	Oral
S41F	36	18	26	21	22	19	24	18	1	18	Oral
S4E	37	19	27	2	23	20	25	14	20	19	Oral
S51C	38	20	17	22	20	1	5	17	21	4	Oral
SH133.79	39	21	29	23	27	10	27	19	1	21	Endocarditis
SH36190	40	22	30	3	17	22	28	6	24	22	Endocarditis
SHB65.12	41	22	32	24	17	22	28	6	25	22	Endocarditis
SHB47.7	42	23	31	3	14	5	5	20	2	23	Endocarditis
SHBS.7	43	24	33	12	28	23	29	8	9	24	Endocarditis
SW12	44	25	1	1	29	22	1	21	16	17	Endocarditis
SW15	45	26	17	25	2	24	31	8	2	12	Endocarditis
SW45	46	27	35	26	30	25	33	22	27	12	Endocarditis
SW61	47	27	35	26	32	27	2	24	29	26	Endocarditis
W1233	48	27	2	2	36	2	2	24	27	29	Oral
SW62	49	28	37	2	27	28	5	14	2	27	Endocarditis
W17BC1	50	29	5	5	35	9	10	26	31	12	Oral
W8-1B1	51	30	40	29	20	30	30	2	21	25	Oral
W1389	52	31	3	3	37	3	3	27	32	30	Oral

Table 1 (*Continued*)

Isolate	ST[1]	Alleles[2]									Origin
		carB[2]	*czcD*	*ddl*	*dnaX*	*gdh*	*rpoB*	*sodA*	*bacA*	*ssaB*	
W1649	53	32	4	4	38	4	4	2	33	31	Oral
F5A7	54	33	41	30	8	31	38	29	8	9	Oral
G2F	55	33	42	30	8	32	37	6	8	13	Oral

Strains labelled SW were isolated from patients with endocarditis treated at Guy's and St Thomas' NHS Trust (London, UK), those labelled SH were provided by Professor Ian Douglas (Sheffield, UK) and those labelled AC were provided by Dr Mark van der Linden (Aachen, Germany). Strain SK36 is the sequenced *S. sanguinis* strain and NCTC 7863 is the type strain of the species. Other strains were isolated by the authors from the oral biofilm of healthy subjects.

[1]ST is sequence type; the numbers in this column indicate the ST number; these numbers are used in Fig. 2.

[2]The numbers in the columns are the allele numbers for each isolate.

(NCTC 7863), from patients with endocarditis and 34 isolates recovered from either supra-gingival or sub-gingival plaque samples taken from medically fit subjects. In addition, the sequences of the select genes were obtained from the genomic sequence of *S. sanguinis* strain, SK36, an oral isolate, and included in the analysis (Xu *et al.*, 2007).

Choice of loci and primer design

The housekeeping genes used to differentiate between strains were *carB* (carbamoyl-phosphate synthetase), *czcD* (Co/Zn/Cd efflux system component), *ddl* (D-alanyl-D-alanine ligase), *dnaX* (DNA polymerase III), *gdh* (glucose-6-phosphate dehydrogenase), *rpoB* (DNA-directed RNA polymerase, beta subunit) and *sodA* (superoxide dismutase) and the putative virulence genes investigated were *bacA* (bacitracin-resistance protein) and *ssaB* (saliva-binding protein). These genes were appropriate for use in an MLST scheme because they yielded sequences of > 450 base pairs, were present as a single copy and were dispersed around the *S. sanguinis* genome reducing linkage between alleles (Maiden *et al.*, 1998). Primers to amplify and sequence these genes were either those reported by Hoshino *et al.* 2005 or were designed using gene sequences from *S. sanguinis* SK36 (Xu *et al.*, 2007). These primers are shown in Table 2.

DNA extraction, polymerase chain reaction conditions and nucleotide sequencing

Bacteria grown anaerobically overnight on FAA agar plates were suspended in 100 μl sterile ultra-high quality (UHQ) water and heated on a heat block (Microtherm Camlab, Cambridge, UK) for 10 min at 100°C with shaking at 50 r.p.m. to disrupt cells. The samples were centrifuged at 15,800 ***g*** for 10 s, the supernatant containing genomic DNA was used directly in the polymerase chain reaction (PCR). The primers used for the initial amplifications and subsequent sequencing reactions are shown in Table 2. The 25 μl reaction mixtures consisted of 22.5 μl 1.1× ReddyMix PCR master mix (1.5 mM $MgCl_2$) (Thermo Scientific ABgene, Epsom, UK), 0.5 μl of each forward and reverse primer (10 μM initial concentration) and 1.5 μl of the DNA template. The cycling parameters were: initial denaturation for 10 min at 94°C followed by 29 cycles of 94°C for 30 s, 49°C for 30 s and 72°C for 90 s. A final extension was carried out for 10 min at 72°C. The PCR amplicons were purified using MicroClean (Microzone, Haywards Heath, UK), rehydrated in the same volume of UHQ water and stored at −20°C. Both strands of the amplicons were sequenced in reactions containing 2 μl PCR product, 0.5 μl BigDye v.3.0 (Applied Biosystems, Warrington, UK), 1.75 μl of 5× solution buffer (Applied Biosystems), 1.75 μl sterile UHQ water and 4 μl primer (3 pmol). The cycling protocol and cleaning of sequencing reaction products were carried out as described in the manufacturer's protocols and sequencing was performed using an ABI 3730xl DNA Analyzer (Applied Biosystems).

Sequence data analysis

The sequences of the individual amplicons were inspected using BioEdit (Hall, 1999; http://www.mbio.ncsu.edu/BioEdit/bioedit.html) and the forward and reverse sequences were aligned. The

Table 2 Forward and reverse polymerase chain reaction (PCR) primers for gene amplification and sequencing

PCR target	Primers	Amplification primers (5′ → 3′) For PCR and sequencing	
carB	carb-F1	ATACGACCAGCGCATTGTCT	This study
	carb-F2	AGGCTGCTGAGTTTGA	
	carb-R1f	GGTCATTGGTTCTGGT	
	carb-R2	AACATTGGGTCACAGG	
czcD	czcd-F1	AGGCAATACACCTTGG	This study
	czcd-F2	TAGTCTTTTAGGGGCC	
	czcd-R1	TGACATTGGTCAGAGCCT	
	czcd-R2	TCTAGACCATCCATGGAC	
ddl	ddl-F	GCYATGGATAAAATYACRAC	Hoshino *et al.* (2005)
	ddl-R	CCACTGGKTRAARCCTGGCAGRGT	
dnaX	dnax-F1	CTGTCCGAATCAAAAGG	This study
	dnax-F2	TTTTCTGGTCCTCGTG	
	dnax-F3	TAAACAAGCGGTGGAG	
	dnax-R1	TAGACAAGGCATCCCGCATA	
	dnax-R2	CGGCAATGTCTGTAGTCA	
	dnax-R3	CAGCTCATTTCCCTGT	
	dnax-R4	TCCTGTGTCAAACTCAAGGC	
gdh	gdh-F	CGTGGYGGCTAYTATGACC	Hoshino *et al.* (2005)
	gdh-R	CYTCRTCCCAGTGRCTRAARTTRG	
rpoB	rpob-F	AARYTIGGMCCTGAAGAAAT	Hoshino *et al.* (2005)
	rpob-R	TGIARTTTRTCATCAACCATGTG	
sodA	soda-F1	TRCAYCATGAYAARCACCAT	Hoshino *et al.* (2005)
	soda-R1	ARRTARTAMGCRTGYTCCCARACRTC	
	soda-F2	GGAGCCTTACATTG	
	soda-R2	TTCGGACGAAGATTGC	
bacA	baca-F	GTCTCCTTCTTCTC	This study
	baca-R	TGGAAACTGGCGGTAA	
ssaB	ssab-F	CCAGACCATTCCCATT	This study
	ssab-R	TGGCGGTTGTCGTTAT	

sequences of each locus were read in-frame and aligned using ClustalW (Thompson *et al.*, 1994) and all unique sequences of a given locus were assigned an allele number and each unique combination of allele numbers was assigned a sequence type (Maiden *et al.*, 1998). The G+C mol% content, the number of alleles, the number of polymorphic sites, the non-synonymous/synonymous ratio (d_N/d_S), which is a measure for changes in nucleotide sequence that lead to alterations in amino acid sequence (Nei & Gojobori, 1986), the nucleotide diversity per site (π), the mutation rate per site (θ) and Tajima's *D* (Tajima, 1989) were calculated using DnaSP (Librado & Rozas, 2009).

Tests for recombination

Split decomposition trees were constructed with 1000 bootstrap replicates based on parsimony splits as implemented in SplitsTree 4.0 (http://www.splitstree.org) (Huson & Bryant, 2006) using the data for individual alleles and for concatenated alleles of all strains with statistical evidence for recombination determined using the pairwise homoplasy index test (Bruen *et al.*, 2006). When these split decomposition trees indicated that individual allele sequences diverged from the majority of sequences these individual sequences were searched using BLAST against the NCBI database (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) to confirm, or otherwise, their identity as being most like other *S. sanguinis* sequences.

Evidence for recombination was also sought by analysing the alleles using the recombination detection and analysis algorithms, implemented in the Recombination Detection Program (RDP) (Martin & Rybicki, 2000; http://darwin.uvigo.es/rdp/rdp.html). Evidence for recombination was accepted if significant ($P < 0.01$) evidence for the same recombination

event was demonstrated with at least two tests. Additionally, evidence for recombination was investigated by the construction of maximum likelihood (ML) phylogenetic trees using the Phylogenetic Analysis Using Parsimony (PAUP) program (Swofford, 2003). The ML trees for each of the nine genes were computed and compared using the Shimodaira–Hasegawa test to identify significant differences between the tree topologies (differences in log likelihood, Δ-ln *L*). The extent of congruence among the ML trees was assessed by randomization tests (Holmes *et al.*, 1999), in which the Δ-ln *L* values for each of the seven loci were compared with the equivalent values computed for 200 random trees created from each locus. The null hypothesis of linkage equilibrium was tested by calculating the standardized index of association, I_A^S (Haubold *et al.*, 1998), as implemented in START2 (Jolley *et al.*, 2001).

Further phylogenetic analysis of the sequence data was performed using ClonalFrame (Didelot & Falush, 2007). To assess the influence of recombination on the phylogenetic relationships between STs, 50% consensus trees were created for all STs derived using seven loci and also derived using the nine loci. ClonalFrame was also used to calculate r/m (Relative impact of recombination in comparison to point mutation in the genetic diversification of the lineage) and ρ/θ (Relative frequency of occurrence of recombination in comparison to point mutation in the history of the lineage) of the concatenated sequences for both sets of loci. Six independent runs were performed for each model, using the default settings except for the value of delta which was set to 0.0005 and a harmonic mean value of 13.989 was used.

RESULTS AND DISCUSSION

The 55 strains included in this study were each unique on the basis of the sequence comparisons of the partial sequences of the seven housekeeping genes (Table 1). Such a high level of ST diversity has been reported previously for other oral bacteria including *Porphyromonas gingivalis*, *S. oralis*, *S. mitis*, *S. mutans*, *Streptococcus salivarius* and *Streptococcus vestibularis* (Enersen *et al.*, 2006; Delorme *et al.*, 2007; Do *et al.*, 2009, 2010). In contrast, pathogenic streptococci, including *Streptococcus pneumoniae*, *Streptococcus uberis*, *Streptococcus agalactiae* and *Streptococcus zooepidemicus* exhibited less diversity with pathogenic clones identified in all species (Enright & Spratt, 1998; Jones *et al.*, 2003; Zadoks *et al.*, 2005; Webb *et al.*, 2008). It had been proposed that the inclusion of virulence determinants might increase the discriminatory power of an MLST scheme based solely on housekeeping genes (Edwards *et al.*, 2008). However, it is apparent that the inclusion of the two potential virulence determinants would not significantly increase the discriminatory power of the housekeeping gene sequence comparisons, as we also found in a study of *S. mutans* (Do *et al.*, 2010). The general properties of all nine alleles are shown in Table 3. The G+C (mol%) varied from 43% (*ssaB* and *sodA*) to 51% (*carB*) while the G+C (mol%) of the sequenced *S. sanguinis* strain was 43.4%. There was no significant relationship between locus length and number of alleles or number of polymorphic sites, which was not expected given that π (nucleotide diversity per site) and θ (mutation rate per site) were similar for all loci. The

Table 3 Gene characteristics for the nine loci in *Streptococcus sanguinis*

Locus	Length (bp)	G+C (mol%)	Alleles	Polymorphic sites	d_N	d_S	d_N/d_S	π	θ	Tajima's *D*
carB	510	51	33	107	0.0044	0.1874	0.024	0.05	0.06	−0.651
czcD	426	47	44	129	0.0064	0.2147	0.029	0.05	0.08	−1.014
ddl	429	49	31	124	0.0078	0.1540	0.051	0.04	0.07	−1.456
dnaX	423	47	38	120	0.0088	0.2654	0.033	0.06	0.07	−0.515
gdh	474	47	34	123	0.0098	0.1634	0.060	0.04	0.06	−0.958
rpoB	522	50	39	103	0.0039	0.1720	0.023	0.04	0.05	−0.371
sodA	390	43	29	105	0.0047	0.1010	0.046	0.03	0.06	−1.957*
bacA	483	48	34	94	0.0068	0.1327	0.051	0.04	0.05	−0.729
ssaB	453	43	31	90	0.0022	0.2173	0.010	0.05	0.05	−0.2305

*Statistical significance ($P < 0.05$).

d_N/d_S ratios were <1 indicting that all loci were subject to stabilizing selection. The significant negative values for Tajima's *D* signifies an excess of low-frequency polymorphisms, indicating population size expansion or positive selection. Here all values were negative, being only significant for *sodA*, suggesting that for this allele there maybe an excess of low-frequency polymorphisms.

Evidence of horizontal gene transfer, inter-species recombination, was apparent from the Neighbour-net trees derived using SplitsTree, as alleles ddl-6, gdh-17 and sodA-21 were outliers (data not shown). When BLAST searched, the ddl-6 exhibited only 89% similarity to *S. sanguinis* SK36 with the closest match being with the *ddl* of *Streptococcus gordonii* str. Challis substr. CH1, but with only 91% similarity. Gdh-17 had only 91% similarity with the *S. sanguinis* SK36 gene, and sodA-21 appeared to be derived from *S. oralis*, exhibiting 98% with *S. oralis* strain 2711. Other loci produced more compact trees and all other alleles exhibited >95% sequence similarity with sequences derived from organisms identified as *S. sanguinis* except for a small number of *ssaB* and *bacA* sequences which in some instances exhibited a similarity as low as 93% with the gene sequences of *S. sanguinis* SK36. We had previously reported sharing of *aroE* and *hexB* alleles between *S. oralis*, *S. mitis* and *S. pseudopneumoniae* (Do *et al.*, 2009). Horizontal gene transfer between *pepO*, *glcK* and *ilvC* alleles from different streptococcal species into *S. salivarius* has been reported (Delorme *et al.*, 2007). Hoshino *et al.* (2005) reported the presence of *rpoB* and *ddl* genes, originating from *Streptococcus sinensis* and *S. gordonii*, respectively, in *S. sanguinis* strains they investigated while Bishop *et al.* (2009) failed to demonstrate foreign genes in *S. sanguinis* isolates. The present data confirm the presence of foreign alleles, derived from other identified or unknown streptococcal species, in *S. sanguinis*.

The standardized index of association (I_A^S), which measures the degree of linkage in MLST data was determined using the concatenated sequences of all isolates and found to be 0.172. I_A^S is a function of the rate of recombination and is zero when all alleles are in linkage equilibrium, indicating that the distribution of any one allele is independent of all others. However, here the value I_A^S suggests a population in which recombination was widespread with all alleles in linkage equilibrium; all alleles are independent of each other. In other species this low value is suggestive of a population with a weak clonal structure similar to that of *Neisseria meningitidis* ($I_A^S = 0.14$) and *Campylobacter jejuni* ($I_A^S = 0.256$) but greater than *Neisseria gonorrhoeae* ($I_A^S = 0.005$), a panmictic species undergoing extensive recombination (Feil *et al.*, 2000; Suerbaum *et al.*, 2001). When the loci were examined individually for intra-species recombination using the pairwise homoplasy index test (Bruen *et al.*, 2006) significant evidence for recombination was found in *carB* ($P = 3.88 \times 10^{-5}$), *czcD* ($P = 0.02$), *ddl* ($P = 3.34 \times 10^{-5}$), *gdh* ($P = 0.029$), *rpoB* ($P = 1.14 \times 10^{-6}$) and *bacA* ($P = 1.7 \times 10^{-3}$). Significant evidence for recombination was found in all loci using the tests implemented in the RDP suite of programs, with one to three unique recombination events detected in individual loci, with the exception of *sodA* in which no significant recombination events were detected. The widespread evidence for recombination is similar to that previously reported for other oral streptococci but the failure to detect recombination in the *sodA* allele may indicate the inability of these methods of analysis to detect recombination when complete allele replacement has occurred.

The Shimodaira–Hasegawa test for congruence demonstrated that all ML trees generated for individual MLST loci were significantly incongruent (Table 4). However, there was evidence of an underlying low level of congruence between all trees, except *sodA*, as indicated by their lower Δ-ln *L* values which tended to be outside the range of values of the random trees (Fig. 1). In *S. oralis* we (Do *et al.*, 2009) reported that the Δ-ln *L* values for individual ML trees were consistently more similar to each other than randomly generated trees and demonstrated the small level of congruence observed was the result of a clonal complex structure. Whereas in *S. mutans* (Do *et al.*, 2010) all genes were incongruent with the Δ-ln *L* values of individual loci not different to the Δ-ln *L* values of random trees. The slight differences in the level of congruence between these three species may represent real differences, *S. oralis* is more competent than the other species, or may be a result of strain selection resulting in a biased result.

The phylogenetic trees resulting from the ClonalFrame analyses of the data with the seven housekeeping loci and all loci are shown in Fig. 2. With an increased number of loci the number of single STs

Table 4 Tests for congruence among *Streptococcus sanguinis* maximum likelihood gene trees and random trees (with $P < 0.05$)

Locus	-ln *L*	Δ-ln *L*	*P*	Δ-ln *L* of random trees
carB	2354.561	813.801–984.146	<0.001	931.621–1170.312
czcD	2558.868	781.885–965.333	<0.001	842.376–1076.000
ddl	2011.759	695.318–917.312	<0.001	750.510–969.400
dnaX	2310.520	1193.677–1361.762	<0.001	1310.378–1632.689
gdh	2080.356	895.828–1044.537	<0.001	924.306–1189.887
rpoB	2211.463	851.627–1064.622	<0.001	943.696–1196.609
sodA	1278.125	402.880–466.064	<0.001	373.742–490.448
bacA	1819.487	724.313–819.818	<0.001	805.018–996.964
ssaB	1786.870	832.756–1129.940	<0.001	960.390–1238.306

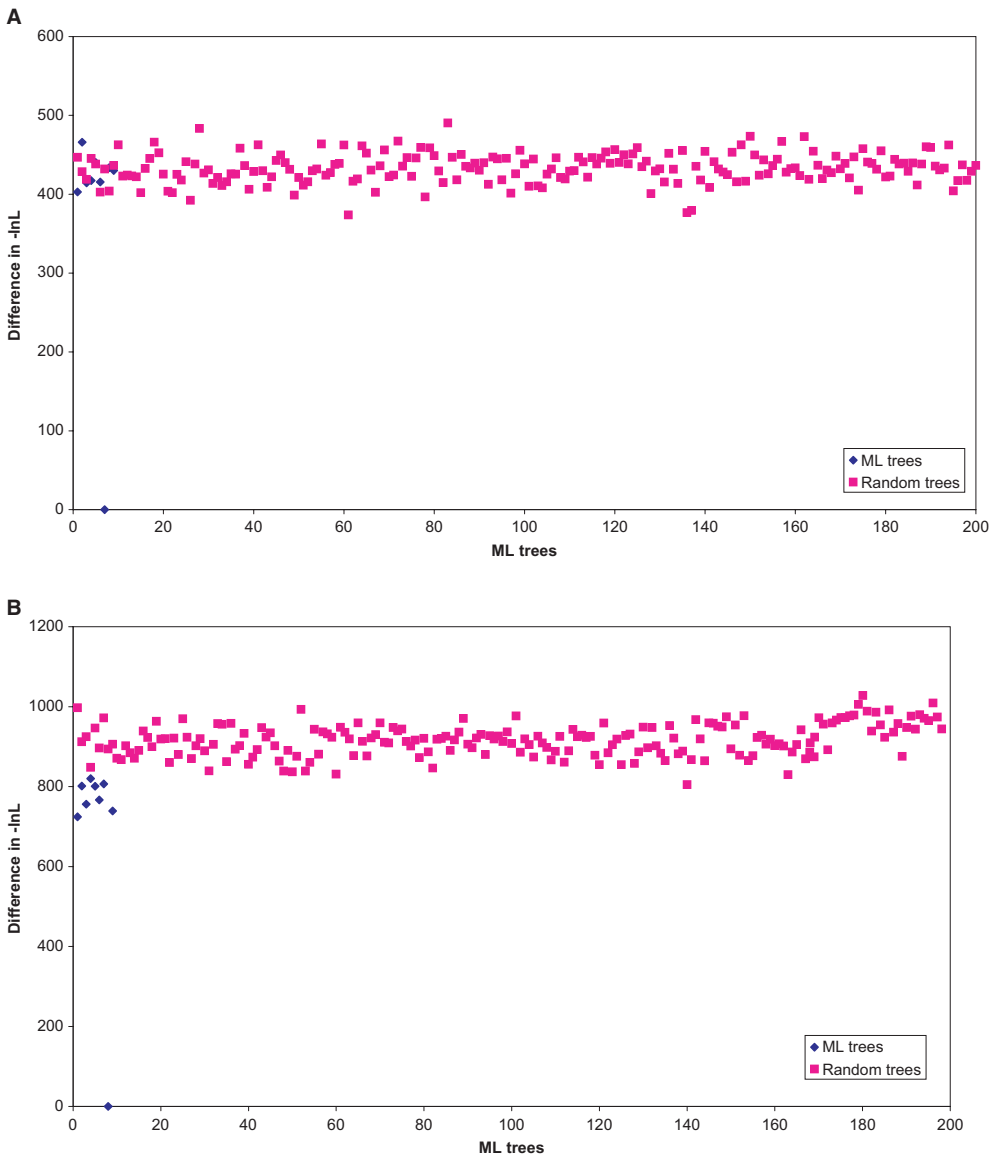

Figure 1 Congruence of the (A) *sodA* and (B) *bacA* maximum likelihood trees with trees generated from other multi-locus sequence type loci and 200 random trees.

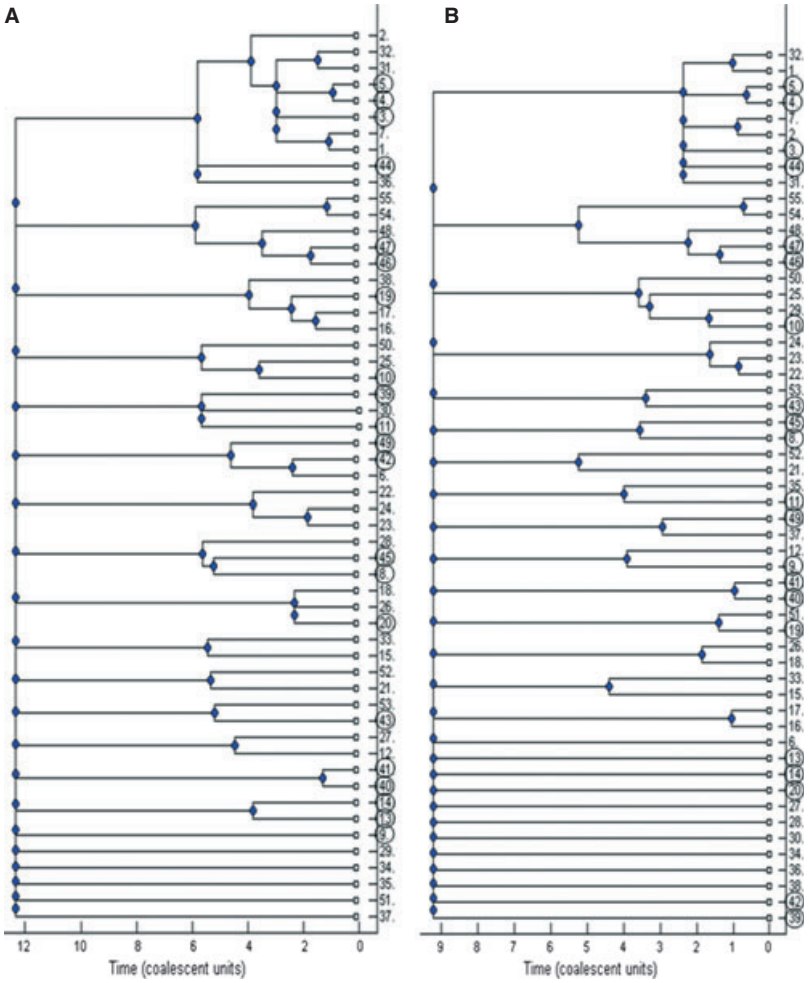

Figure 2 Phylogenetic trees of the *Streptococcus sanguinis* sequence types. A 50% majority-rule consensus phylogenetic tree derived from six trees generated with the ClonalFrame program for the seven housekeeping loci. Phylogenetic trees were constructed for (A) seven housekeeping loci, (B) all nine loci. Endocarditis strain sequence type numbers are circled.

increased from 6 to 12 as would be expected. However, there was weak clonal structure apparent in both trees with many STs forming distinct branches similar to the situation we have previously observed with *S. oralis* and to a lesser extent with *S. mutans*. The cluster comprising STs 2, 32, 31, 3, 4, 5, 7, 1, 44 and 36 in the housekeeping dendrogram is apparently more coherent in the nine-loci tree although ST 36 is no longer a member of the group, being now found as a singleton.

From the ClonalFrame analyses the r/m ratio (Relative impact of recombination in comparison to point mutation in the genetic diversification of the lineage) for the concatenated housekeeping gene sequences was 1.67 (95% confidence limits 1.25–2.72) smaller than that of other streptococci including *S. pneumoniae* (17.2), *S. mutans* (8.3) and *Streptococcus pyogenes* (23.1) but greater than some other *Firmicutes* including *Clostridium difficile* (0.2), *Enterococcus faecalis* (0.6), *Lactobacillus casei* (0.1) and *Enterococcus faecium* (1.1) (Vos & Didelot, 2009). The r/m for the putative virulence determinants of *S. sanguinis* was higher (3.99, 95% confidence limits 1.61–8.72) than the r/m for the housekeeping genes therefore recombination was more important than mutation in the diversification of the housekeeping genes. By contrast the r/m of two virulence genes in *S. mutans* [*gtfB* (glucosyltransferase B) and *spaP* (surface protein antigen I/II)] was 0.67 and lower than both the value for the *S. mutans* housekeeping genes and the virulence genes of *S. sanguinis*. This was also the case in the two putative virulence genes of *S. sanguinis*

but in *S. mutans* mutation was the driver (r/m = 0.67, 95% confidence interval 0.3–1.15) of diversification in the virulence determinants. We had previously suggested that positive selection of amino acids was expected in the *S. mutans* proteins as similar surface proteins also exhibit high variability and mosaic allele structures with evidence for positive selection mediated by immunogenic pressure (McGraw *et al.*, 1999; Lachenauer *et al.*, 2000). This also suggests that immune selection, mediated by mutation, was not the driver for the variation observed in the *S. sanguinis* putative virulence genes. The ratio ρ/θ Relative frequency of occurrence of recombination in comparison to point mutation in the history of the lineage) of the concatenated housekeeping genes was 0.13 (95% confidence limits 0.097–0.18), similar to that of *Listeria monocytogenes* Lineage I (0.13, 95% confidence limits 0.03–0.25) and *Campylobacter coli* (0.25) but less than that of *S. pneumoniae* (2.1) and *S. uberis* (9.05). Indicating that recombination in the housekeeping genes of *S. sanguinis* is relatively rare. In contrast, ρ/θ for the putative virulence genes was 1.45 (95% confidence limits 0.55–3.74) suggesting that the frequency of recombination was greater than the frequency of mutation.

The number of isolates used in this study was small, and they represented a convenience sample selected from several sources. Nevertheless, there was no evidence that the endocarditis strains formed a discrete cluster(s) distinct from the oral isolates, which supports the widely held opinion that infecting *S. sanguinis* strains are opportunistic human pathogens.

ACKNOWLEDGEMENTS

The authors acknowledge support from the Dental Institute, King's College London and from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust.

REFERENCES

Aas, J.A., Paster, B.J., Stokes, L.N., Olsen, I. and Dewhirst, F.E. (2005) Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**: 5721–5732.

Arbique, J.C., Poyart, C., Trieu-Cuot, P. *et al.* (2004) Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. *J Clin Microbiol* **42**: 4686–4696.

Beighton, D., Hardie, J.M. and Whiley, R.A. (1991) A scheme for the identification of viridans streptococci. *J Med Microbiol* **35**: 367–372.

Bishop, C.J., Aanensen, D.M., Jordan, G.E., Kilian, M., Hanage, W.P. and Spratt, B.G. (2009) Assigning strains to bacterial species via the internet. *BMC Biol* **7**: 3.

Bruen, T.C., Philippe, H. and Bryant, D. (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.

Burnette-Curley, D., Wells, V., Viscount, H. *et al.* (1995) *FimA*, a major virulence factor associated with *Streptococcus parasanguis* endocarditis. *Infect Immun* **63**: 4669–4674.

Chen, C.C., Teng, L.J., Kaiung, S. and Chang, T.C. (2005) Identification of clinically relevant viridans streptococci by sequence analysis of the 16S-23S ribosomal DNA spacer region. *J Clin Microbiol* **42**: 2651–2657.

Cole, J.R., Wang, Q., Cardenas, E. *et al.* (2009) The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

Das, S., Kanamoto, T., Ge, X. *et al.* (2009) Contribution of lipoproteins and lipoprotein processing to endocarditis virulence in *Streptococcus sanguinis*. *J Bacteriol* **191**: 4166–4179.

Delorme, C., Poyart, C., Ehrlich, S.D. and Renault, P. (2007) Extent of horizontal gene transfer in evolution of Streptococci of the salivarius group. *J Bacteriol* **189**: 1330–1341.

Dewhirst, F.E., Chen, T., Izard, J. *et al.* (2010) The human oral microbiome. *J Bacteriol* **192**: 5002–5017.

Didelot, X. and Falush, D. (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**: 1251–1266.

Do, T., Jolley, K.A., Maiden, M.C. *et al.* (2009) Population structure of *Streptococcus oralis*. *Microbiology* **155**: 2593–2602.

Do, T., Gilbert, S.C., Clark, D. *et al.* (2010) Generation of diversity in *Streptococcus mutans* genes demonstrated by MLST. *PLoS ONE* **5**: e9073.

Drancourt, M., Roux, V., Fournier, P.E. and Raoult, D. (2004) *rpoB* gene sequence-based identification of aerobic gram-positive cocci of the genera *Streptococcus, Enterococcus*, *Gemella*, *Abiotrophia*, and *Granulicatella*. *J Clin Microbiol* **42**: 497–504.

Edwards, M.T., Fry, N.K. and Harrison, T.G. (2008) Clonal population structure of *Legionella pneumophila* inferred from allelic profiling. *Microbiology* **154**: 852–864.

Enersen, M., Olsen, I., van Winkelhoff, A.J. and Caugant, D.A. (2006) Multilocus sequence typing of *Porphyromonas gingivalis* strains from different geographic origins. *J Clin Microbiol* **44**: 35–41.

Enright, M.C. and Spratt, B.G. (1998) A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* **144**: 3049–3060.

Feil, E.J., Enright, M.C. and Spratt, B.G. (2000) Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Res Microbiol* **151**: 465–469.

Ganeshkumar, N., Hannam, P.M., Kolenbrander, P.E. and McBride, B.C. (1991) Nucleotide sequence of a gene coding for a saliva-binding protein (SsaB) from *Streptococcus sanguis* 12 and possible role of the protein in coaggregation with actinomyces. *Infect Immun* **59**: 1093–1099.

Garnier, F., Gerbaud, G., Courvalin, P. and Galimand, M. (1997) Identification of clinically relevant viridans group streptococci to the species level by PCR. *J Clin Microbiol* **35**: 2337–2341.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95–98.

Haubold, B., Travisano, M., Rainey, P.B. and Hudson, R.R. (1998) Detecting linkage disequilibrium in bacterial populations. *Genetics* **150**: 1341–1348.

Holmes, E.C., Urwin, R. and Maiden, M.C. (1999) The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol Biol Evol* **16**: 741–749.

Hoshino, T., Fujiwara, T. and Kilian, M. (2005) Use of phylogenetic and phenotypic analyses to identify nonhemolytic streptococci isolated from bacteremic patients. *J Clin Microbiol* **43**: 6073–6085.

Huson, D.H. and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Jolley, K.A., Feil, E.J., Chan, M.S. and Maiden, M.C. (2001) Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**: 1230–1231.

Jones, N., Bohnsack, J.F., Takahashi, S. *et al.* (2003) Multilocus sequence typing system for group B streptococcus. *J Clin Microbiol* **41**: 2530–2536.

Kawamura, Y., Whiley, R.A., Shu, S.-E., Ezaki, T. and Hardie, J.M. (1999) Genetic approaches to the identification of the mitis group with the genus *Streptococcus*. *Microbiology* **145**: 2605–2613.

Kilian, M., Mikkelsen, L. and Henrichsen, J. (1989) Taxonomic study of viridans streptococci; description of *Streptococcus gordonii* sp. nov. and emended descriptions of *Streptococcus sanguis* (White and Niven 1946), *Streptococcus oralis* (Bridge and Sneath, 1982) and *Streptococcus mitis* (Andrewes and Horder, 1906). *Int J Syst Bacteriol* **39**: 471–484.

Lachenauer, C., Creti, S.R., Michel, J.L. and Madoff, L.C. (2000) Mosaicism in the alphalike protein genes of group B streptococci. *Proc Natl Acad Sci U S A* **97**: 9630–9635.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Maiden, M.C., Bygraves, J.A., Feil, E. *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**: 3140–3145.

Martin, D.P., Williamson, C. and Posada, D. (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**: 260–262.

McGraw, E.A., Li, J., Selander, R.K. and Whittam, T.S. (1999) Molecular evolution and mosaic structure of alpha beta and gamma intimins of pathogenic *Escherichia coli*. *Mol Biol Evol* **16**: 12–22.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous andnonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.

Okwumabua, O., O'Connor, M. and Shull, E. (2003) A polymerase chain reaction (PCR) assay specific for *Streptococcus suis* based on the gene encoding the glutamate dehydrogenase. *FEMS Microbiol Lett* **218**: 79–84.

Paik, S., Brown, A., Munro, C.L., Cornelissen, C.N. and Kitten, T. (2003) The *sloABCR* operon of *Streptococcus mutans* encodes an Mn and Fe transport system required for endocarditis virulence and its Mn-dependent repressor. *J Bacteriol* **185**: 5967–5975.

Poyart, C., Quesne, G., Coulon, S., Berche, P. and Trieu-Cuot, P. (1998) Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J Clin Microbiol* **36**: 41–47.

Poyart, C., Quesne, G. and Trieu-Cuot, P. (2002) Taxonomic dissection of the *Streptococcus bovis* group by

analysis of manganese-dependent superoxide dismutase gene (*sodA*) sequences: reclassification of '*Streptococcus infantarius* subsp. *coli*' as *Streptococcus lutetiensis* sp. nov. and of *Streptococcus bovis* biotype 11.2 as *Streptococcus pasteurianus* sp. nov. *Int J Syst Evol Microbiol* **52**: 1247–1255.

Suerbaum, S., Lohrengel, M., Sonnevend, A., Ruberg, F. and Kist, M. (2001) Allelic diversity and recombination in *Campylobacter jejuni*. *J Bacteriol* **183**: 2553–2559.

Swofford, D. (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 Beta 10*. Sunderland, MA: Sinauer Associates.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics* **123**: 585–595.

Teng, L.J., Hsueh, P.R., Tsai, J.C. *et al.* (2002) *groESL* sequence determination, phylogenetic analysis, and species differentiation for viridans group streptococci. *J Clin Microbiol* **40**: 3172–3178.

Thompson, J.D., Higgins, D.G. and 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.

Vos, M. and Didelot, X. (2009) A comparison of homologous recombination rates in bacteria and archaea. *ISME J* **3**: 199–208.

Webb, K., Jolley, K.A., Mitchell, Z. *et al.* (2008) Development of an unambiguous and discriminatory multilocus sequence typing scheme for the *Streptococcus zooepidemicus* group. *Microbiology* **154**: 3016–3024.

White, J.C. and Niven, C.F. (1946) Streptococcus s.b.e.: a streptococcus associated with subacute bacterial endocarditis. *J Bacteriol* **51**: 717–722.

Xu, P., Alves, J.M., Kitten, T. *et al.* (2007) Genome of the opportunistic pathogen *Streptococcus sanguinis*. *J Bacteriol* **189**: 3166–3175.

Zadoks, R.N., Schukken, Y.H. and Wiedmann, M. (2005) Multilocus sequence typing of *Streptococcus uberis* provides sensitive and epidemiologically relevant subtype information and reveals positive selection in the virulence gene *pauA*. *J Clin Microbiol* **43**: 2407–2417.