molecular oral microbiology

molecular oral microbiology

Physiological and molecular characterization of genetic competence in *Streptococcus sanguinis*

A.M. Rodriguez[1,2*], J.E. Callahan[1,2], P. Fawcett[3,4], X. Ge[1], P. Xu[1,2,4] and T. Kitten[1,2,4]

1 The Philips Institute of Oral and Craniofacial Molecular Biology, Virginia Commonwealth University, Richmond, VA, USA
2 Department of Microbiology and Immunology, Virginia Commonwealth University, Richmond, VA, USA
3 Department of Internal Medicine, Virginia Commonwealth University, Richmond, VA, USA
4 Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA, USA

Correspondence: Todd Kitten, The Philips Institute of Oral and Craniofacial Molecular Biology, Virginia Commonwealth University, 521 North 11th Street, Richmond, VA 23298-0566, USA Tel.: + 1 804 628 7010; fax: + 1 804 828 0150; E-mail: tkitten@vcu.edu
* Present address: PPD Inc., 2244 Dabney Road, Richmond, VA 23230, USA.

Keywords: *Streptococcus*; transformation; viridans
Accepted 16 December 2010
DOI: 10.1111/j.2041-1014.2011.00606.x

SUMMARY

Streptococcus sanguinis **is a major component of the oral flora and an important cause of infective endocarditis. Although *S. sanguinis* is naturally competent, genome sequencing has suggested significant differences in the *S. sanguinis* competence system relative to those of other streptococci. An *S. sanguinis* mutant possessing an in-frame deletion in the *comC* gene, which encodes competence-stimulating peptide (CSP), was created. Addition of synthetic CSP induced competence in this strain. Gene expression in this strain was monitored by microarray analysis at multiple time-points from 2.5 to 30 min after CSP addition, and verified by quantitative reverse transcription–polymerase chain reaction. Over 200 genes were identified whose expression was altered at least two-fold in at least one time point, with the majority upregulated. The 'late' response was typical of that seen in previous studies. However, comparison of the 'early' response in *S. sanguinis* with that of other oral streptococci revealed unexpected differences with regard to the number of genes induced, the nature of those genes, and their putative upstream regulatory sequences. *Streptococcus sanguinis* possesses a comparatively limited early response, which may define a minimal streptococcal competence regulatory circuit.**

INTRODUCTION

Streptococcus sanguinis is one of the earliest colonizers of teeth and is an abundant component of dental plaque (Carlsson, 1965; Aas *et al.*, 2005). It may be beneficial in the oral environment by interfering with colonization of teeth by *Streptococcus mutans* and other pathogens (Hillman *et al.*, 1985; Stingu *et al.*, 2008). When introduced into the bloodstream, however, *S. sanguinis* and other oral streptococci can prove harmful as a cause of infective endocarditis (Di Filippo *et al.*, 2006).

Streptococcus sanguinis possesses the ability to take up and incorporate free extracellular DNA from the environment (Gaustad & Håvarstein, 1997), a trait known as genetic competence. This property is shared by certain other bacteria, including several oral streptococcal species (Martin *et al.*, 2006). Genetic competence is best understood in *Streptococcus pneumoniae*, in which it was first discovered (Johnsborg & Håvarstein, 2009). Seven 'early' *com* genes are essential for induction and regulation of the *S. pneumoniae* competence system (Johnsborg &

Håvarstein, 2009). The *comC* gene encodes a 41-amino-acid precursor peptide that is cleaved and exported by the products of the *comAB* operon, ComA and ComB (Hui *et al.*, 1995). The mature 17-amino-acid ComC peptide, termed 'competence-stimulating peptide' (CSP), is sensed by the ComD sensor kinase, which presumably responds by phosphorylating the ComE response regulator. The latter two proteins are encoded along with ComC in the *comCDE* operon (Pestova *et al.*, 1996). ComE activates expression of the *comAB* and *comCDE* operons (Ween *et al.*, 1999; Johnsborg & Håvarstein, 2009). These genes form a 'quorum sensing' system, which results in amplification of the component gene products when the concentration of cells (and, hence, extracellular CSP) reaches a critical density. ComE also activates the *comX* gene, which is present in two identical copies. ComX serves as an alternative sigma factor, directing the expression of 'late' genes required for uptake and incorporation of DNA (Lee & Morrison, 1999). The *comW* gene product has been shown to stabilize ComX and activate it by an additional, undefined mechanism (Sung & Morrison, 2005). Studies in *Streptococcus gordonii* have shown that the genetic and physiological requirements for competence in this species are similar to those in *S. pneumoniae* (Håvarstein *et al.*, 1996; Lunsford, 1998).

Microarray analyses in *S. pneumoniae* (Peterson *et al.*, 2000, 2004; Dagkessamanskaia *et al.*, 2004) have extended earlier findings of transient, sequential gene expression upon competence induction (Alloing *et al.*, 1998; Rimini *et al.*, 2000). A similar analysis has been performed in *S. gordonii* (Vickerman *et al.*, 2007). In both species, addition of synthetic CSP to the culture medium produced two temporal waves of transcriptional induction. Genes categorized as 'early' were upregulated within 5 min of exposure to CSP, and included *comAB*, *comCDE*, and two identical copies of genes encoding the competence-specific sigma factor (ComX in *S. pneumoniae* and ComR in *S. gordonii*) (Peterson *et al.*, 2004; Vickerman *et al.*, 2007). Genes categorized as 'late' showed maximal expression 10–15 min after CSP exposure and most were dependent on the competence sigma factor for induction (Alloing *et al.*, 1998; Peterson *et al.*, 2004; Vickerman *et al.*, 2007). This category includes genes encoding proteins required for uptake and recombination of DNA. Transcription of both early and late genes returned to near basal levels in both species by 30–40 min. The kinetics of genetic competence are consistent with this expression pattern, with transformation frequencies peaking 10–20 min after CSP addition and decreasing dramatically by 40 min in both species (Håvarstein *et al.*, 1995b; Peterson *et al.*, 2004; Vickerman *et al.*, 2007).

Streptococcus mutans has also been the subject of a recent microarray analysis examining CSP-dependent gene expression (Perry *et al.*, 2009). Genetic competence in this species has been shown to differ from that in *S. pneumoniae* and *S. gordonii* in several respects (Martin *et al.*, 2006), including delayed induction of competence gene expression (Ahn *et al.*, 2006) and transformation (Kreth *et al.*, 2005) in response to CSP. The results were notable for the identification of a large number of upregulated bacteriocin genes (Perry *et al.*, 2009), in keeping with the earlier suggestion that the *comC* and *comDE* genes of *S. mutans* function primarily in induction of bacteriocin production rather than genetic competence (van der Ploeg, 2005; Martin *et al.*, 2006).

Because *S. sanguinis* is closely related to the fellow mitis group members *S. pneumoniae* and *S. gordonii*, it was expected to possess a similar competence system (Martin *et al.*, 2006).Yet, when its genome was sequenced (Xu *et al.*, 2007), it was discovered that two of the early genes indispensable for competence induction in *S. pneumoniae* (Peterson *et al.*, 2004) and *S. gordonii* (Lunsford & London, 1996), *comA* and *comB*, were seemingly absent from SK36. The genome sequence also confirmed that ComC of SK36, like that of *S. sanguinis* NCTC 7863 (Håvarstein *et al.*, 1997), lacks the Gly-Gly motif common to all other streptococcal ComC sequences examined to date, which serves as the recognition sequence for ComA-mediated cleavage and concomitant CSP secretion (Håvarstein *et al.*, 1995a; Ishii *et al.*, 2006). Fundamental components of the competence regulatory system in *S. sanguinis* therefore appear to differ from those in other mitis group species, as well as other transformable streptococci such as *S. mutans*, suggesting the need to examine competence more closely in *S. sanguinis*. We report here the initial characterization of genetic competence in *S. sanguinis* at the physiological and molecular levels. We have performed a comparison of the transcriptional response of *S. sanguinis* to CSP to that of *S. pneumoniae*, *S. gordonii* and *S. mutans*, allowing for the identification of common and unique features.

Streptococcus sanguinis exhibited a remarkably limited early gene response, characterized by fewer competence regulatory genes and non-competence genes than found in other streptococci.

METHODS

Bacterial strains, plasmids and growth conditions

A list of strains and plasmids used in this study is shown in Table 1. Strain SK36 of *S. sanguinis* was selected for this study based on the availability of its genome sequence (Xu *et al.*, 2007) and its demonstrated genetic competence (Paik *et al.*, 2005). SK36 and derivatives were routinely grown in brain–heart infusion broth (Bacto, Sparks, MD) at 37°C under reduced oxygen conditions (7% H_2, 7% CO_2, 80% N_2, 6% O_2). For transformation, strains were grown aerobically overnight at 37°C in Todd–Hewitt broth (Bacto) supplemented with 2.5% (volume/volume) horse serum (TH-HS; Invitrogen, Grand Island, NY). For examination of pH effects on growth, SK36 was grown overnight in TH-HS (pH 7.6) at 37°C then diluted 1 : 100 in TH-HS or TH-HS acidified with HCl in a 96-well plate and incubated at 37°C in a FLUOstar plate reader (BMG Labtechnologies, Durham, NC) where the optical density at 450 nm (OD_{450}) was recorded every 10 min immediately following agitation. Wells containing medium alone were used for background subtraction.

Transformation

Transformation was performed as previously described (Paik *et al.*, 2005) for mutant generation, or using a modified protocol for competence assessment. Unless otherwise indicated, overnight cultures of each strain grown as described above were diluted 1 : 200 into pre-warmed TH-HS, and incubated at 37°C until the OD_{660} reached approximately 0.07. Aliquots of 330 μl were then transferred into pre-warmed 0.7-ml microfuge tubes containing 70 ng

Table 1 Bacterial strains, plasmids, and primers used in this study

Strain, plasmid or primer	Description	References
Strains		
S. sanguinis		
SK36	Human dental plaque isolate	(Xu *et al.*, 2007)
JFP41	Δ*comC*; derived from SK36	This study
Plasmid		
pJFP16	Cm^r and Kn^r; containing *nrdD::magellan2*	(Turner *et al.*, 2009)
comC mutagenesis primers		
ComC-A	CGAACAAGAATACCTGATT	This study
ComC-B	TCCAATTATTTCAAGTGACATTAACTATCTCCTATCTTTTTA	This study
ComC-C	TCACTTGAAATAATTGGATTA	This study
ComC-D	TGGCACTCTCAATAGCA	This study
RT-PCR primers		
0195-FWD	CCTGCAGCACCGAAATTTAT	This study
0195-RVS	CTGAGAATATGGGCCAGCAT	This study
comD-FWD	CGAGAAAGAAGAGCGTTTGC	This study
comD-RVS	AATTTGCATCAGCCAAGACC	This study
comE-FWD	TCAAGTCAGGATGGAAACGAC	This study
comE-RVS	TTCCTCCCCCTTAATGTCAA	This study
comX-FWD_3	CGCCAGCCTTATGAAGAAGT	This study
comX-RVS_3	TCGCTTCTCTGAAGGCAACT	This study
comEA-FWD	CTGAGAGCAGGAGCGAGAGT	This study
comEA-RVS	GCCTCTGTCGCCGTATTTAG	This study
comYD-FWD	TACCAGGAGACGCAGAGGTT	This study
comYD-RVS	ACTGTCCTGTCCTCCGTCTG	This study
gapA-FWD	GCTTTCCGTCGTATCCAAAA	This study
gapA-RVS	ACACCGTCAGTAGCCCAGTC	This study
tuf-FWD	GCTTCAACTGACGGACCAAT	This study
tuf-RVS	CGTCACCTGGGAAGTCGTAT	This study

S. sanguinis CSP (Håvarstein *et al.*, 1997) and 10 ng pJFP16, a suicide plasmid containing the chloramphenicol (Cm) acetyl transferase gene flanked on either side by approximately 1 kb of the SK36 *nrdD* locus (Table 1). Preliminary studies revealed that 10 ng of this plasmid was saturating for transformation. Cells were incubated at 37°C for 20 min before addition of DNase I (10.5 ng μl^{-1}) and 10 min of further incubation to eliminate extracellular DNA. Cells were then serially diluted in phosphate-buffered saline, transferred to 12.5 ml tryptic soy broth (Bacto) supplemented with 1% (weight/volume) sterile molten low-melting agarose (Promega, Madison, WI), and dispensed into Petri dishes for solidification. Plates were then incubated for 2 h at 37°C to allow expression of antibiotic resistance. A top layer of 12.5 ml tryptic soy agar ± Cm (10 μg ml^{-1}) was then added, plates were again allowed to solidify, and then transferred to 37°C. Transformation frequency was determined after 48 h of incubation and calculated as colony-forming units (CFU) ml^{-1} from plates with Cm divided by CFU ml^{-1} from plates with no antibiotic.

Construction of a *comC* deletion mutant

An overlap extension polymerase chain reaction (PCR) technique (Ho *et al.*, 1989) was used to fuse the start codon of the *comC* gene to the second codon of *comD*, so deleting the *comC* gene and intervening sequence (Dagkessamanskaia *et al.*, 2004). The upstream fragment was generated with primers ComC-A and ComC-B, and the downstream fragment with primers ComC-C and ComC-D. The two fragments were fused in a final reaction containing ComC-A and ComC-D. All primers used for this study are listed in Table 1. The purified 2.9-kb fusion PCR product was introduced into SK36 by transformation, as described above. Mutants were identified via PCR screening using primers flanking the deletion. One mutant, designated JFP41, was selected for further study. DNA sequence analysis confirmed that it possessed the desired deletion and had no other mutations in the flanking sequences.

Microarray analysis

JFP41 was cultured in TH-HS, as for competence assessment. An aliquot was harvested for RNA isolation, and the remainder of the culture was treated with *S. sanguinis* CSP (Gaustad & Håvarstein, 1997). Additional samples were harvested at timed intervals. Detailed methods for RNA isolation, cDNA synthesis and labeling, microarray hybridization, and data analysis are provided in the Appendix S1. Microarray data have been submitted to the NCBI Gene Expression Omnibus (GEO) repository under accession number GSE21619.

Quantitative reverse transcription PCR

For analysis of SSA_0195 expression, 12.5-ml aliquots of JFP41 cultures were incubated for 5 min with 25 ml RNAprotect (Qiagen, Valencia, CA), then centrifuged for 20 min at 8000 ***g***. Cells were lysed according to the manufacturer's protocol for gram-positive bacteria, except that a Fastprep beadbeater with lysing matrix B (MP Biomedicals, Solon, OH) was used in place of disruption with acid-washed beads. RNA was isolated via the RNeasy Mini kit (Qiagen), with the optional DNase I treatment. Purified RNA was subjected to a second DNase treatment (Ambion, Austin, TX), quantified, and analysed for integrity as described in the Appendix S1. First-strand cDNA synthesis was performed with SuperScript III Reverse transcriptase (200 units μl^{-1}; Invitrogen). For analysis of *com* gene orthologs, aliquots of RNA samples previously isolated for microarray analysis were subjected to an additional DNase treatment (Ambion) and purified as per the manufacturer's protocol. The cDNA templates were created via Ready-To-Go™ You-Prime First-Strand Beads (Amersham Biosciences, Pittsburgh, PA) as described in the manufacturer's protocol. For both analyses, quantitative reverse trasncription (qRT-) PCR were carried out in a 7500 Fast RT-PCR system (Applied Biosystems, Carlsbad, CA) using RT^2 Real-Time™ SYBR Green/Rox PCR master mix (Qiagen) and the primers listed in Table 1. Control reactions performed in the absence of reverse transcriptase confirmed the absence of contaminating genomic DNA in all samples. Normalization was performed with the *tuf* or *gapA* genes. A standard curve was generated for each gene from DNA templates of known concentration.

Bioinformatic and phylogenetic analysis of putative upstream regulatory sites

For identification of potential ComE binding sites, conserved sequences (Ween *et al.*, 1999) upstream

from the *comC* gene in *S. pneumoniae*, *Streptococcus mitis*, *Streptococcus oralis*, *Streptococcus cristatus*, *S. gordonii* and *S. sanguinis* were entered into the program MEME for motif identification (http://meme.nbcr.net/) (Bailey & Elkan, 1994). The 31-bp position-specific scoring matrix (PSSM) produced by the program from the input sequences was used to search a database of nucleotide sequences extending 1000 bp upstream and 200 bp downstream from the start site of all annotated genes in *S. sanguinis* SK36 using the program MAST (Bailey & Gribskov, 1998). Default settings were employed except that reverse complement strands were treated as separate sequences. The same PSSM was used for a MAST search of *S. pneumoniae* (strain TIGR4), *S. gordonii* (strain CH1) and *S. mutans* (strain UA159) databases. The process was repeated after adding the *S. mutans* SMU.1914c-proximal sequence (van der Ploeg, 2005) to the six *comC* sequences for PSSM construction. Sites from each species identified in either search were used for further rounds of PSSM creation and database searching until no additional significant matches from that species emerged. Potential combox sites (Campbell *et al.*, 1998) upstream from *S. sanguinis* late genes were initially sought by manual inspection. A 200-bp region upstream from each gene identified in this way was then used for PSSM creation and iterative searching, as described for ComE. The search was discontinued when no additional significant matches were obtained. Phylogenetic analysis of putative ComE-binding sites was performed with the MEGA4 program (Tamura *et al.*, 2007). Phylogenetic trees were created using the neighbor-joining method with evolutionary distances computed using the Jukes–Cantor method.

RESULTS

Competence induction in *S. sanguinis*

Given the unique features of *S. sanguinis* competence regulatory genes determined by DNA sequencing, we were interested in further characterizing genetic competence in this species. A notable feature of *S. pneumoniae* is a rapid cell-density-dependent increase and decay of competence (Håvarstein *et al.*, 1995b). Competence was assessed in *S. sanguinis* SK36 grown to different densities as described in the Methods section, by addition of a saturating amount of pJFP16, a plasmid that confers Cm resistance by integration into the chromosomal *nrdD* locus (Turner *et al.*, 2009). As shown in Fig. 1(A), *S. sanguinis*

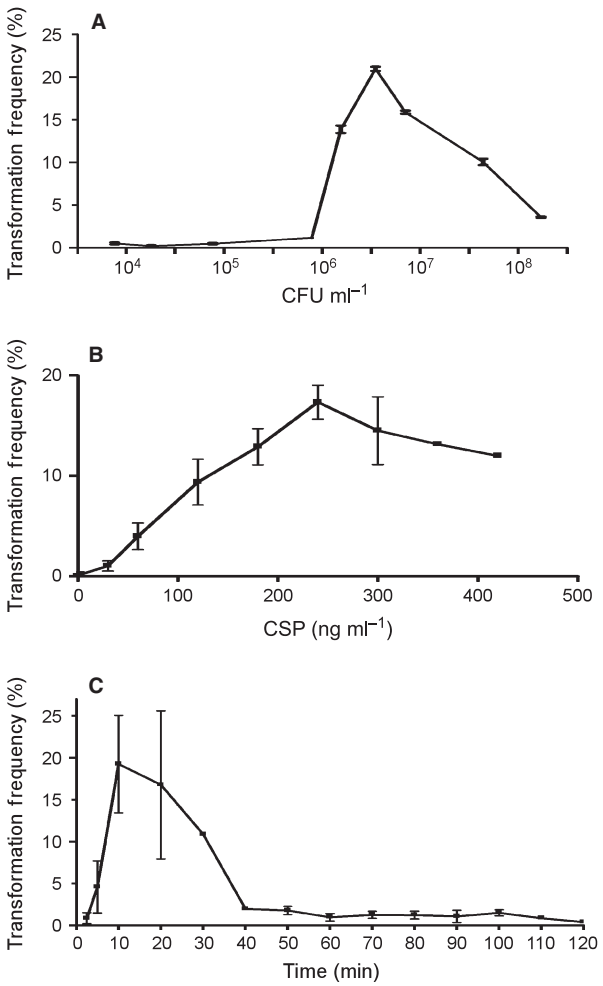

Figure 1 Characterization of competence development in SK36 and a *comC* mutant, JFP41. (A) The effect of culture density on transformation in SK36. Cells were cultured and assessed for transformation frequency as described in the Methods section. Error bars represent the SD derived from one experiment that was representative of six. Mean and SD values from separate experiments could not be combined because it was not possible to obtain identical cell densities in separate experiments. (B) The effect of competence-stimulating peptide (CSP) concentration on transformation frequency in JFP41. Exogenous CSP was added to cultures at the concentration indicated before transformation and plating. (C) Transformation of JFP41 after various durations of exposure to 220 ng ml^{-1} CSP. CSP was added at time 0, transforming DNA was added at the time indicated, and incubation was continued for 10 min before addition of DNase I and plating. Error bars in (B) and (C) represent SDs from the combined data from three separate experiments.

likewise demonstrated rapid induction of competence, in this case, when the cell density approached 10^6 CFU ml^{-1} (corresponding to an OD_{660} of 0.068–0.074). As the cell density surpassed 10^7 CFU ml^{-1}, competence decreased dramatically (Fig. 1A). Density dependence of transformation has also been demonstrated for *S. mutans* (Perry & Kuramitsu, 1981) and *S. gordonii* (Pakula & Walczak, 1963).

A primary goal of this study was to perform a transcriptomic analysis of gene regulation in response to competence induction. Previous studies performed with *S. gordonii* and *S. pneumoniae* have employed growth in acidic conditions to inhibit natural competence, allowing for the controlled induction of competence by addition of exogenous CSP (Peterson *et al.*, 2000, 2004; Vickerman *et al.*, 2007). The acidic conditions did not affect the growth of either species (Peterson *et al.*, 2000; Vickerman *et al.*, 2007). To determine whether this strategy could be used with *S. sanguinis*, SK36 was cultured in normal (pH 7.6) and acidic (pH 6.9) TH-HS before transformation with pJFP16. The transformation frequency decreased from approximately 12% at pH 7.6 to below the level of detection (3.4×10^{-5}%) at pH 6.9, suggesting that natural competence was effectively inhibited. It was observed, however, that cultures grown in the pH 6.9 medium had significantly lower cell densities than those grown in neutral conditions (data not shown). When an intermediate condition of pH 7.3 was used, growth was only moderately affected; however, competence was also partially restored (data not shown).

Given these results, we sought to create a *comC* deletion mutant in which competence could be induced by the addition of exogenous CSP. The *comC* gene in *S. sanguinis* is presumably co-transcribed with the adjacent *comD* and *comE* genes, as in *S. pneumoniae* and *S. gordonii*. To minimize effects on expression of these downstream genes, which are required for competence induction in other streptococci, a PCR overlap extension technique (Ho *et al.*, 1989) was used to fuse the start codon of *comC* to the second codon of *comD*, deleting *comC*, as well as the 29-bp *comCD* intergenic region. Test transformations were performed as described above. In the absence of added CSP, transformation of JFP41 could not be detected. When synthetic CSP was added, competence was restored to near wild-type levels (data not shown). The results demonstrated two critical phenotypes expected for the mutant: first, that natural competence was drastically reduced or eliminated, as would be expected for a mutant incapable of CSP production; and second, that competence was restored upon the addition of exogenous CSP, suggesting that the *comD* and *comE* genes were subject to normal CSP-dependent induction. An identical strategy was used previously for microarray analysis of CSP-dependent gene expression in *S. pneumoniae* (Dagkessamanskaia *et al.*, 2004).

The effect of varying concentrations of CSP on JFP41 transformation frequency was also examined. As shown in Fig. 1(B), competence increased with increasing CSP concentration until saturation, peaking at approximately 200–240 ng ml^{-1}. The concentration of 220 ng ml^{-1} was chosen for the remaining studies.

Finally, the timing of competence induction following addition of exogenous CSP to JFP41 cultures was examined. JFP41 cells at an OD_{660} of approximately 0.07 were exposed to CSP, and at varying intervals, transforming DNA was added and the cultures were examined for transformation frequency as above, except that incubation was continued for only 10 min before addition of DNase. The results are shown in Fig. 1(C). JFP41 competence peaked after approximately 10 min of exposure to CSP and returned to near baseline levels by 40 min. These results are similar to those observed in *S. pneumoniae* (Håvarstein *et al.*, 1995b; Rimini *et al.*, 2000; Peterson *et al.*, 2004) and *S. gordonii* (Vickerman *et al.*, 2007), and differ from *S. mutans*, in which a 2-h delay was noted between addition of CSP and development of competence (Kreth *et al.*, 2005).

Microarray analysis of transcriptional response to CSP

Conditions were chosen for microarray analysis based on the results shown in Fig. 1. JFP41 cells were grown to an OD_{660} of 0.068–0.074 (3.2×10^6–4.2×10^6 CFU ml^{-1}), and a culture aliquot was removed for RNA isolation as described in the Appendix S1. CSP was then added to 220 ng ml^{-1}, and additional culture aliquots harvested for RNA isolation 2.5–30 min later. The RNA isolation protocol employed hot acid phenol to quickly halt transcription, as with previous microarray analyses of the CSP

response in *S. pneumoniae* (Peterson *et al.*, 2000, 2004) and *S. gordonii* (Vickerman *et al.*, 2007).

Data were collected from microarrays and filtered based on signal quality and significance (see Appendix S1). Of the 2394 genes represented on the array, 205 were identified whose expression was statistically verified and altered at least two-fold for at least one time point relative to the untreated sample from the same culture; 122 genes were upregulated, and 83 were downregulated. Normalized values for the *S. sanguinis* genes, expressed as the $\log_2$ of the ratio of the values for CSP-treated samples relative to untreated samples, are shown in Table S1. These data are displayed in a colorimetric graph (Fig. 2) in

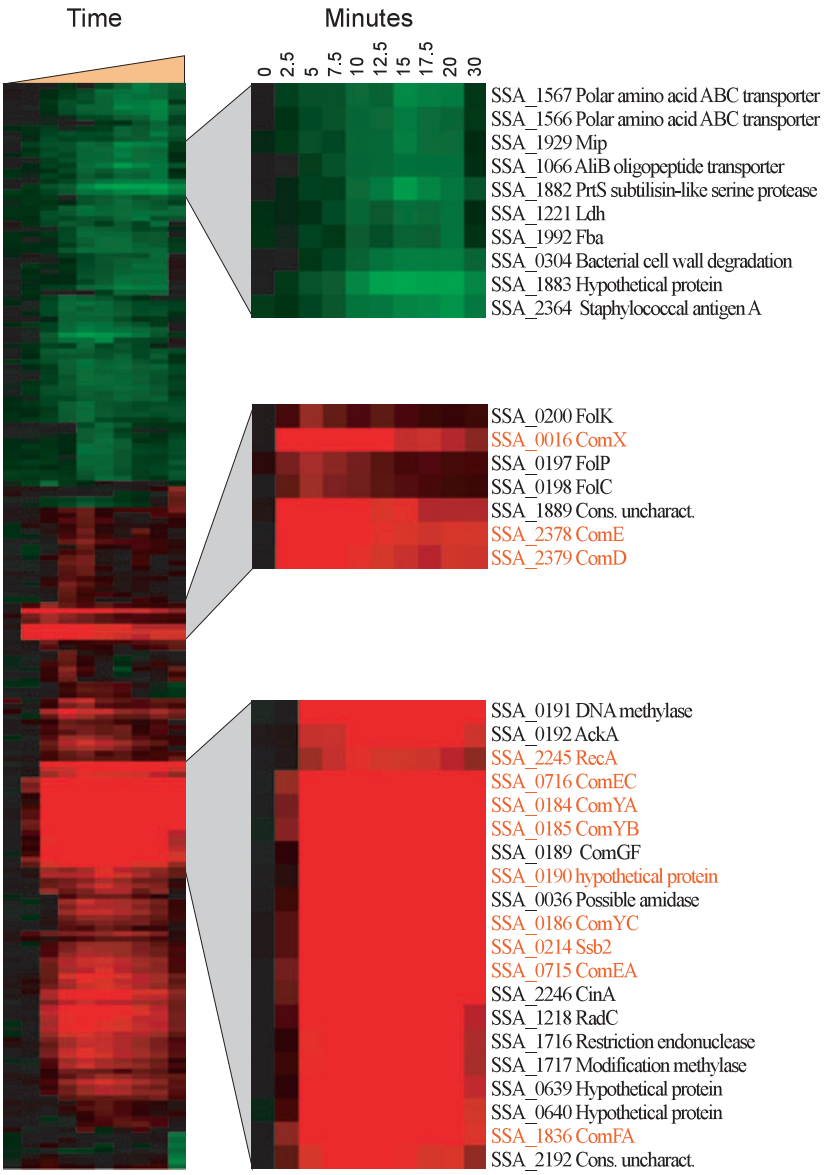

Figure 2 Colorimetric map of *Streptococcus sanguinis* genes whose expression was statistically validated and regulated at least two-fold by competence-stimulating peptide (CSP). Each column represents RNA sampled from one time point, and each row represents the expression of a single gene relative to the '0' time point. Red squares indicate upregulation and green squares indicate downregulation, with the intensity of the color being proportional to the degree of change. Black squares indicate no significant difference, and gray squares indicate incomplete data. Genes were clustered based on similarity of expression patterns using the CLUSTER 3.0 program (Eisen *et al.*, 1998) and visualized using TREEVIEW (Eisen *et al.*, 1998). Data were derived from three separate experiments. Enlargements indicate clusters that include genes with the highest levels of regulation. Labels in red indicate orthologs of genes that have been shown to be required for transformation in *Streptococcus pneumoniae* (Peterson *et al.*, 2004).

which genes are clustered based on similarity of expression patterns.

Early-induced genes

The induced genes were categorized as 'early' or 'late', with the early category comprising four genes that demonstrated greater than five-fold upregulation at 2.5 min, peak expression within 10 min, and reduced expression at later time points: the known competence regulatory genes, *comD*, *comE*, *comX*; and a gene encoding a conserved uncharacterized protein, SSA_1889 (Fig. 2). This group corresponds kinetically to the 'early' competence gene category identified previously in *S. gordonii* (Vickerman *et al.*, 2007) and a previous *S. pneumoniae* analysis (Peterson *et al.*, 2004). (Additional microarray analyses in *S. pneumoniae* and *S. mutans* did not apply this classification. Dagkessamanskaia *et al.*, 2004; Perry *et al.*, 2009) Table 2 lists the early genes from *S. sanguinis*, *S. pneumoniae* and *S. gordonii*, along with any orthologs present in the other species, which were identified as mutual best hits by BLASTP, which also had alignment *e* values $<10^{-5}$. The most striking feature of this comparison is the lower number of genes assigned to this category in *S. sanguinis*: five genes, in comparison to 28 for *S. pneumoniae* (Peterson *et al.*, 2004) and 35 for *S. gordonii* (Vickerman *et al.*, 2007). SK36 lacks recognizable genes encoding the CSP-transport proteins ComA and ComB, as well as a duplicate copy of the competence sigma factor gene (*comX* in *S. pneumoniae* or *comR* in *S. gordonii*), all of which occur in the other two species. Also, the *S. sanguinis comC* gene is not detectable in our study because of its deletion from the test strain. The greater contribution to this discrepancy, however, lies in the large number of early genes with no defined role in competence induction in the other two species: 20 in *S. pneumoniae* and 27 in *S. gordonii*. As shown in Table 2, these genes are remarkable in that their upregulation is almost entirely species-specific. SP_1918 in *S. pneumoniae* and SGO_1750 of *S. gordonii* encode orthologous ABC transport proteins, both of which belong to the early response. Every other gene is upregulated in only one of the three species.

We wondered whether this difference might be explained by stricter criteria used in our study for categorization of genes as early vs. late. Table 2 indicates that this is not the case. Only two genes, SSA_1184, encoding DNA topoisomerase I, and SSA_0036, encoding a putative cell wall amidase, were identified as late in *S. sanguinis* while their orthologs in one of the other two species (SGO_1197 and SGO_2094, respectively) were identified as early. Moreover, the two *S. gordonii* genes are preceded by combox sequences characteristic of late gene expression, as described below, and were expressed more highly at 15 and 40 min than at 5 min – another characteristic of late genes (Vickerman *et al.*, 2007). We also noted that a previous microarray analysis performed in *S. pneumoniae* employed a mixture of cDNA derived from all time points in place of a '0' time point reference to increase the reliability of identifying genes with low basal expression levels (Peterson *et al.*, 2004). We re-examined our data and noted a single gene that would have been classified as early had it not been filtered out, partly because of low expression at time 0 (see Appendix S1). This gene, SSA_0195, encoding a hypothetical protein, exhibited kinetics consistent with early gene expression, appearing upregulated more than 15-fold at 2.5 min and more than four-fold at all later time points, suggesting that it might be an early gene (data not shown). Expression of this gene was therefore examined by qRT-PCR, using *gapA* for normalization. SSA_0195 expression at time 0 was barely detectable. In comparison, mean normalized $\log_2$ values for expression of SSA_0195 at 2.5, 5 and 15 min post-CSP addition were 32.6, 36.1 and 30.6, respectively, indicating strong and rapid induction. Inclusion of SSA_0195 along with the missing *comC* gene suggests a total of six early genes: *comCDE*, *comX*, SSA_1889 and SSA_0195.

Late-induced genes

Apart from the early genes indicated above, all other upregulated genes (Table S1) were classified as 'late'. All were expressed at levels lower than the early genes at 2.5 min. Those genes whose expression was induced most strongly and persistently are indicated in Fig. 2. Most are orthologs of genes previously associated with competence, including nine of the 14 late genes shown to be essential for transformation in *S. pneumoniae* (Peterson *et al.*, 2004). The remainder are orthologs of CSP-induced genes previously identified in *S. pneumoniae*, *S. gordonii* or *S. mutans*, with the exception of SSA_0639 and SSA_0640, which encode hypothetical proteins and are found only in *S. sanguinis*.

Table 2 Early-induced genes and orthologs in *Streptococcus pneumoniae*, *Streptococcus sanguinis* and *Streptococcus gordonii*

S. pneumoniae		*S. sanguinis*		*S. gordonii*	
Locus[1]	Gene	Locus	Gene	Locus	Gene
SP_0014	***comX1***	**SSA_0016**	***comX***	**SGO_1707**	***comR1***
SP_2006	***comX2***			**SGO_2130**	***comR2***
SP_0018	***comW***	–		–	
SP_0019	***purA***	SSA_2185	*purA*	SGO_1989	*purA*
SP_0042	***comA***	–		**SGO_2097**	***comA***
SP_0043	***comB***	–		**SGO_2096**	***comB***
SP_0429		–		–	
SP_0430		–		–	
SP_0530	***blpA***	–		–	
SP_0545	***blpY***	–		–	
SP_0546	***blpZ***	–		–	
SP_0547		–		–	
SP_0635		–		–	
SP_1110		SSA_0936	*mreA*	SGO_1050	*ribF*
SP_1547		SSA_1395		SGO_1307	
SP_1548		–		–	
SP_1549	***def-2***	SSA_0458		SGO_1754	
SP_1716		–		–	
SP_1717		–		–	
SP_1918		SSA_0462		**SGO_1750**	
SP_1942		SSA_1909		SGO_0535	
SP_1943		SSA_1910		SGO_0534	
SP_1944		SSA_1911		SGO_0533	
SP_1945		–		–	
SP_2156		SSA_1965		SGO_1730	
SP_2235	***comE***	**SSA_2378**	***comE***	**SGO_2145**	***comE***
SP_2236	***comD***	**SSA_2379**	***comD***	**SGO_2146**	***comD***
SP_2237	***comC2***	–		–	
–		**SSA_1889**		–	
–		**SSA_0195**[2]		–	
–		–		**SGO_0097**	
SP_0158		SSA_2263	*nrdI*	**SGO_0098**	
–		–		**SGO_0115**	***sthB***
–		–		**SGO_0376**	
–		–		**SGO_0377**	
–		SSA_0621		**SGO_0647**	
SP_0785		SSA_1532	*acrA*	**SGO_0750**	
SP_0786		SSA_1531	*salX*	**SGO_0751**	
SP_0787		SSA_1530		**SGO_0752**	
SP_1381		SSA_1026		**SGO_0798**	
SP_1380		SSA_1027		**SGO_0799**	
SP_1162		SSA_1175	*acoC*	**SGO_1131**	***sucB***
SP_1263	*topA*	SSA_1184[3]	*topA*[3]	**SGO_1197**	***topA***
SP_0868		SSA_1955		**SGO_1721**	***sufD***
SP_2000		SSA_1972		**SGO_1731**	
SP_2001		SSA_1973		**SGO_1732**	
SP_2002		SSA_1974		**SGO_1733**	
SP_0338		SSA_2096	*clpL*	**SGO_1856**	
SP_1722		SSA_0456	*scrA*	**SGO_1857**	
–		SSA_0860		**SGO_2013**	
–		–		**SGO_2086**	
–		–		**SGO_2088**	

Table 2 (Continued)

S. pneumoniae		*S. sanguinis*		*S. gordonii*	
Locus[1]	Gene	Locus	Gene	Locus	Gene
SP_0054	*purK*	SSA_0040	*purK*	**SGO_2089**	***purK***
SP_0051	*purD*	SSA_0037	*purD*	**SGO_2092**	***purD***
–		–		**SGO_2093**	
–		SSA_0036[3]		**SGO_2094**	
–		–		**SGO_2095**	***comX***
–		–		**SGO_2147**	***comC***

[1]Genes listed on the same line are orthologs as determined by bidirectional BLASTP comparison in all combinations. Bold text indicates classification as an early gene in this study, that of Peterson *et al.* (2004), or Vickerman *et al.* (2007).
[2]Gene identified as early by quantitative reverse transcription polymerase chain reaction.
[3]Genes that were classified as belonging to the late competence response.

In *S. pneumoniae*, the late response was followed by a 'delayed' response, which was characterized by a continued increase in expression beyond that of the early and late genes (Peterson *et al.*, 2004). The delayed class contained several genes encoding chaperones and other stress-response proteins. A comparable response was not evident in *S. gordonii* (Vickerman *et al.*, 2007). In *S. sanguinis*, four orthologs of *S. pneumoniae* delayed-response genes, including *hrcA*, *grpE*, *dnaJ* and *groES* were significantly upregulated. Their expression did not display delayed kinetics, however, peaking 10–15 min after CSP addition.

The commonest function of the late genes, as assessed by their Cluster of Orthologous Group (Tatusov *et al.*, 2001) classification, was 'DNA replication, recombination and repair'. Seventeen genes were assigned to this category; fewer than half that number were assigned to any other. Table S2 presents a comparison of the late/delayed response of *S. pneumoniae* (Peterson *et al.*, 2004), *S. sanguinis* and *S. gordonii* (Vickerman *et al.*, 2007), as well as CSP-induced genes identified in *S. mutans* (Perry *et al.*, 2009). Orthologous genes upregulated in at least two of these species are shown, along with any orthologs present in the other species. Approximately half of the late genes (65 of 118) contained orthologs that were also upregulated by CSP in *S. pneumoniae*, *S. gordonii* or *S. mutans*. We suspected that many of the apparent differences among the species were the result of technical issues with individual arrays or induction levels falling just below the level of significance. In particular, we noted that the five SK36 orthologs of late genes identified previously as required for transformation in *S. pneumoniae* (Peterson *et al.*, 2004) that were not identified as late genes in our study were deleted during the filtering process rather than because of unchanged expression. Nevertheless, we wanted to confirm that this was not indicative of a fundamental difference in transformation machinery in the two species. We took advantage of an existing collection of SK36 allelic exchange mutants in which genes were replaced by the *aphA-3* gene encoding kanamycin resistance (Xu *et al.*, manuscript in preparation) to test the effect on the transformation efficiency of mutating 15 genes that were either orthologs of genes essential for transformation in *S. pneumoniae*, or were located in the same gene cluster. Transformation was assessed in SK36 and the mutants as above, except that the incubation with CSP and DNA was begun at a lower OD and extended to 3 h for increased sensitivity, and surface plating was used. The results are presented in Table S2, along with previously published data for *S. pneumoniae* (Peterson *et al.*, 2004) and *S. mutans* (Merritt *et al.*, 2005). Our results were comparable to those published previously. The combined mutant transformation data from *S. pneumoniae*, *S. sanguinis* and *S. mutans* suggest that the *comY* operon contains seven genes required for transformation. If so, this would suggest a shared core regulon of 16 late genes required for streptococcal transformation (see Table S2).

The comparison also suggests gene clusters with no obvious connection to competence that are shared among a subset of species. For example, the *fruRBK* gene cluster encoding genes involved in fructose uptake and metabolism is conserved in all four species, and has been shown to be co-transcribed in *S. mutans* (Wen *et al.*, 2001) and *S. gordonii* (Loo

et al., 2003). Yet, the entire operon is upregulated in *S. sanguinis* and *S. gordonii*, but not the other two species. The significance of these differences is unknown.

Downregulated genes

Genes significantly downregulated by CSP were fewer in number than those upregulated (Fig. 2 and Table S1). The most common categorizations of these genes were as encoding hypothetical proteins (14 genes), ribosomal proteins (13 genes), or ABC transporters (11 genes). Genes belonging to these three categories were also prevalent among the genes downregulated by CSP in *S. pneumoniae*, *S. gordonii* and *S. mutans* (Peterson *et al.*, 2004; Vickerman *et al.*, 2007; Perry *et al.*, 2009). These commonalities in broad functional categories generally did not extend to individual orthologous genes, however. Among hypothetical protein genes, SSA_1941 was orthologous to SGO_1699, but there were no other orthologs in any of the three species that were downregulated. Similarly, fewer than half of the downregulated *S. sanguinis* ABC transporters had orthologs that were downregulated in any of the other three species. Most tellingly, there were no orthologous genes that were downregulated in all four species (data not shown). The combined results suggest that CSP-mediated gene downregulation in these species may result primarily from non-specific responses to changes in growth rate and other cellular properties rather than from concerted, gene-specific repression.

Examination of selected competence genes by qRT-PCR

To confirm the results observed from the microarrays, qRT-PCR were performed with selected orthologs of known competence genes. Three early response genes, *comD*, *comE* and *comX*, were examined, along with the late gene *comEA* and the suspected late gene, *comYD*. Orthologs of this last gene are required for competence in *S. pneumoniae* and are CSP-induced in *S. pneumoniae*, *S. gordonii* and *S. mutans*. The microarray data suggested that the *tuf* gene, encoding elongation factor Tu, was expressed highly and constitutively at all time points. This gene was therefore chosen as a normalization control. Another housekeeping gene frequently used for normalization, *gapA*, was included as an additional control. As shown in Fig. 3, the results corresponded to those obtained from the microarray, with early gene expression peaking at 5 min and the late gene *comEA* peaking later and declining less during the course of the study. The *comYD* gene was not identified as a late gene in the microarray analysis, despite upregulation ≥32-fold at 5 min and beyond, because it was deleted during the filtering process (data not shown). Figure 3 indicates, however, that its expression pattern was almost identical to that of *comEA*, confirming its expected status as a late gene. This is also consistent with our finding that *comYD* is required for competence. Expression of the *gapA* gene was similar at all time points, as expected, suggesting that either the *tuf* or *gapA* genes can be used as normalization controls for these studies.

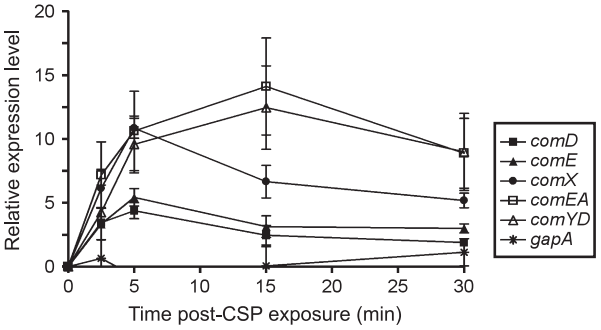

Figure 3 Examination of the expression of selected competence-related genes by quantitative reverse transcription–polymerase chain reaction. Values are presented in units of $\log_2$ of the ratio of each gene's normalized expression relative to that at T_0. Samples were examined in triplicate in two or three separate experiments. Error bars indicate SEM.

Conserved upstream regulatory sequences

The simplest model for CSP-dependent upregulation of early genes entails direct binding of phosphorylated ComE to upstream regulatory sites. By aligning the sequences upstream from the *comC* genes of *S. pneumoniae*, *S. mitis*, *S. oralis*, *S. cristatus*, *S. gordonii* and *S. sanguinis*, Ween *et al.* (1999) identified a consensus motif consisting of two approximately 9-bp imperfect direct repeats, which was predicted to serve as a ComE binding site. This pattern is consistent with known binding sites for transcriptional regulators in the AlgR/AgrA/LytR family, to which ComE belongs (Nikolskaya & Galperin, 2002).

To identify possible ComE binding sites in *S. sanguinis*, the alignment produced by Ween *et al.* (1999) from the *comC*-proximal sequences of six mitis group species was used to create a PSSM, which was then searched against an *S. sanguinis* SK36 nucleotide database, as described in the Methods section. Statistically significant matches were found upstream of three genes: SSA_2394 (*comC*); SSA_1889; and SSA_0195 (Fig. 4A). The first two *S. sanguinis* genes were identified as early response genes by the microarray analysis (Fig. 2), and the third by qRT-PCR. The SSA_1889 match overlapped the gene's start codon; however, use of an alternative ATG start codon 78 bp downstream from that annotated results in a match that is similarly spaced from a consensus −10 element (Fig. 4A) and start codon (not shown). This start codon results in a protein with an N terminus similar or identical to that of its two closest homologs in GenBank, as determined by BLASTP analysis (data not shown), suggesting its authenticity.

A second PSSM was created from the preceding three matches to allow for the possibility of *S. sanguinis*-specific sequence variations. The only additional significant match occurred upstream from SSA_0016, encoding *comX* – the final early response gene identified in our analysis. Interestingly, however, the site was farther upstream than the other genes and in the reverse orientation (Fig. 4A). Examination of the *comX* upstream sequence in the positive orientation uncovered an alignment that had less similarity, but was appropriately spaced relative to a −10 sequence, as shown in Fig. 4(A). Inclusion of the

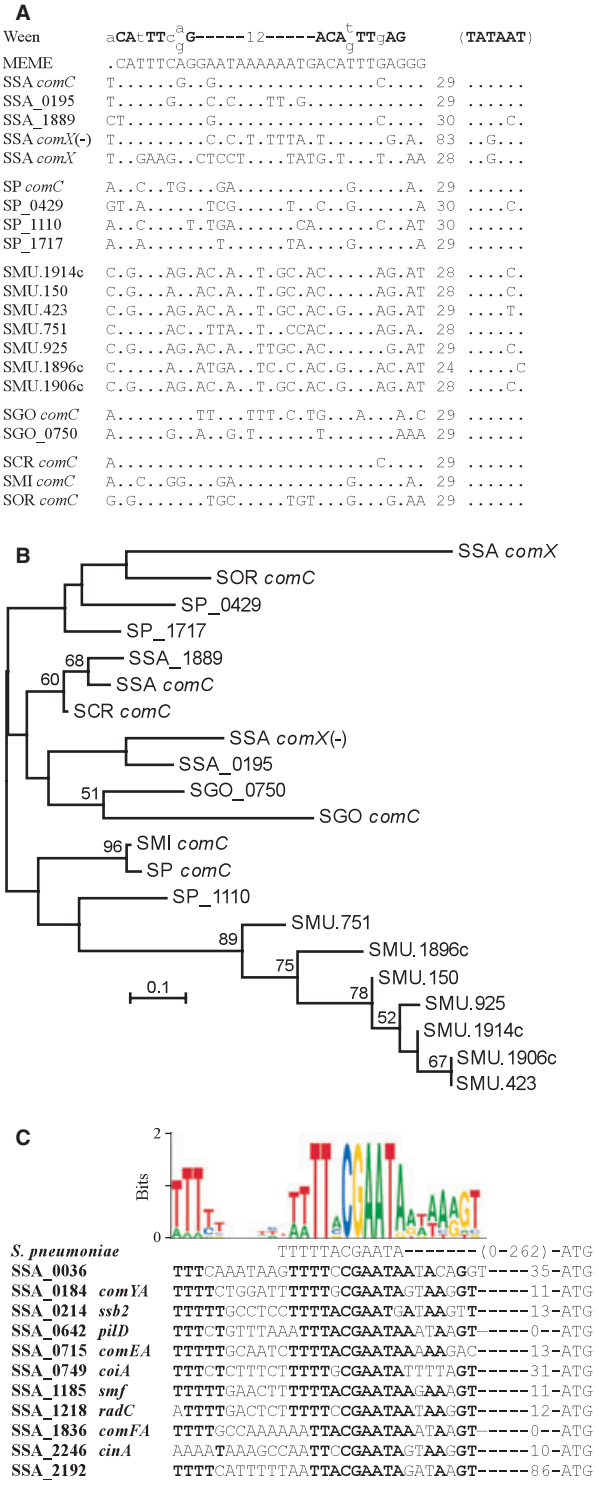

Figure 4 Examination of putative upstream competence regulatory sites. Genes are indicated by name and/or locus tag. Streptococcal species included are: SSA, *S. sanguinis*; SP, *S. pneumoniae*; SMU, *S. mutans*; SGO, *S. gordonii*; SCR, *S. cristatus*; SMI, *S. mitis*; and SOR, *S. oralis*. (A) Alignment of putative ComE-binding sites and −10 promoter elements. The motif identified by Ween *et al.* (1999) is indicated for reference; MEME, a simplified representation of the position-specific scoring matrix (PSSM) created by the MEME program. Dots indicate identity to the PSSM or conserved −10 sequences; differences are shown. (B) Phylogenetic tree of the putative ComE binding sites shown in A (without −10 elements). Bootstrap values ≥50% from 2000 replicates are indicated next to the branches. Scale indicates the number of base substitutions per site. (C) Alignment of putative *S. sanguinis* combox sequences. Conservation is indicated by the sequence logo, with the overall height of each stack indicating the sequence conservation at that position, as measured in bits, and the height of each symbol reflecting the relative frequency of the corresponding nucleotide (Crooks *et al.*, 2004). Nucleotides present in more than half of the aligned sequences are indicated in bold. Numbers indicate distances to the downstream start codon. The *S. pneumoniae* consensus combox sequence (Peterson *et al.*, 2004) is shown for comparison.

comX sequence in either orientation for another round of PSSM creation and searching resulted in no additional significant matches. *Streptococcus pneumoniae* (Peterson *et al.*, 2004), *S. gordonii* (Vickerman *et al.*, 2007) and *S. mutans* (van der Ploeg, 2005) have also been examined for ComE-binding sites using different methods. To better compare the four species, the same iterative PSSM strategy used above was applied to the other three species, except that the site between the *S. mutans comC* gene (SMU.1915) and SMU.1914c was also included in the PSSM. Significant matches from these species, along with the *Streptococcus cricetus*, *S. mitis* and *S. oralis comC* sites, are indicated in Fig. 4(A). The degree of sequence divergence among the sites is depicted in Fig. 4(B).

Previous studies have also identified a conserved 'combox' or 'cinbox' site upstream from a number of late competence genes, which is thought to serve as the binding site for the competence-specific sigma factor – ComX in *S. pneumoniae* (Campbell *et al.*, 1998; Peterson *et al.*, 2004) and *S. mutans* (Okinaga *et al.*, 2010) or ComR in *S. gordonii* (Vickerman *et al.*, 2007). Combox sequences were readily detected preceding 19 late genes or gene clusters in *S. pneumoniae* TIGR4 (Peterson *et al.*, 2004), 12 in *S. gordonii* (Vickerman *et al.*, 2007) and 15 in *S. mutans* (Okinaga *et al.*, 2010). Nine of these sites were observed in all three studies, and a 10th was shared in the last two. The same 10 sites were identified in *S. sanguinis*, as shown in Fig. 4(C). An 11th site was observed upstream of SSA_2192, an *S. sanguinis* late gene with no orthologs in *S. pneumoniae* or *S. mutans*. Further examination using a variety of strategies yielded no further matches to the combox in the *S. sanguinis* genome.

DISCUSSION

This study has characterized the transcriptional response associated with competence induction in *S. sanguinis*, providing an unprecedented comparison of streptococcal competence regulatory systems. The magnitude of the overall transcriptional response to CSP in *S. sanguinis* – 122 genes upregulated, and 83 downregulated – was similar to that seen previously in other streptococci. One *S. pneumoniae* study identified 124 and 64 upregulated and downregulated genes, respectively (Peterson *et al.*, 2004), whereas the corresponding numbers for another were 105 and 42 (Dagkessamanskaia *et al.*, 2004). In *S. gordonii*, 162 and 89 were upregulated and downregulated (Vickerman *et al.*, 2007), and 202 and 75 were similarly classified in *S. mutans* (Perry *et al.*, 2009). In addition, the genes identified as downregulated or upregulated late in *S. sanguinis* were similar to those identified in previous studies with regard to general function.

Given this result, it was surprising to see the sparseness of the early CSP-induced transcriptional response in *S. sanguinis* relative to *S. pneumoniae* and *S. gordonii*. There are a number of possible explanations for this. One possibility is that this difference stems from different growth conditions employed in the studies. Previous studies identifying early genes in *S. pneumoniae* (Peterson *et al.*, 2004) and *S. gordonii* (Vickerman *et al.*, 2007) employed growth in acidic conditions to prevent natural competence induction whereas we used a *comC* deletion mutant and growth at neutral pH. However, an identical strategy to ours was used previously in another *S. pneumoniae* study (Dagkessamanskaia *et al.*, 2004), and although genes were not categorized as early or late or examined at 2.5 min, 16 of the 23 previously identified early genes (Peterson *et al.*, 2004) were significantly upregulated at 5 min, suggesting a similar early response. Our use of CSP at 220 ng ml^{-1} is also comparable to the CSP concentrations of 100 ng ml^{-1} for *S. gordonii* (Vickerman *et al.*, 2007) and 200 ng ml^{-1} (Peterson *et al.*, 2004) or 500 ng ml^{-1} (Dagkessamanskaia *et al.*, 2004) for *S. pneumoniae*, suggesting that this was not a factor. Another possibility is that this discrepancy results from differences in statistical procedures employed in the different studies. This again seems unlikely. Examination of the raw data revealed no evidence of induction of any additional early genes, apart from those already indicated in Table 2. Moreover, the set of early genes identified in *S. sanguinis* coincides almost exactly with the intersection of early genes shared by *S. pneumoniae* and *S. gordonii*. This suggests that the *S. sanguinis* early response is confined almost entirely to genes required for competence induction.

It was also interesting that our bioinformatic analysis detected a putative ComE-binding site upstream from every early gene operon identified by expression analysis, and at no other sites in the genome. Peterson *et al.* (2004) identified 13 gene clusters in

S. pneumoniae as belonging to the early response by microarray analysis. A ComE motif was noted upstream from nine of these, suggesting both direct and indirect control of early gene expression by ComE. In contrast, Vickerman *et al.* (2007) failed to detect a ComE motif upstream from any of the 18 transcriptional units identified by microarray analysis as belonging to the early response in *S. gordonii*. ComE binding sites were not examined in the *S. mutans* microarray analysis (Perry *et al.*, 2009), although such sites have been noted previously upstream from five bacteriocin genes (van der Ploeg, 2005). When we applied the same bioinformatic strategy to the other three species, four significant matches were detected in *S. pneumoniae*; three coincided with ComE sites noted previously (Peterson *et al.*, 2004), but the fourth occurred upstream from SP_0429, which was also classified as an early gene by microarray analysis (Peterson *et al.*, 2004). We detected all five bacteriocin genes noted previously in *S. mutans* (van der Ploeg, 2005), one site upstream from the bacteriocin gene SMU.1896c, and a non-bacteriocin gene upregulated 1.74-fold by CSP, SMU.751 (Perry *et al.*, 2009). SMU.1896c was not identified as CSP regulated by microarray analysis (Perry *et al.*, 2009), and a transcriptional reporter analysis has indicated that SMU.1896c expression is not affected by ComE or CSP (van der Ploeg, 2005). Although a – 10 element was not part of the PSSM used for these analyses, the alignment in Fig. 4(A) indicates that a highly conserved – 10 element is located 28–30 bp upstream from every gene in all four species except SMU.1896c. The absence of a consensus – 10 element for SMU.1896c suggests an explanation for its lack of induction by ComE and CSP, as well as supporting the specificity of the PSSM.

Only two potential ComE-binding sites were identified by the PSSM in *S. gordonii* (Fig. 4A), both of which preceded genes identified as belonging to the early response (Vickerman *et al.*, 2007). The phylogenetic analysis in Fig. 4(B) indicates that these sites are more closely related to the other mitis group sequences than are most of the *S. mutans* sites. It is therefore remarkable that so few sites were identified upstream from *S. gordonii* early genes. This suggests that there is either more similarity in ComE binding sites among species than among sites within this single species, or that there are very few early genes regulated directly by ComE in *S. gordonii*. Our findings with *S. sanguinis* suggest more divergence in mechanisms of early gene regulation within the mitis group than previously envisioned.

It has recently been reported that *comX* expression and transformation are upregulated by a stand-alone transcriptional regulator, termed ComR, and a small peptide, named ComS in *Streptococcus thermophilus* and *Streptococcus salivarius* (Fontaine *et al.*, 2010) and by an orthologous system found in *S. mutans* and the pyogenic and bovis group streptococci (Mashburn-Warren *et al.*, 2010). Related consensus sequences identified upstream from *comX* were predicted to serve as binding sites for ComR in the two orthologous systems (Mashburn-Warren *et al.*, 2010). Neither of the consensus sequences were identified by the authors upstream of *comX* in *S. sanguinis* or other mitis group species examined (Mashburn-Warren *et al.*, 2010). We confirmed these findings for *S. sanguinis* using an identical analysis, as well as a PSSM analysis similar to that described earlier for detection of ComE and ComX binding sites. It would therefore seem unlikely that *comX* expression is controlled by ComRS orthologs in *S. sanguinis* or the other mitis group species examined to date.

One goal of this study was to identify the CSP transporter(s) in SK36. In *S. pneumoniae*, early induction of the *comAB* genes encoding the CSP transporter is important for competence development, because CSP transport appears to be rate-limiting for transformation (Martin *et al.*, 2000). Although ComAB orthologs are clearly not encoded in the *S. sanguinis* genome (Xu *et al.*, 2007), we reasoned that there might exist early-induced genes encoding ComAB analogs that would serve to export *S. sanguinis* CSP. SSA_0195 and SSA_1889 are obvious candidates. SSA_0195 is annotated as a hypothetical protein with weak homology to a putative lipoprotein in *Streptococcus* sp. M143 (accession no ZP_01835889.1). It also possesses weak similarity to the *comM* gene of *S. pneumoniae* (spr1762 or SP_1945), which has been shown to be an early CSP-induced gene that provides immunity to pneumococcal lysins (Håvarstein *et al.*, 2006). SSA_1889 is annotated as a conserved uncharacterized protein (Xu *et al.*, 2007) whose closest match in GenBank is another conserved hypothetical protein in *S. oralis* (ZP_06611623.1). It had no mutually best BLASTP matches with any other sequences in GenBank. Both SSA_0195 and SSA_1889 have weak homology to

CAAX N-terminal protease family proteins (Pei & Grishin, 2001), although the two proteins share no significant similarity. To determine whether either gene played a role in CSP processing and transport, these genes were deleted individually and together by allelic exchange mutagenesis. None of the three mutants exhibited a greater than two-fold reduction in transformation efficiency in our standard assay, in the presence or absence of exogenous CSP (data not shown). It therefore appears that neither gene encodes a CSP transporter nor do they have any other essential role in transformation. Other potential roles of these genes are under investigation.

The preceding results suggest that *S. sanguinis* CSP may be processed and exported by a *com*-independent transporter with a dual function. A logical candidate would be a bacteriocin transporter. A large number of bacteriocin-like peptides are similar to non-*S. sanguinis* ComC sequences in possessing a Gly-Gly motif, and are processed and transported similarly to CSP by ComAB homologs (Håvarstein *et al.*, 1995a). The genes originally identified as encoding the *S. mutans* CSP transporter were later shown to serve as a bacteriocin transporter (Hale *et al.*, 2005), and the ComAB proteins of *S. gordonii* have been shown to process and secrete both CSP and bacteriocin-like peptides (Heng *et al.*, 2007). Moreover, bacteriocin production is closely tied to competence induction in *S. pneumoniae* (Peterson *et al.*, 2004; Martin *et al.*, 2006), *S. gordonii* (Lunsford & London, 1996; Vickerman *et al.*, 2007) and *S. mutans* (Hale *et al.*, 2005; van der Ploeg, 2005; Perry *et al.*, 2009). However, *S. sanguinis* differs from these other three species in lacking not only a Gly-Gly-containing ComC protein, but also any recognizable bacteriocin genes. Early reports of bacteriocin production by *S. sanguinis* employed isolates that are currently classified as *S. gordonii* (Schlegel & Slade, 1974) or are of uncertain taxonomy (Fujimura & Nakamura, 1979), and a database mining program (de Jong *et al.*, 2006) failed to identify any classical bacteriocin candidates in the SK36 genome (data not shown). If *S. sanguinis* does encode bacteriocins, it is possible that the unidentified ComC transporter is also responsible for their transport. It is also possible that *S. sanguinis* CSP undergoes *sec*-dependent secretion, although no recognizable signal sequence is present (data not shown). If CSP is secreted by a housekeeping transporter, the *comCDE* and *comX* genes of *S. sanguinis* may represent the minimal dedicated system required for CSP regulation of competence gene expression. Studies are currently underway to identify the CSP transporter in *S. sanguinis*.

ACKNOWLEDGEMENTS

We thank Nicai Zollar for technical assistance. Microarray slides were provided by the joint National Institute of Dental and Craniofacial Research Oral Microbe Microarray Initiative and National Institute of Allergy and Infectious Diseases Pathogen Functional Genomics Resource Center. We thank Don Morrison for providing advice on RNA isolation and microarray procedures, and Meg Vickerman and Celine Lévesque for providing additional data to facilitate inter-species comparisons. This work was supported by a grant from the Hazel Thorpe and George Gay Carman Trust to T.K., and grants from the National Institute of Dental and Craniofacial Research to J.E.C. (F31DE020963) and P.X. (R01DE18138).

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.

Ahn, S.-J., Wen, Z.T. and Burne, R.A. (2006) Multilevel control of competence development and stress tolerance in *Streptococcus mutans* UA159. *Infect Immun* **74**: 1631–1642.

Alloing, G., Martin, B., Granadel, C. and Claverys, J.-P. (1998) Development of competence in *Streptococcus pneumoniae*: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. *Mol Microbiol* **29**: 75–83.

Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28–36.

Bailey, T.L. and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48–54.

Campbell, E.A., Choi, S.Y. and Masure, H.R. (1998) A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Mol Microbiol* **27**: 929–939.

Carlsson, J. (1965) Zooglea-forming streptococci, resembling *Streptococcus sanguis*, isolated from dental plaque in man. *Odontol Revy* **16**: 348–358.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190.

Dagkessamanskaia, A., Moscoso, M., Hénard, V. *et al.* (2004) Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of *ciaR* mutant cells. *Mol Microbiol* **51**: 1071–1086.

Di Filippo, S., Delahaye, F., Semiond, B. *et al.* (2006) Current patterns of infective endocarditis in congenital heart disease. *Heart* **92**: 1490–1495.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868.

Fontaine, L., Boutry, C., de Frahan, M.H. *et al.* (2010) A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J Bacteriol* **192**: 1444–1454.

Fujimura, S. and Nakamura, T. (1979) Sanguicin, a bacteriocin of oral *Streptococcus sanguis*. *Antimicrob Agents Chemother* **16**: 262–265.

Gaustad, P. and Håvarstein, L.S. (1997) Competence-pheromone in *Streptococcus sanguis*. Identification of the competence gene comC and the competence pheromone. *Adv Exp Med Biol* **418**: 1019–1021.

Hale, J.D.F., Heng, N.C.K., Jack, R.W. and Tagg, J.R. (2005) Identification of *nlmTE*, the locus encoding the ABC transport system required for export of nonlantibiotic mutacins in *Streptococcus mutans*. *J Bacteriol* **187**: 5036–5039.

Håvarstein, L., Diep, D. and Nes, I. (1995a) A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol* **16**: 229–240.

Håvarstein, L.S., Coomaraswamy, G. and Morrison, D.A. (1995b) An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci USA* **92**: 11140–11144.

Håvarstein, L.S., Gaustad, P., Nes, I.F. and Morrison, D.A. (1996) Identification of the streptococcal competence-pheromone receptor. *Mol Microbiol* **21**: 863–869.

Håvarstein, L.S., Hakenbeck, R. and Gaustad, P. (1997) Natural competence in the genus *Streptococcus*: evidence that streptococci can change pherotype by interspecies recombinational exchanges. *J Bacteriol* **179**: 6589–6594.

Håvarstein, L.S., Martin, B., Johnsborg, O., Granadel, C. and Claverys, J.-P. (2006) New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. *Mol Microbiol* **59**: 1297–1307.

Heng, N.C.K., Tagg, J.R. and Tompkins, G.R. (2007) Competence-dependent bacteriocin production by *Streptococcus gordonii* DL1 (Challis). *J Bacteriol* **189**: 1468–1472.

Hillman, J.D., Socransky, S.S. and Shivers, M. (1985) The relationships between streptococcal species and periodontopathic bacteria in human dental plaque. *Arch Oral Biol* **30**: 791–795.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**: 51–59.

Hui, F.M., Zhou, L. and Morrison, D.A. (1995) Competence for genetic transformation in *Streptococcus pneumoniae*: organization of a regulatory locus with homology to two lactococcin A secretion genes. *Gene* **153**: 25–31.

Ishii, S., Yano, T. and Hayashi, H. (2006) Expression and characterization of the peptidase domain of *Streptococcus pneumoniae* ComA, a bifunctional ATP-binding cassette transporter involved in quorum sensing pathway. *J Biol Chem* **281**: 4726–4731.

Johnsborg, O. and Håvarstein, L.S. (2009) Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. *FEMS Microbiol Rev* **33**: 627–642.

de Jong, A., van Hijum, S.A.F.T., Bijlsma, J.J.E., Kok, J. and Kuipers, O.P. (2006) BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res* **34**: W273–W279.

Kreth, J., Merritt, J., Shi, W. and Qi, F. (2005) Co-ordinated bacteriocin production and competence development: a possible mechanism for taking up DNA from neighbouring species. *Mol Microbiol* **57**: 392–404.

Lee, M.S. and Morrison, D.A. (1999) Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. *J Bacteriol* **181**: 5004–5016.

Loo, C.Y., Mitrakul, K., Voss, I.B., Hughes, C.V. and Ganeshkumar, N. (2003) Involvement of an inducible fructose phosphotransferase operon in *Streptococcus gordonii* biofilm formation. *J Bacteriol* **185**: 6241–6254.

Lunsford, R.D. (1998) Streptococcal transformation: essential features and applications of a natural gene exchange system. *Plasmid* **39**: 10–20.

Lunsford, R.D. and London, J. (1996) Natural genetic transformation in *Streptococcus gordonii*: *comX* imparts spontaneous competence on strain wicky. *J Bacteriol* **178**: 5831–5835.

Martin, B., Prudhomme, M., Alloing, G., Granadel, C. and Claverys, J.-P. (2000) Cross-regulation of competence pheromone production and export in the early control of transformation in *Streptococcus pneumoniae*. *Mol Microbiol* **38**: 867–878.

Martin, B., Quentin, Y., Fichant, G. and Claverys, J.-P. (2006) Independent evolution of competence regulatory cascades in streptococci? *Trends Microbiol* **14**: 339–345.

Mashburn-Warren, L., Morrison, D.A. and Federle, M.J. (2010) A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Mol Microbiol* **78**: 589–606.

Merritt, J., Qi, F. and Shi, W. (2005) A unique nine-gene *comY* operon in *Streptococcus mutans*. *Microbiology* **151**: 157–166.

Nikolskaya, A.N. and Galperin, M.Y. (2002) A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res* **30**: 2453–2459.

Okinaga, T., Xie, Z., Niu, G., Qi, F. and Merritt, J. (2010) Examination of the *hdrRM* regulon yields insight into the competence system of *Streptococcus mutans*. *Mol Oral Microbiol* **25**: 165–177.

Paik, S., Senty, L., Das, S. *et al.* (2005) Identification of virulence determinants for endocarditis in *Streptococcus sanguinis* by signature-tagged mutagenesis. *Infect Immun* **73**: 6064–6074.

Pakula, R. and Walczak, W. (1963) On the nature of competence of transformable streptococci. *J Gen Microbiol* **31**: 125–133.

Pei, J. and Grishin, N.V. (2001) Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends Biochem Sci* **26**: 275.

Perry, D. and Kuramitsu, H.K. (1981) Genetic transformation of *Streptococcus mutans*. *Infect Immun* **32**: 1295–1297.

Perry, J.A., Jones, M.B., Peterson, S.N., Cvitkovitch, D.G. and Lévesque, C.M. (2009) Peptide alarmone signalling triggers an auto-active bacteriocin necessary for genetic competence. *Mol Microbiol* **72**: 905–917.

Pestova, E.V., Håvarstein, L.S. and Morrison, D.A. (1996) Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol* **21**: 853.

Peterson, S., Cline, R.T., Tettelin, H., Sharov, V. and Morrison, D.A. (2000) Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. *J Bacteriol* **182**: 6192–6202.

Peterson, S.N., Sung, C.K., Cline, R. *et al.* (2004) Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol Microbiol* **51**: 1051–1070.

van der Ploeg, J.R. (2005) Regulation of bacteriocin production in *Streptococcus mutans* by the quorum-sensing system required for development of genetic competence. *J Bacteriol* **187**: 3980–3989.

Rimini, R., Jansson, B., Feger, G. *et al.* (2000) Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. *Mol Microbiol* **36**: 1279–1292.

Schlegel, R. and Slade, H.D. (1974) Alteration of macromolecular synthesis and membrane permeability by a *Streptococcus sanguis* bacteriocin. *J Gen Microbiol* **81**: 275–277.

Stingu, C.-S., Eschrich, K., Rodloff, A.C., Schaumann, R. and Jentsch, H. (2008) Periodontitis is associated with a loss of colonization by *Streptococcus sanguinis*. *J Med Microbiol* **57**: 495–499.

Sung, C.K. and Morrison, D.A. (2005) Two distinct functions of ComW in stabilization and activation of the alternative sigma factor ComX in *Streptococcus pneumoniae*. *J Bacteriol* **187**: 3052–3061.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V. *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**: 22–28.

Turner, L.S., Das, S., Kanamoto, T., Munro, C.L. and Kitten, T. (2009) Development of genetic tools for in vivo virulence analysis of *Streptococcus sanguinis*. *Microbiology* **155**: 2573–2582.

Vickerman, M.M., Iobst, S., Jesionowski, A.M. and Gill, S.R. (2007) Genome-wide transcriptional changes in *Streptococcus gordonii* in response to competence signaling peptide. *J Bacteriol* **189**: 7799–7807.

Ween, O., Gaustad, P. and Havarstein, L.S. (1999) Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Mol Microbiol* **33**: 817–827.

Wen, Z.T., Browngardt, C. and Burne, R.A. (2001) Characterization of two operons that encode components of

fructose-specific enzyme II of the sugar: phosphotransferase system of *Streptococcus mutans*. *FEMS Microbiol Lett* **205**: 337–342.

Xu, P., Ge, X., Chen, L., Dou, Y., Wang, X., Patel, J.R., Trinh, M., Evans, K., Stone, V., Kitten, T., Bonchev, D., and Buck, G. (2007) Genome of the opportunistic pathogen *Streptococcus sanguinis*. *J Bacteriol* **189**: 3166–3175.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table S1. Microarray expression data in xls format.

Table S2. Comparison of late genes and late gene orthologs in *Streptococcus sanguinis*, *Streptococcus pneumoniae*, *Streptococcus gordonii* and *Streptococcus mutans* in xls format.

Appendix S1. Microarray Methods in pdf format.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.