Subpopulation-specific transcriptome analysis of competence-stimulating-peptide-induced Streptococcus mutans
Subpopulation specific transcriptome analysis of CSP induced *Streptococcus mutans*

Running title: Competence development in *Streptococcus mutans*

André Lemme*, Lothar Gröbe, Michael Reck, Jürgen Tomasz and Irene Wagner-Döbler

Helmholtz Centre for Infection Research, Research Group Microbial Communication, Inhoffenstr. 7, D-38124 Braunschweig, Germany

* Corresponding author: André Lemme, Helmholtz-Centre for Infection Research, Research Group Microbial Communication, Inhoffenstr. 7, D-38124 Braunschweig, Germany. Phone: +49 (0)531 6181 3082 Fax: +49 (0)531 6181 3096 Email: Andre.Lemme@helmholtz-hzi.de
Abstract

CSP mediated competence development in *Streptococcus mutans* is a transient and biphasic process, since only a subpopulation induces expression of ComX in the presence of CSP and activation of the DNA uptake machinery in this fraction shuts down ~3-4 hours post induction. Here we combine, to our knowledge, for the first time in bacteria flow cytometric sorting of cells and subpopulation specific transcriptome analysis of both the competent and non-competent fraction of CSP treated *S. mutans* cells. Sorting was guided by a ComX-GFP reporter and the transcriptome analysis demonstrated the successful combination of both methods because a strong enrichment of transcripts for *comX* and its downstream genes was achieved. Three two component systems were expressed in the competent fraction, among them ComDE. Moreover, the recently identified regulator system ComR/S was expressed exclusively in the competent fraction. By contrast, expression of bacteriocin related genes was at the same level in all cells. GFP reporter strains for ComE and CipB (mutacin V) confirmed this expression pattern on the single cell level. Fluorescence microscopy revealed that some ComX expressing cells committed autolysis in an early stage of competence initiation. In viable ComX expressing cells uptake of DNA could be shown on the single cell level. This study demonstrates that all cells in the population respond to CSP through activation of bacteriocin related genes. Some of these cells start to activate ComX expression but then again segregate into two subpopulations, one becoming competent and another one that lyses, resulting in intra population diversity.
Introduction

Competence development is a complex process involving complex regulatory networks that trigger the capacity to take up exogenous DNA from the environment (7, 8, 11, 16). This phenomenon occurs in gram negative and gram positive bacteria and is frequently encountered in bacteria of the oral cavity, e.g. *Streptococcus mutans* (39). *S. mutans* is considered the major etiological agent of dental caries because it is associated with the initiation and progression of dental caries. So far, development of genetic competence is best characterized in *Bacillus subtilis* and the human pathogen *Streptococcus pneumoniae* and differs significantly between these two species (8, 11, 16).

At the onset of the stationary phase, cells of *B. subtilis* differentiate into several distinct cell types, e.g. ~10% of the cells initiate competence development. This phenomenon of phenotypic variation of a clonal population has been referred to as “bistability”, due to the dual stable pattern of gene expression in genetically identical cells (12).

According to Ferrell, two mechanisms exist to manifest bistability in a bacterial population (13). One is the principle mechanism of competence initiation in *B. subtilis* and is described below. It requires the presence of a positively auto regulated transcriptional activator, which responds to itself in a non-linear manner. Specifically, above a certain threshold of the regulator a hypersensitive change in gene expression is induced due to the activation of a positive auto feedback loop. Consequently, this results in high levels of the regulator, and its controlled genes will be activated (2). Cells that do not exceed the threshold will remain inactive, resulting in segregation of the clonal population. In *B. subtilis*, ComK fulfils these criteria and induces the expression of all genes in the competence pathway in a subpopulation. Whether a cell exceeds the threshold of the regulator concentration to activate the positive feedback loop is determined by noise, or random fluctuations, in the expression of the *comK* gene. In *B. subtilis*, this can be artificially manipulated by inducing the expression of *comK* or
by promoting the stabilization of ComK. Due to the non-linear autoregulatory properties of
ComK, bistability will remain but the number of competence initiating cells can be varied in
that way (2).
So far, there is no evidence for bistability in competence development in *S. pneumoniae*. With
laboratory strains, routinely ~100% of competent cell could be obtained (6).
During growth, *S. pneumoniae* releases the competence stimulating peptide CSP (encoded by
*comC*), using the ComAB secretion apparatus. Interaction of CSP with the membrane
embedded histidine kinase ComD results in activation of the cognate response regulator
ComE through phosphoryl transfer. Activation of the ComDE two component system
leads to the expression of ~20 early competence genes (5, 15, 34, 42). Coordinated induction
of competence in the whole population is achieved by a positive feedback loop, since ComE
activates expression of the *comAB* and *comCDE* operons. However, this feedback loop in this
case does not lead to bifurcation of the population. Martin *et al.* discovered that
transcriptional read through of the tRNA<sup>Arg5</sup> gene upstream of *comCDE* was crucial to
maintain a sufficient level of ComDE, ensuring that all cells can initiate competence when the
threshold of CSP is exceeded. Inactivation of this transcriptional read through resulted in
segregation of the culture and only 25% started expression of late competence genes (24).
Among the early competence genes is the alternate sigma factor ComX. ComX is required for
the induction of the late competence genes, which includes e.g. the genes necessary for DNA
uptake and processing.
*S. pneumoniae* contains a paralogue of the ComABCDE system, designated BlpABCRH.
BlpRH constitutes a two component system which is activated by the auto-inducing peptide
BlpC (similar to CSP and ComDE). BlpAB was shown to be necessary for the transport of
BlpC. The BlpR regulon was shown to comprise several bacteriocin like peptides (10, 37).
The ComCDE and ComAB system (transport of CSP) of *S. mutans* is more related to the Blp system of *S. pneumoniae*, than to ComABCDE, based on genomic organization and blast identity (25). Interestingly, in *S. mutans* the ComCDE system combines the action of the two orthologues in *S. pneumoniae*. Activation of ComE, through CSP and its receptor kinase ComD, leads to induction of competence through the alternate sigma factor ComX and at the same time ComE directly induces a set of bacteriocin related genes (Fig. 1) (1, 17-19, 23, 41). Interestingly, disruption of *comE* in *S. pneumoniae* completely abolishes transformation, suggesting that ComE is exclusively responsible for ComX activation. Deletion of *comE* in *S. mutans* affects only the CSP inducible competence initiation, a basal level of competence remains (referred as to CSP independent competence), suggesting that the CSP/ComDE system is one of several signalling pathways used to activate ComX (1).

Indeed, under conditions of biofilm growth the HdrRM system was shown to contribute to competence development through activation of ComX by a yet unknown signal (Fig. 1) (29). Moreover, microarray analysis revealed that both regulators, ComE and HdrR, activate a large set of overlapping genes. (29, 30). Recently, Xie et al. identified another regulatory system, designated BsrRM, that primarily regulates bacteriocin related genes but also affects the HdrMR system and thus indirectly contributes to competence development (44) (Fig. 1).

In *S. mutans*, no binding motif for ComE is present in the promoter region of ComX, suggesting a missing link between both regulators and ComX. This gap was filled by a recent study, which identified a new peptide regulator system (ComS/R) that acts downstream of ComE and activates directly ComX (26). Deletion of ComR completely abolished competence, suggesting that signals from all regulatory circuits (CSP dependent and independent) are integrated at the level of ComR. ComR activates expression of the peptide precursor ComS. ComS is secreted, processed and internalised through the peptide transporter Opp. The processed peptide, designated XIP (sigX-inducing peptide), modulates the activity of ComR which activates expression of ComS and ComX (Fig. 1).
Using GFP reporter constructs it was shown that in *S. mutans* after addition of CSP mutacin IV was expressed population wide, but that expression of ComX was restricted to some cells, suggesting a bistable switch downstream of bacteriocin induction (18, 33). The scope of this study was to determine the CSP dependent activation of competence on the single cell level. Therefore we used flow cytometry to separate competent from non-competent cells, guided by a ComX-GFP reporter. RNA of both subpopulations was subjected to microarray analysis to disclose transcriptional changes. To our knowledge, such a combination of flow cytometry and microarray analysis has never been carried out before in bacteria. Data from GFP reporter strains for ComE and CipB (mutacin V) and for DNA uptake on the single cell level indicate another bifurcation within the ComX expressing subpopulation into cells committing autolysis and cells that develop competence. Our data provide a first view on the segregation of a CSP induced clonal population of *S. mutans* into three phenotypically distinct subpopulations and show the level within the competence signalling cascade where it occurs and thus expand our knowledge for competence development in *S. mutans*. 
**Materials and Methods**

**Bacterial strains, plasmids, and culture conditions**

Bacterial strains and plasmids and their relevant characteristics are listed in table 1.

*Escherichia coli* was routinely cultured in Luria Bertani (LB, Carl-Roth, Karlsruhe, Germany) medium at 37°C. *E. coli* strains carrying plasmids were selected with 50 µg ml⁻¹ spectinomycin, 200 µg ml⁻¹ erythromycin, 20 µg ml⁻¹ chloramphenicol, or 20 µg ml⁻¹ tetracycline. All *S. mutans* strains were cultivated in Todd Hewitt Broth medium supplemented with 0.5% (w/v) yeast extract (THBY, Becton Dickinson, Heidelberg, Germany). *S. mutans* strains were grown at 37°C without agitation aerobically (5% CO₂ enriched). Selection of mutant strains was carried out with 10 µg ml⁻¹ or 4 µg ml⁻¹ (pALSM derivates) erythromycin and 500 µg ml⁻¹ spectinomycin.

The CSP peptide was synthesized in the department of “Chemical Biology” at the Helmholtz-Centre for Infection Research. The peptide was purified by HPLC and solved in water to a concentration of 1 mM. The sequence is SGSLSTFFRLFNRSFTQALGK and the peptide was synthesized having a free N-terminus and a carboxylic group at the C-terminus.

**Construction of reporter plasmids and strains**

For the construction of pALEC50 and pALEC52, the suicide vector pALEC15 (21) was used as backbone. Upstream regions of *comE* (SMU.1917) and *comX* (SMU.1997) were amplified using primers PcomEF/R and PcomXF/R digested with NcoI and ligated into pALEC15. The plasmids were transformed into the wildtype of UA159 to generate strains SMLuccomE and SMLuccomX by single cross-over recombination.

For the construction of the GFP fluorescence reporter plasmids, a fragment containing the *recA* promoter from *S. pyogenes* and GFP was released from pVA-EGFP2 (28) by XbaI and religated, yielding pALSM01. The eGFP from pKRC12 was amplified using primer GFPF/R,
cut with XbaI/NheI and ligated to pALSM01, giving pALSM02. Afterwards the vector was digested with PciI and EcoRI, blunted and religated to release the lac promoter, giving pALSM03. Subsequently, the upstream regions of comX (PcomX2F/R), cipB (PcipBF/R) and comE (PcomE2F/R) were amplified and cloned into the EcoRV site. The plasmids were confirmed by sequencing and transformed into the wildtype of UA159.

**Fluorescence microscopy**

Cells were pelleted, washed and resuspended with PBS. For chain disruption, cells were sonicated (20x 0.5 sec pulse, 0.5 sec break, MS72 probe with 10% power; Bandelin Sonoplus HD2200, Berlin, Germany). For imaging an Olympus BX60 microscope, equipped with a colorview II camera and a 100/1.3 oil immersion objective was used. The filter U-MWIBA3 (excitation 460-495 nm, emission 510-550 nm, dichromatic filter 505 nm) from Olympus (Seelze, Germany) was used to visualize eGFP with an exposure time of 5 sec. The filter 20HE (excitation 546 ± 12 nm, emission 607 ± 80 nm, dichromatic filter 560 nm) from Carl-Zeiss (Zaventum, Belgium) was used to visualize Cy3 with an exposure time of 10 sec. The filter U-MWIY2 (excitation 545-580 nm, emission 610IF, dichromatic filter 600 nm) from Olympus was used to visualize PI (propidium iodide) with an exposure time of 1 sec. Images were recorded under the same conditions, using the cellB software from Olympus. For better visualizations on print-outs, brightness and contrast were modified equally for all images using Adobe Photoshop.

**Visualization of DNA uptake**

Overnight cultures were 1:10 diluted and incubated as described above. CSP (0.2 µM) was added after 1h and cultures were incubated further for 2 hours. 100 µl of cells were withdrawn and a ~ 2 kb DNA fragment (PCR amplified), labelled with Cy3 (ULS labelling, see below), was added at a final concentration of 1µg/ml and incubated for 15 min at 37°C. Afterwards,
30 Units of DNase I (Roche, Mannheim, Germany) were added and incubated for 5 min at 37°C. Cells were washed, resuspended in PBS and visualized using fluorescence microscopy.

**Luciferase assay**

Luciferase assays were performed by withdrawing 100 µl culture for measurement of the optical density, using a Victor3 Wallac plate reader and luciferase measurement. Samples were mixed with 3x assay buffer (75 mM tricine, 15 mM MgSO$_4$, 1.5 mM EDTA, 1.5 mM DTT, 900 µM ATP, 3 mg/ml BSA, and 3% (w/v) D-Glucose, pH = 7.8) and incubated 10 min prior to injection of 100 µl D-luciferin (120 µM final concentration) solved in 20 mM tricine (pH 7.8). D-Luciferin (Carl-Roth, Karlsruhe, Germany) was resuspended in 20 mM tricine (pH = 7.8, 1 mg/ml), aliquoted and stored at −70°C until use. Luminescence was recorded for 30 s (Victor Wallac Luminescence reader, Perkin Elmer Life-Sciences) and normalized against the OD$_{620}$ to calculate the relative light units (RLU). All measurements were done at least for two biological replicates.

**FACS analysis and subpopulation sorting**

All media used for flow cytometry were filtered through a 0.22 µm filter. For subpopulation sorting an overnight culture of *S. mutans* SMGFPcomX was diluted 1:10 into fresh medium and incubated at 37°C. After one hour, induction with 0.2 µM CSP was carried out. Cells were further incubated for two hours and sonicated as described for disrupting cell chains for fluorescence microscopy prior to flow cytometry. This procedure was repeated every half hour and every sample was subjected to sorting for 30 minutes. In total 10 individual samples were sorted, yielding a single biological sample. Control cells of a mixed induced and not induced culture were sonicated and kept in RNA protect until the sorting was finished. Flow cytometric sorting was performed on a FACSArta™ II Cell Sorter System (Becton Dickinson). Filtered PBS was used as the sheath fluid. Cells were kept at 37°C during sorting
and were collected directly in RNA protect. Forward and side scatter were set to log with a threshold of 200 on both parameters. Detection of GFP fluorescence was through a 525/50 nm bandpass filter. PMT Voltage was 370 for FSC (forward scatter), 205 for SSC (side scatter), and 467 for GFP. Despite sonication, chains of two or more cells still occurred in the culture.

We used pulse-width (-W) and –area (-A) of FSC and SSC to identify these doublets and chains. When a cell/particle passes through the laser beam, it will generate a signal pulse (signal over time), which has a height (-H), width (-W) and integrated area (-A). Single cells of different sizes will show a constant pulse width and strongly correlated height to area ratio. Doublets and other aggregates will show a similar pulse height, but can be identified by their increased pulse width and area. According to the FSC-W and SSC-W plots, gates were set more stringent to allow sorting of single cells. Not induced cells reached a fluorescence intensity of approx. 800 (arbitrary units), thus cells with a higher value were sorted as GFPplus and cells with a lower value as GFPminus. Before cells were collected in RNA protect, the efficiency of sorting was determined and both subpopulations were subjected again for analysis. The whole experiment was done in duplicate and each replicate was used separately for transcriptome analysis. To measure the GFP intensity of all reporter strains, cells were induced and treated as described above and 100,000 cells were analysed. The median values of GFP positive cells (intensity > 400 units) from two biological replicates was determined.

**RNA extraction**

RNA protect samples (Qiagen, Hilden, Germany) containing sorted or mixed cells were filtered through a 0.22 µm filter (Millipore, Cork, Ireland). The filter was stored at –70°C until RNA extraction. The filter was covered with 1 ml TE buffer (pH 8) containing 2.5 mg lysozyme and 50 U mutanolysin and incubated at room temperature for 25 min. RLT buffer (RNeasy kit, Qiagen) containing sterile, acid washed glass beads (diameter 106 µm) was
added and vortexed for 3 min. Subsequent RNA extraction was carried out using the RNeasy micro kit (Qiagen) according to the manufacturer’s instructions. Genomic DNA was removed using the DNAse I (Qiagen) on column digestion protocol. Absence of genomic DNA was controlled by a standard PCR. The quality of the total RNA was controlled on a denaturing formaldehyde agarose gel.

Microarray

A customized whole genome microarray of *S. mutans* UA159 was used to carry out microarray experiments and was previously described (45). Briefly, the array contained probes for all open reading frames (ORF) and non-coding regions of *S. mutans*. Each ORF and intergenic region was covered with three antisense probes (60 bp in length) in duplicate. Positive and negative controls from Agilent were included in the design.

RNA extraction was carried out as described above and labelling was achieved using the ULS fluorescent labelling kit (Kreatech, Germany). Reference RNA from the not CSP induced mixed culture was labelled using Cy3. The three remaining RNA populations (CSP induced mixed and sorted by flow cytometry) were labelled with Cy5. The degree of labelling was calculated using a web form on the manufacturer’s homepage (http://www.kreatech.com/Portals/kreatech/downloads/labeling/27_DoL%20calculator_28082007.xls). For each condition RNA of two individual biological samples was used. 450 ng of reference RNA were mixed with 450 ng of each Cy5-RNA and subjected to fragmentation according to the protocols provided by Agilent. Hybridisation was carried out at 65°C for 17 hours using the Agilent Hybridisation chamber. Scanning was achieved using the Agilent DNA Microarray scanner. Background correction was performed using the Agilent algorithms and raw data were extracted using the agilent feature extraction software. Data processing was carried out with the Bioconductor – Linear Models for Microarray analysis (LIMMA) package (43) using the R language (http://www.r-project.org). We compared the
expression data of the three CSP induced RNA populations with RNA from the non-CSP induced mixed culture (“fold mixed” refers to induced mixed culture vs. non-induced mixed culture, “GFP+” refers to induced sorted competent subpopulation vs. non-induced mixed culture, “GFP-” refers to induced sorted non-competent subpopulation vs. non-induced mixed culture). Moreover, we compared directly the transcription of the CSP induced sorted competent and non-competent subpopulations, referred to as “GFP+/GFP-“. The P-values for differential expression were adjusted for false-discovery rate using the method by Benjamini and Hochberg (BH) (4). Genes with log2 fold change value > 0.85 and a P-value < 0.05 were selected for further analysis.

**Microarray data accession number.** Microarray data have been deposited at NCBI-GEO under the accession number GSE25284.

**Quantitative real-time RT PCR**

Synthesis of cDNA was carried out with 500 ng of total RNA using the Quantitect Reverse transcription kit (Qiagen) according to the manufacturer’s protocols. Quantitative real-time PCR was performed using the LightCycler 480 system (Roche, Mannheim, Germany) and the reaction mixtures were prepared using the Quantitect SYBR Green PCR Kit (Qiagen).

Changes in the level of gene expression were calculated automatically by the LightCycler 480 software using the ΔΔC_T method. The gyrase A gene (SMU.1114) was used as the housekeeping reference gene. All steps were performed according to the manufacturer’s protocols. RNA extraction and reverse transcription were carried out from two independent biological samples. All quantitative real time PCR measurements were done in duplicate.
Results

Kinetics of gene expression of key competence genes in response to CSP.

To obtain a population wide overview of gene expression of key competence genes, we performed a time series experiment to determine the response in gene expression to the addition of CSP using quantitative real-time PCR. Table 3 lists the genes that were chosen for analysis. Induction of the ComE direct target genes \(cipB/I\) (mutacin V and its cognate immunity protein CipI) reached almost maximal levels already 15 minutes after addition of CSP, showing that low levels of ComE at the time of induction are sufficient to achieve the maximal induction in response to CSP. The regulatory genes \((comE, comS, comR)\) acting upstream of ComX were already induced 15 minutes after induction and achieved their highest expression two hours post CSP addition. By contrast, induction of the alternate sigma factor ComX occurred with a delay being first induced 30 minutes after CSP addition, confirming the observations of Ahn et al. (1). Similar to its upstream regulators expression of \(comX\) was highest two hours post induction.

Expression of \(comE\) and \(comX\) was transiently activated and showed separation into two subpopulations.

To determine the induction pattern of the \(comE\) and \(comX\) genes upon CSP addition, we constructed firefly luciferase reporter plasmids for both genes. Figure 2 shows that both genes were expressed transiently, reaching their highest values around 2-2.5 hours post induction, confirming the qPCR data. The delayed induction of ComX compared to ComE cannot be seen because the first sample for luciferase measurement was taken 30 minutes after CSP addition, where both genes were already induced. Using a ComX-GFP fusion, we showed that expression of ComX was restricted to a part of the culture. This phenotypic variation in ComX expression and the averaged GFP intensity of the whole culture is shown in figure 3.
Throughout all the experiments we observed that approximately 30-50% of the cells were expressing GFP. In comparison to the luciferase reporter it took approximately 30 minutes longer (1h post induction) until the first cells became detectable in the fluorescence microscope after addition of exogenous CSP. Additional 30 minutes for chromophore maturation were needed to obtain a detectable signal in a fluorescence microtitre plate reader. Since mRNA levels were highest after 2h post induction and a suitable amount of detectable GFP had been produced, we decided to separate both subpopulations using flow cytometry between 2-2.5 hours after addition of 0.2 µM CSP.

Both subpopulations could be successfully separated using flow cytometry.

Prior to flow cytometry, the chains of *S. mutans* were disrupted using a sonication treatment. In pre-experiments we tested the ability to disrupt chains by sonication and checked for membrane integrity by staining the cells with propidium iodide (PI, a fluorescence dye staining DNA which does not penetrate intact membranes and is therefore used for detection of damaged/dead cells). No influence of sonication was noticed, e.g. there was no increased number of PI stained cells. Using sonication it was possible to break down the majority of chains to single or doublet cells (Supplementary figure 1).

For calibration of the FACSria™ II, we first analysed sterile uninoculated media to determine the particle background of the media (Fig. 4a). Afterwards, media containing bacteria were used to determine cells/particles of inoculated media and for setting the voltages and gates (Fig. 4b). For discrimination of chains or other aggregates FSC-W and SSC-W were enabled and gating was carried out more stringently. Not induced cells reached a GFP fluorescence intensity of approx. 800 units (Fig. 4c). To ensure not to collect GFPminus cells with a high auto fluorescence, we set the gate for GFPplus cells at this threshold. The GFPminus population was gated below 300 GFP fluorescence units. The fraction between gate P4 and P5 was not collected.
The histogram in figure 4d represents the GFP intensity of all analysed cells of gate P3. A bi-modal distribution of “dark” and GFP expressing cells became clearly apparent. However, with respect to the amount of cells in both fractions, this pattern was quite dynamic, depending on the fitness of the pre-culture (Supplementary figure 2). Analysis of both separated subpopulation is shown in figure 4d. The majority of “dark” or GFP expressing cells could be successfully separated from each other.

Microarray analysis showed an enrichment and depletion pattern of \textit{comX} transcripts in the subpopulations, compared to the mixed population.

The aim of the ComX-GFP guided separation of both subpopulations was to compare their individual gene expression with the transcriptome of induced and non-induced mixed populations, respectively. Below, GFP expressing cells will be designated GFPplus or competent fraction/subpopulation and non GFP expressing cells will be referred to GFPminus or non-competent fractions, respectively.

Sorting resulted in an enrichment (13.8 fold) of \textit{comX} transcription in the GFPplus population, and a depletion (2 fold) in the GFPminus cells compared to the mixed population (3.6 fold) (table 4 and supplementary table 1). The two fold induction of \textit{comX} in the GFPminus population may reflect a possible carryover of GFPplus cells during the sorting process; alternatively, it might have been caused by naturally competent cells (independent of CSP), whose \textit{comX} transcription level was insufficient for detectable GFP fluorescence.

Another possibility would be stress triggered competence induction through the process of sorting, which cannot occur to the mixed populations because they were not subjected to flow cytometry.

Genes that were significantly changed in at least one of the populations are listed in supplementary table 1 and a selection of genes is presented in table 4. Highlights are discussed below. Additionally, a direct comparison of the GFPplus and GFPminus
subpopulation was carried out. The direct comparison of both fractions did not reflect the large increase of gene expression in response to CSP in the competent (GFPplus) fraction in a proper way due to the weak enrichment of comX, its downstream genes and other known regulators (see below) in the GFPminus subpopulation. Thus, the fold change values are considerably smaller compared to the non CSP-induced control, and smaller changes may be overlooked. For better comparison, the values of all four conditions are listed together. Genes that were significantly changed (fold change +/- 1.8, P-value < 0.05) are marked in bold. The results for some genes were confirmed using quantitative real time PCR and are listed in supplementary table 1. The enrichment or depletion of transcripts in the GFPplus or GFPminus population could be confirmed by qPCR. However, there was a considerable difference in the amount of transcripts in both biological samples yielding a high standard deviation. For are better comparison we inserted the individual results of the qPCR for both biological samples in the supplementary table 1.

Three two component systems showed a similar enrichment/depletion pattern as comX. Our microarray analysis revealed that transcripts for the two component system comDE were enriched in the GFPplus fraction, similar to comX. Moreover, we observed an enrichment of two additional two component systems, namely HK/RR 4 (SMU.927-928) and HK/RR 9 (SMU.1964-1965). The induction of these three TCS upon CSP addition is in agreement with a population wide transcriptome analysis carried out by Perry et al. (33). Interestingly, a deletion of HK9 resulted in an decreased transformation efficiency, whereas inactivation of HK4 showed no effect (22). However, the new finding that these TCS are induced in the competent cell fraction suggests that they may take part in the competence development, but their exact role needs further experimental exploration. Using qPCR we confirmed the microarray data for comE and comX (supplementary table 1).
Interestingly, we found several intergenic regions (IGR) that were differentially regulated in both subpopulations independently of their adjacent genes. Table 5 lists all those IGRs that were either differentially expressed in the mixed population, or GFPplus, or GFPminus subpopulation compared to the non-induced control (cut-off criteria fold change > 2, P-value < 0.05), respectively. Since their neighbouring genes were not affected, these intergenic regions may represent regulatory RNA molecules and a detailed analysis is under way. In addition supplementary table 2 lists all differentially expressed intergenic regions that were changed comparing the CSP induced samples with the non induced control (cut-off criteria fold change > ±1.8, P-value < 0.05).

**Induction of ComS and ComR was restricted to the competent fraction.**

In accordance with the recent finding that XIP(ComS)/ComR is the proximal regulator for *comX* induction, both genes showed the same pattern of enrichment and depletion in the two subpopulations (table 4) (26). Transcription of ComR has been shown to be 2-3 fold increased due to the presence of CSP, loss of *hdrM*, or overexpression of *hdrR* (30, 33), confirming the weak induction we observed in our transcriptome analysis. Since *comS* is a newly annotated open reading frame it was not directly included in our array design, but the intergenic region harbouring *comS* was 81 fold induced in the GFPplus fraction and only 6.6 fold in the GFPminus fraction.

**Transformasome genes were highly enriched in the competent subpopulation.**

The alternative sigma factor ComX is responsible for transcriptional activation of DNA uptake and processing genes, the so-called transformasome (8). As a consequence of *comX* enrichment in the GFPplus subpopulation, the gene transcripts responsible for building and assembling the transformasome were highly enriched (up to 164 fold) in the GFPplus fraction,
compared to the non induced mixed population. Due to the weak enrichment of comX transcripts in the GFPminus fraction, we observed a weak induction of ComX downstream genes in this fraction, too. However, by comparing the ratio of both separated subpopulations, we obtained a transcriptional enrichment of up to ~10 fold for the ComX downstream target genes, demonstrating the success of the applied approaches.

Table 4 lists the genes for DNA uptake and processing based on their orthologues in S. pneumoniae. Recently Okinaga et al., identified a gene (SMU.836), which is not part of the transformation machinery of S. pneumoniae, but was shown to be essential for DNA uptake in S. mutans (30). This gene was also enriched in the competent population. Beside other DNA processing proteins, the five genes that are essential for the processing of transformed DNA in S. pneumoniae (CoiA, DprA, RecA, SsbB and RadA) followed the same pattern of enrichment and depletion. (radA [SMU.327] was only slightly induced [1.7 fold, GFPplus] and thus did not fulfil the criteria of the microarray data analysis).

**Bacteriocin related genes were similarly transcribed in both subpopulations.**

In S. mutans, bacteriocin production is connected with competence development. The (putative) non-lantibiotic bacteriocins SMU.423, SMU 1906c, mutacin IV (nlmAB), CipB (mutacin V) and its cognate immunity protein (CipI) (33), were shown to be directly regulated by ComE, due to the presence of a ComE binding motif in the promoter regions (17, 41). Since comDE was enriched in the GFPplus subpopulation, it would have been expected that its direct target genes would have been enriched as well. Table 4 lists the putative bacteriocin related genes (according to the oralgen database, http://www.oralgen.lanl.gov/), that were induced by CSP in our study. Surprisingly, in contrast to the transformasome and DNA processing genes, there was no clear enrichment of bacteriocins in the competent fraction. The ratio of both subpopulations shows that there was only little enrichment for the mutacin V immunity protein CipI (cipI; SMU.925). By contrast, the self-acting mutacin CipB (mutacin
V), as well as the other bacteriocin related genes were equally induced in both subpopulations. To increase the sensitivity of detection, we used qPCR to determine the transcriptional level and confirmed the data for *nlmA*, *cipB* and *cipI* (supplementary table 1). This finding was surprising to us, since the majority of these genes are directly controlled by ComE because of the presence of a ComE binding motif. Due to the enrichment of *comE* transcripts, one would expect a similar pattern for its bacteriocin target genes. Addition of CSP reduces the growth rate of *S. mutans*. Perry *et al.* demonstrated that this decrease in growth was due to autolysis caused by the action of CipB. Furthermore, they showed that the majority of cells expressing ComX (visualized with a GFP reporter) were susceptible for propidium iodide staining, presumably due to an increased expression of CipB or to an imbalance of CipB/CipI (33). According to their observations, the results from our transcriptional analysis were unexpected. We did not perform a time series experiment to determine the transcriptional pattern of both genes. But at least for the time point under the condition analysed here, it seems very unlikely that cells undergoing competence development triggered their cell death at the same time.

**CSP induced autolysis is correlated with weak ComX expression.**

To evaluate the correlation of ComX expression and PI staining in more depth, we counterstained CSP induced cells (SMGFPComX) with PI and visualized them using fluorescence microscopy. The overlay image in figure 5 represents a typical experiment. When we first analysed the fluorescence images we did not see a correlation between green and red cells, which confirmed our transcriptome data. However, when we increased the green fluorescence intensity for the whole image, weakly fluorescing green cells became visible. The important observation was, that the majority of PI stained cells was weakly fluorescing green and thus expressed GFP to a small extent. In addition to the data shown in figure 5, we tested four
different conditions: addition of 0.2 µM or 2 µM CSP, immediately and one hour after inoculation, respectively. Under all tested conditions, we did observe this correlation of weak GFP and PI fluorescence (data not shown). Several large field images can be found in the supplementary data (supplementary table 3).

Autolysis through CipB is presumably mediated through an overexpression of CipB or to an imbalance of CipB/CipI. If some cells produce CipB to a greater extent, one would expect a bimodal distribution of cipB expression when single cells are analysed. Using flow cytometry, we observed an unimodal distribution of CipB expression after CSP addition (Fig. 6b).

Consequently, autolysis must occur due to a different amount of the cognate immunity protein CipI. This conclusion is supported by our transcriptome data which show an enrichment of CipI in the ComX induced subpopulation but an equal expression of CipB in both subpopulations. This is in full accordance with our microscopic observations, showing that moderate to strong GFP expression did not correlate with cell death, because these cells were protected by CipI. The data suggest the presence of an additional subpopulation among the ComX induced cells. During competence development of the CSP responsive subpopulation, some cells were apparently selected to undergo autolysis. This would explain the small amount of GFP because these cells already started the competence cascade and were expressing ComX/GFP. The possibility that some of the cells expressing GFP only weakly were collected in the GFPminus fraction would represent an additional explanation for the weak induction of the competence genes in this fraction.

Fluorescence microscopy confirmed that CipB is expressed in all cells but ComE only in a subpopulation.

Analysis of the subpopulation specific transcription data had indicated that phenotypic variation occurred upstream of ComX, presumably at the step of ComE induction. To confirm these results, we constructed GFP reporter strains for comE and cipB (mutacin V), which
provide a tool to measure protein synthesis rather than gene transcription. Fluorescence microscopy confirmed the transcriptional data, showing that ComE-GFP was only expressed in a subpopulation, whereas CipB-GFP was induced in all cells (Fig. 6a). Without addition of CSP, no GFP fluorescence could be observed, showing that the synthesis of CipB and ComE under these conditions was due to CSP addition, not to other stimuli. To confirm that cells expressing ComE-GFP or ComX-GFP were actually competent, we visualized the uptake of Cy3-labelled DNA. As expected, only cells expressing GFP were able to internalise DNA (yellow cells in Fig. 6a), thus confirming their competent state.

**Synthesis of GFP from the cipB promoter was weak compared to that of comE and comX.**

During microscopy, we observed that cells carrying pALSM04 (PcomX-GFP) and pALSM34a (PcomE-GFP) produced more GFP than cells carrying pALSM28 (PcipB-GFP). To further confirm this observation, we analysed cells of CSP induced cultures of these strains by flow cytometry.

For measuring the GFP intensity, we set the gate for GFPplus > 400 fluorescence units, to obtain the smallest overlap between low amounts of GFP and high autofluorescence. As expected, all cells expressing CipB-GFP exhibited higher fluorescence upon induction by CSP and shifted to the right on the x-axis (Fig. 6b). By contrast, only a part of the total population of the ComE-GFP and ComX-GFP cells, respectively, showed an increased fluorescence and shifted to the right upon CSP addition. The median intensities for the gated cells (>400 units) were 734 ± 75 for CipB-GFP, 911 ± 33 for ComE-GFP, and 1328 ± 196 for ComX-GFP, confirming the microscopic observation that ComE-GFP and ComX-GFP produced higher amounts of GFP than CipB-GFP and that these genes are exclusively induced in a subpopulation.
The finding that relatively little GFP protein was produced from the $cipB$-$gfp$ mRNA was surprising considering the relative amounts of transcripts for each gene. Figure 6c shows the relative cDNA abundance, determined by qPCR for both biological replicates of non-induced cells and the two induced subpopulations separated by flow cytometry. The relative amount of $cipB$ cDNA was ~385 fold higher than that of $gyrA$ (in both subpopulations). The level of $comE$ in the GFPplus fraction was approximately 18 fold compared to that of $gyrA$ and that of $comX$ was even lower. Therefore, the strength of the $comE$ promoter under these conditions was apparently moderate, and that of $cipB$ was very high. It is worth mentioning, that the amount of $cipB$ mRNA/cDNA was also very high in the absence of CSP.

Based on the high level of transcription of $cipB$, one would have expected that the $cipB$ reporter strain would produce more GFP than the other two. For each reporter strain, the coding sequence of each gene was replaced by GFP, but promoter sequence and ribosomal binding site remained unchanged (except for an addition of ATC in front of the start codon). Translational efficiency of any mRNA is highly dependent on the nucleotide composition in the translation initiation region (30 bp upstream of the start codon), which determines mRNA conformation and ribosome affinity. Apparently the transcript of $cipB$ had a low translational efficiency compared to that of $comE$ and $comX$ in our reporter construct, but we cannot currently infer the translational strength of the native $cipB$ mRNA. Alternatively, posttranscriptional regulatory events could have taken place, e.g. mediated by small regulatory RNAs.
Discussion

Competence development in *S. mutans* is a bistable system and results in intra-population diversity. The analysis of differential gene expression to obtain detailed insights into the different regulation of the subpopulations requires the use of single cell and RNA techniques. Kreth *et al.* and Perry *et al.* for the first time used GFP reporter strains to visualize gene expression on the single cell level in *S. mutans* to study competence development (18, 33). However, microarray analysis was averaged over the entire population.

To our knowledge, the combination of flow cytometry and transcriptome analysis of the separated subpopulations has never been done in bacteria. Many efforts have been carried out to improve the existing fluorescence proteins but a remaining problem for coupling reporter gene expression with cellular transcriptome analysis is the delay between synthesis of mRNA, translation and subsequent chromophore maturation. For the GFP used here, it was shown that under control of the arabinose utilizing system at full de-repression the onset of detectable fluorescence was ~ 16 ± 2.5 minutes post induction. However, with decreasing inducer concentration the time until fluorescence became detectable increased up to ~34 ± 10 min (27).

Thus, the time course of CSP mediated competence development in *S. mutans* allowed the combination of both techniques. Competence development in *S. mutans* is a slow process, since highest levels of mRNA for the key regulators *comE* and *comX* are obtained approximately two hours post induction and competence is maintained for another two hours, while competence in *S. pneumoniae* is only observed in a time window of about 40 min after CSP induction (42). Thus, there is enough time for GFP maturation in *S. mutans*. We demonstrate here the applicability and the success of combining both approaches to compare the differential gene expression of a competent and non-competent fraction derived from the
same population, since mRNA levels for the key regulators ComDE, ComR, ComS ComX and the transformasome genes were highly enriched in the GFPplus fraction. We show that all bacteriocin related genes were similarly expressed in both populations upon the addition of CSP, particularly CipB (mutacin V). Perry and co-workers demonstrated that CipB is a self-acting bacteriocin and that the protein CipI confers immunity against it. Moreover, they showed that the majority of the ComX expressing cells underwent autolysis (33). By contrast, we did not see a correlation of strong ComX expression and autolysis. But under all tested conditions we found that autolysis was indeed correlated with ComX expression but these cells did not produce GFP to a great amount. Perry and co-workers showed that the immunity protein CipI is additionally regulated by the LiaSR two component system in a cell density dependent manner. Furthermore, they demonstrated that the time of CSP addition to planktonically growing cells was very important, because it influenced the amount of cells that underwent autolysis (32). Different experimental conditions may explain that they observed that nearly all CSP responsive cells committed autolysis. Nevertheless, our data strongly indicate that the CSP responsive subpopulation itself bifurcates into two subpopulations: one that triggers autolysis and another one that develops the full competence response. The altruistic behaviour of autolysis would not be favoured by evolution without providing benefits to the remaining bacterial population. One advantage of autolysis is the release of DNA which provides building material for the biofilm matrix and thus increases the ability to form biofilms, which has already been shown for S. mutans (32, 35). Competence is measured by determining the number of cells that integrate added DNA into the chromosome. In S. mutans a transformation frequency of not more than 1-2% can be obtained, regardless of the employed conditions or the source of donor DNA. We show here that the amount of ComX expressing cells in the total population was roughly 30-50%. Our transcriptome data indicate that the DNA uptake machinery in these cells was highly expressed but nevertheless
transformation efficiencies in these dimensions could not be achieved. Possibly, internalised DNA was not used for recombination but for repair as it is the case for *S. pneumoniae* (9) or as a source of nutrients. The latter possibility seems unlikely because natural competence in *S. mutans* develops in the exponential growth phase, where nutrients are not exhausted. Another role for the DNA uptake machinery apart from internalising DNA has been demonstrated by Petersen and co-workers (35). They showed that the pilin-like proteins of the DNA uptake protein complex, which are evolutionarily related to type IV pili, are necessary for biofilm formation.

Similar observations were also reported from other bacteria, e.g. *Pseudomonas aeruginosa*, where type IV pili determined the shape of the biofilm architecture (3). The transformasome protein ComYB of *S. mutans* is needed for correct assembly of the pilin-like proteins. Deletion of ComYB not only resulted in decreased DNA uptake of 88%, also the formation of the biofilm was severely affected since the remaining components of the multi protein complex could not be assembled correctly (35).

The two component system ComDE represents the major signal transduction system for CSP mediated competence development in *S. mutans*. Deletion of either *comD* or *comE* resulted in no competence induction through the addition of CSP, concluding that activation and induction of ComDE is the first step in CSP mediated competence development (1). The CSP signal is further transmitted to the ComR regulator, which directly activates expression of ComX/SigX (26). Our study demonstrates for the first time that bifurcation into competent and non-competent cells occurred before induction of ComX, since the ComE-GFP reporter was expressed only in a subpopulation. It has to remain open which mechanism is responsible for the bifurcation of the population at the level of ComE. It is not known how the transcription of *comE* is regulated, except that it is induced by CSP. The action of ComR can be precluded due to the lack of the highly conserved consensus ComR recognition site in the promoter region of *comE*. Similarly, ComX can be excluded, since there is no binding motif
for ComX in the promoter region of comE and knock-out of ComX does not affect CSP induced comE expression (33).

Adapted from Mashburn-Warren et al. (26), we present a model that summarizes the current knowledge of CSP mediated competence development (Fig. 7) and is described below. Before the natural accumulation or exogenous addition of CSP the ComDE system is expressed at a basal level but is not activated. Above the threshold of CSP, ComE molecules become phosphorylated and activate expression of the bacteriocin related genes. We assume that due to the presence of the ComE binding box in their promoter regions, a low amount of activated ComE protein is sufficient to drive full expression. Because ComDE is present in all cells at a low concentration, bacteriocins are expressed in the whole population at a high rate. CSP mediated induction of comDE transcription, which occurs only in a subpopulation, may require an unknown factor, possibly in combination with population wide variability in the expression of comE. Due to the lack of the known ComE binding site in the comE promoter region, the presence of a currently unknown factor that is needed for comE induction is likely. We suggest that competence initiating cells are selected by the noise in ComE expression and that cells exceeding a certain threshold activate a positive feedback by interaction with an unknown factor. A similar mechanism has been found to control competence in B. subtilis, except that ComK directly activates a positive autoregulatory loop (12). We show here that despite a moderate amount of comE transcripts a high number of GFP proteins was obtained from the comE-gfp promoter fusion. Ozbudak et al. demonstrated that a high translational efficiency coupled to a low transcriptional rate resulted in the highest amount of noise (31). According to their observations and calculations, large noise in ComE expression would have to be expected supporting the hypothesis that cells might have been selected due to noise in ComE expression. We suggest that two levels of ComE expression exist in the whole population. Expression below threshold induces bacteriocin synthesis, while above threshold expression triggers
induction of \textit{comDE} and \textit{comR} and selects the cells for subsequent activation of ComX. We further suggest that cells that establish a positive feedback for ComE activate (directly or indirectly) ComR, which then induces the expression of \textit{comS} and \textit{comX} and subsequent expression of the late competence genes.

Our data indicate a second bifurcation, inducing autolysis in some cells through the action of CipB. According to the studies of Perry \textit{et al.}, the immunity determinant CipI regulated by LiaR was shown to dictate the cell fate of autolysis (32, 33), which is in accordance with the enrichment of \textit{cipI} transcripts in the ComX expressing subpopulation. Beside its autolytic properties CipB is a potent mutacin, targeting mainly non streptococcal species (14). Thus, CSP mediated competence initiation liberates DNA from competitors in the oral cavity, providing a source either for genetic diversity, or for replenishment of the nucleotide pool that can be used for DNA repair. Finally the liberated DNA represents a building material for the biofilm matrix. The additional DNA binding capacity of the pseudo type IV pili of the DNA uptake complex further promotes the formation of biofilm, encasing non-CSP responsive cells as well. In case that surrounding competitors are not sensitive to the released mutacins, autolysis represents a safe mechanism for liberating DNA that can be used as biofilm building material. Thus the phenotypic diversity and the altruistic behaviour of lytic cells provides benefit to the whole population.

In summary, we performed a subpopulation specific transcriptome analysis comparing CSP induced competent and non-competent cells. The results show that flow cytometry and subsequent transcriptome analysis of \textit{S. mutans} cells were successfully combined, resulting in a strong enrichment of mRNA of \textit{comX} and its downstream targets, e.g. the transformasome related genes. We demonstrated for the first time, that bifurcation into competent and non-competent cells occurred at the level of ComE and not ComX, but that expression of bacteriocin related genes occurred similarly in all cells, particular the self acting mutacin CipB. We suggest that due to noise in the basal level of ComE expression some cells exceed a
threshold of CSP activated ComE protein molecules, enabling a positive feedback loop and
entering the competent state. Moreover, our data indicate a second bifurcation within the
competent fraction, where some cells undergo autolysis presumably due to a decreased
amount of CipI. Intra population diversity during competence development apparently has
several advantages. It enables \textit{S. mutans} to cope with competitors, produces membrane
embedded DNA binding proteins, and liberates DNA that can be used for genetic diversity,
repair or as biofilm building material, and thus reflects its successful adaptation to its
ecological niche.
Acknowledgements

We would like to thank Andreas Podbielski for providing the pFW5 plasmid.


### Table 1: Strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strain/plasmid</th>
<th>Relevant characteristics</th>
<th>Reference/source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH10β</td>
<td>General cloning strain</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td><em>S. mutans</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UA159</td>
<td>Wild-type, Erm&lt;sup&gt;s&lt;/sup&gt;, Sp&lt;sup&gt;s&lt;/sup&gt;</td>
<td>ATCC 700610</td>
</tr>
<tr>
<td>SMLuccomE</td>
<td>UA159::φ(comE&lt;sub&gt;P-luc&lt;/sub&gt;), Sp&lt;sup&gt;r&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>SMLuccomX</td>
<td>UA159::φ(comX&lt;sub&gt;P-luc&lt;/sub&gt;), Sp&lt;sup&gt;r&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>SMGFPcomX</td>
<td>UA159, carrying pALSM04</td>
<td>This study</td>
</tr>
<tr>
<td>SMGFPcomE</td>
<td>UA159, carrying pALSM28</td>
<td>This study</td>
</tr>
<tr>
<td>SMGFPcipB</td>
<td>UA159, carrying pALSM34a</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKRC12</td>
<td>Donor of eGFP</td>
<td>K. Riedel (38)</td>
</tr>
<tr>
<td>pFW5</td>
<td>Suicide vector, Sp&lt;sup&gt;r&lt;/sup&gt;,</td>
<td>A. Podbielski (36)</td>
</tr>
<tr>
<td>pALEC15</td>
<td>Derivate pFW5, luc, Sp&lt;sup&gt;r&lt;/sup&gt;</td>
<td>A. Lemme (21)</td>
</tr>
<tr>
<td>pALEC50</td>
<td>pALEC15 + φ(comE&lt;sub&gt;P-luc&lt;/sub&gt;), Sp&lt;sup&gt;r&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pALEC52</td>
<td>pALEC15 + φ(comX&lt;sub&gt;P-luc&lt;/sub&gt;), Sp&lt;sup&gt;r&lt;/sup&gt;</td>
<td>B. Kunze (20)</td>
</tr>
<tr>
<td>pALSM01</td>
<td>modified pAT18, Erm&lt;sup&gt;r&lt;/sup&gt;</td>
<td>P. Trieu-Cuot</td>
</tr>
<tr>
<td>pALSM02</td>
<td>pALSM01, eGFP, Erm&lt;sup&gt;r&lt;/sup&gt;</td>
<td>I.M. Nakagawa (28, 40)</td>
</tr>
<tr>
<td>pALSM03</td>
<td>pALSM03, Removal of lac&lt;sub&gt;P&lt;/sub&gt;, Erm&lt;sup&gt;r&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pALSM04</td>
<td>pALSM03 + φ(comX&lt;sub&gt;P-eGFP&lt;/sub&gt;), Erm&lt;sup&gt;r&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>PALSM28</td>
<td>pALSM03 + φ(cipB&lt;sub&gt;P-eGFP&lt;/sub&gt;), Erm&lt;sup&gt;r&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>PALSM34a</td>
<td>pALSM03 + φ(comE&lt;sub&gt;P-eGFP&lt;/sub&gt;), Erm&lt;sup&gt;r&lt;/sup&gt;</td>
<td>This study</td>
</tr>
</tbody>
</table>

Erm, erythromycin; Sp, spectinomycin; luc, luciferase
Table 2: Primers used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'→3')</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>PcomEF</td>
<td>TTTTTTCCATGGGAGGTTGCATCGATCTTCCT</td>
<td>Construction pALEC50</td>
</tr>
<tr>
<td>PcomER</td>
<td>TTTTTTCCATGGTATTTTCTCCTTTTAATCTTCTAT</td>
<td>Construction pALEC50</td>
</tr>
<tr>
<td>PcomXF</td>
<td>AAAAAACCAATGGTCCAAAAATAAGTGACTAAGG</td>
<td>Construction pALEC52</td>
</tr>
<tr>
<td>PcomXR</td>
<td>AAAAAACCAATGGCTATTAGATGACCTTTTTTTCCT</td>
<td>Construction pALEC52</td>
</tr>
<tr>
<td>GFPF</td>
<td>AAAAAATCTAGAAGATTATCTAGGTTAAGGAGAAGACTTTTCATTG</td>
<td>eGFP</td>
</tr>
<tr>
<td>GFPR</td>
<td>AAAAAACCATAGGCCTATTGTATAGTCTCCTATTCCATGCCCATG</td>
<td>eGFP</td>
</tr>
<tr>
<td>PcomX2F</td>
<td>TCCAAAAATAAGTGACTAAGG</td>
<td>Promoter comX</td>
</tr>
<tr>
<td>PcomX2R</td>
<td>CTATTAGATGACCTCTTTT</td>
<td>Promoter comX</td>
</tr>
<tr>
<td>PcipBF</td>
<td>TTTATATCTCCTTTTTTTTGGATTATAATTAGTATACTA</td>
<td>Promoter cipB</td>
</tr>
<tr>
<td>PcipBR</td>
<td>ATGATAAATACCCCTCTCCCACCTTTTAGGT</td>
<td>Promoter cipB</td>
</tr>
<tr>
<td>PcomE2F</td>
<td>TTATATACATTTGGACACAGGCTGAT</td>
<td>Promoter comE</td>
</tr>
<tr>
<td>PcomE2R</td>
<td>TTTTTCCTTTAATCTTCTAT</td>
<td>Promoter comE</td>
</tr>
</tbody>
</table>

* a restriction sites in bold
Table 3: Fold change of gene expression compared to time zero in response to 0.2 µM CSP determined by quantitative RT-PCR.

<table>
<thead>
<tr>
<th>gene</th>
<th>15 min</th>
<th>30 min</th>
<th>60 min</th>
<th>120 min</th>
<th>240 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>comE</td>
<td>1.75 ± 0.14</td>
<td>2.91 ± 0.35</td>
<td>8.54 ± 1.28</td>
<td>18.58 ± 2.23</td>
<td>1.98 ± 0.22</td>
</tr>
<tr>
<td>comR</td>
<td>1.38 ± 0.12</td>
<td>2.68 ± 0.27</td>
<td>2.24 ± 0.29</td>
<td>2.97 ± 0.48</td>
<td>0.97 ± 0.05</td>
</tr>
<tr>
<td>comS</td>
<td>2.17 ± 0.13</td>
<td>20.88 ± 2.71</td>
<td>68.3 ± 12.98</td>
<td>144.08 ± 30.26</td>
<td>8.97 ± 1.17</td>
</tr>
<tr>
<td>comX</td>
<td>0.88 ± 0.03</td>
<td>1.99 ± 0.12</td>
<td>5.29 ± 0.58</td>
<td>12.48 ± 1.75</td>
<td>0.15 ± 0.02</td>
</tr>
<tr>
<td>cipB</td>
<td>27.49 ± 3.02</td>
<td>33.86 ± 4.74</td>
<td>34.91 ± 5.24</td>
<td>41.02 ± 6.56</td>
<td>94.99 ± 16.15</td>
</tr>
<tr>
<td>cipI</td>
<td>8.45 ± 0.59</td>
<td>11.44 ± 1.94</td>
<td>10.83 ± 1.63</td>
<td>15.52 ± 2.48</td>
<td>14.22 ± 1.99</td>
</tr>
</tbody>
</table>

Average gene expression and standard deviation of two biological samples.
Table 4: Selected genes that were significantly changed (log2 fold change > 0.85 and P-value < 0.05) after cytometric sorting.

<table>
<thead>
<tr>
<th>Locus Tag SMU.#</th>
<th>Description</th>
<th>Gene symbol</th>
<th>Fold Ind. mixed</th>
<th>Fold GFP-</th>
<th>P-Value Induced mixed</th>
<th>Fold GFP-</th>
<th>P-Value GF+</th>
<th>Fold GF+</th>
<th>P-Value GF+/GFP-</th>
<th>P-Value GF+/GFP-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulatory functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>ComR</td>
<td>comR</td>
<td>1.4</td>
<td>1.70E-02</td>
<td>1.1</td>
<td>2.27E-01</td>
<td>1.8</td>
<td>1.31E-03</td>
<td>1.4</td>
<td>4.9E-01</td>
</tr>
<tr>
<td>Shp61</td>
<td>ComS within intergenic region 51</td>
<td>comS</td>
<td>15.37</td>
<td>1.11E-04</td>
<td>6.59</td>
<td>5.86E-04</td>
<td>81.03</td>
<td>1.26E-05</td>
<td>11.76</td>
<td>7.48E-04</td>
</tr>
<tr>
<td>1916</td>
<td>histidine kinase of the competence regulon, ComD</td>
<td>comD</td>
<td>3.95</td>
<td>9.95E-04</td>
<td>2.62</td>
<td>4.45E-03</td>
<td>6.99</td>
<td>2.13E-04</td>
<td>2.09</td>
<td>2.36E-02</td>
</tr>
<tr>
<td>1917</td>
<td>response regulator of the competence regulon, ComE</td>
<td>comE</td>
<td>3.87</td>
<td>4.23E-04</td>
<td>2.54</td>
<td>2.16E-03</td>
<td>7.31</td>
<td>7.47E-05</td>
<td>2.31</td>
<td>7.37E-03</td>
</tr>
<tr>
<td>1997</td>
<td>ComX1, transcriptional regulator of competence-specific genes</td>
<td>comX1</td>
<td>3.56</td>
<td>2.58E-02</td>
<td>2.01</td>
<td>1.40E-01</td>
<td>13.79</td>
<td>1.50E-03</td>
<td>4.96</td>
<td>2.48E-02</td>
</tr>
<tr>
<td>Transformsome related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>498</td>
<td>late competence protein</td>
<td>comF</td>
<td>9.37</td>
<td>3.86E-04</td>
<td>3.60</td>
<td>4.25E-03</td>
<td>31.81</td>
<td>5.38E-05</td>
<td>6.93</td>
<td>2.28E-03</td>
</tr>
<tr>
<td>499</td>
<td>late competence protein</td>
<td>-</td>
<td>12.60</td>
<td>3.65E-04</td>
<td>4.82</td>
<td>2.88E-03</td>
<td>47.99</td>
<td>5.36E-05</td>
<td>8.35</td>
<td>8.07E-03</td>
</tr>
<tr>
<td>539c</td>
<td>signal peptidase type IV</td>
<td>-</td>
<td>2.85</td>
<td>4.14E-03</td>
<td>1.77</td>
<td>3.91E-02</td>
<td>8.75</td>
<td>1.75E-04</td>
<td>4.32</td>
<td>3.73E-03</td>
</tr>
<tr>
<td>625</td>
<td>competence protein</td>
<td>comEA</td>
<td>20.29</td>
<td>2.10E-04</td>
<td>7.12</td>
<td>1.38E-03</td>
<td>70.87</td>
<td>4.32E-05</td>
<td>8.26</td>
<td>4.83E-03</td>
</tr>
<tr>
<td>626</td>
<td>competence protein</td>
<td>-</td>
<td>18.62</td>
<td>1.64E-04</td>
<td>6.64</td>
<td>1.12E-03</td>
<td>55.17</td>
<td>3.88E-05</td>
<td>6.63</td>
<td>5.23E-03</td>
</tr>
<tr>
<td>836</td>
<td>hypothetical protein</td>
<td>-</td>
<td>9.93</td>
<td>7.66E-04</td>
<td>3.83</td>
<td>7.22E-03</td>
<td>34.45</td>
<td>1.11E-04</td>
<td>6.61</td>
<td>3.26E-03</td>
</tr>
<tr>
<td>1979c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>16.28</td>
<td>1.38E-04</td>
<td>5.69</td>
<td>1.13E-03</td>
<td>44.87</td>
<td>3.35E-05</td>
<td>6.83</td>
<td>3.61E-03</td>
</tr>
<tr>
<td>1980c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>37.48</td>
<td>1.92E-05</td>
<td>11.83</td>
<td>1.11E-04</td>
<td>111.95</td>
<td>5.66E-06</td>
<td>8.30</td>
<td>1.50E-03</td>
</tr>
<tr>
<td>1981c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>41.42</td>
<td>1.52E-05</td>
<td>13.06</td>
<td>8.33E-05</td>
<td>134.85</td>
<td>4.25E-06</td>
<td>8.47</td>
<td>7.14E-05</td>
</tr>
<tr>
<td>1982c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>35.24</td>
<td>1.53E-05</td>
<td>11.54</td>
<td>8.57E-05</td>
<td>116.79</td>
<td>3.99E-06</td>
<td>8.52</td>
<td>6.48E-04</td>
</tr>
<tr>
<td>1983</td>
<td>putative competence protein ComYD</td>
<td>comYD</td>
<td>23.54</td>
<td>1.14E-04</td>
<td>7.52</td>
<td>8.42E-04</td>
<td>74.40</td>
<td>2.75E-05</td>
<td>8.08</td>
<td>3.90E-03</td>
</tr>
<tr>
<td>1984</td>
<td>putative competence protein ComYC</td>
<td>comYC</td>
<td>40.28</td>
<td>1.58E-05</td>
<td>12.21</td>
<td>9.46E-05</td>
<td>140.37</td>
<td>4.12E-06</td>
<td>9.17</td>
<td>5.54E-04</td>
</tr>
<tr>
<td>1985</td>
<td>ABC transporter ComYB</td>
<td>comYB</td>
<td>30.59</td>
<td>2.92E-05</td>
<td>9.49</td>
<td>1.97E-04</td>
<td>91.35</td>
<td>8.13E-06</td>
<td>7.93</td>
<td>5.71E-04</td>
</tr>
<tr>
<td>1987</td>
<td>ComYA; late competence gene</td>
<td>comYA</td>
<td>44.78</td>
<td>1.51E-05</td>
<td>13.28</td>
<td>8.81E-05</td>
<td>164.36</td>
<td>3.86E-06</td>
<td>9.63</td>
<td>7.10E-04</td>
</tr>
<tr>
<td>Putative Bacteriocins, Immunity Proteins, Bacteriophage Island</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>hypothetical protein</td>
<td>nlmA</td>
<td>6.53</td>
<td>3.84E-04</td>
<td>5.65</td>
<td>5.49E-04</td>
<td>7.29</td>
<td>2.98E-04</td>
<td>1.10</td>
<td>7.16E-01</td>
</tr>
<tr>
<td>151</td>
<td>hypothetical protein</td>
<td>nlmB</td>
<td>6.45</td>
<td>5.51E-04</td>
<td>5.67</td>
<td>7.56E-04</td>
<td>7.01</td>
<td>4.55E-04</td>
<td>1.04</td>
<td>8.81E-01</td>
</tr>
<tr>
<td>423</td>
<td>hypothetical protein</td>
<td>-</td>
<td>6.66</td>
<td>2.34E-04</td>
<td>5.92</td>
<td>3.11E-04</td>
<td>7.25</td>
<td>1.92E-04</td>
<td>1.04</td>
<td>8.58E-01</td>
</tr>
<tr>
<td>925</td>
<td>hypothetical protein</td>
<td>cipI</td>
<td>6.18</td>
<td>5.92E-05</td>
<td>5.20</td>
<td>9.33E-05</td>
<td>8.57</td>
<td>2.66E-05</td>
<td>1.38</td>
<td>1.10E-01</td>
</tr>
<tr>
<td>1892c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>0.87</td>
<td>6.18E-01</td>
<td>0.55</td>
<td>7.56E-02</td>
<td>0.46</td>
<td>3.52E-02</td>
<td>0.71</td>
<td>1.82E-01</td>
</tr>
<tr>
<td>1902c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>2.83</td>
<td>7.64E-04</td>
<td>2.80</td>
<td>8.09E-04</td>
<td>3.13</td>
<td>5.09E-04</td>
<td>0.96</td>
<td>8.32E-01</td>
</tr>
<tr>
<td>1903c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>4.90</td>
<td>1.04E-04</td>
<td>4.83</td>
<td>1.08E-04</td>
<td>5.69</td>
<td>6.90E-05</td>
<td>1.00</td>
<td>9.90E-01</td>
</tr>
<tr>
<td>1904c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>4.52</td>
<td>3.74E-04</td>
<td>4.35</td>
<td>4.19E-04</td>
<td>5.04</td>
<td>2.73E-04</td>
<td>0.99</td>
<td>9.47E-01</td>
</tr>
<tr>
<td>1905c</td>
<td>putative bacteriocin secretion protein</td>
<td>-</td>
<td>4.83</td>
<td>2.74E-04</td>
<td>4.75</td>
<td>2.86E-04</td>
<td>5.40</td>
<td>2.01E-04</td>
<td>0.96</td>
<td>8.53E-01</td>
</tr>
<tr>
<td>1906c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>6.63</td>
<td>1.91E-04</td>
<td>6.47</td>
<td>2.03E-04</td>
<td>6.37</td>
<td>2.10E-04</td>
<td>0.84</td>
<td>5.51E-01</td>
</tr>
<tr>
<td>1908c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>6.38</td>
<td>1.91E-04</td>
<td>6.13</td>
<td>2.11E-04</td>
<td>6.40</td>
<td>1.90E-04</td>
<td>0.88</td>
<td>5.48E-01</td>
</tr>
<tr>
<td>1909c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>6.23</td>
<td>2.40E-04</td>
<td>5.93</td>
<td>2.72E-04</td>
<td>6.03</td>
<td>2.61E-04</td>
<td>0.88</td>
<td>6.32E-01</td>
</tr>
<tr>
<td>1910c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>6.78</td>
<td>2.47E-04</td>
<td>6.60</td>
<td>2.64E-04</td>
<td>6.60</td>
<td>2.64E-04</td>
<td>0.87</td>
<td>6.56E-01</td>
</tr>
<tr>
<td>1912c</td>
<td>hypothetical protein</td>
<td>-</td>
<td>6.07</td>
<td>1.52E-04</td>
<td>5.61</td>
<td>1.86E-04</td>
<td>6.06</td>
<td>1.53E-04</td>
<td>0.92</td>
<td>6.98E-01</td>
</tr>
<tr>
<td>1913c</td>
<td>putative immunity protein</td>
<td>-</td>
<td>5.69</td>
<td>2.78E-04</td>
<td>5.49</td>
<td>3.06E-04</td>
<td>5.86</td>
<td>2.58E-04</td>
<td>0.89</td>
<td>6.10E-01</td>
</tr>
</tbody>
</table>
whose adjacent genes were not affected.

IGR that are significantly changed (log2 fold change > 0.85 and P-value < 0.05) are in bold

<table>
<thead>
<tr>
<th>Genomic Position</th>
<th>IGR between ORFs (SMU.1+)</th>
<th>Fold Induced mixed</th>
<th>P-Value Induced mixed</th>
<th>Fold GFP-</th>
<th>P-Value GFP-</th>
<th>Fold GFP+ vs. GFP-</th>
<th>Fold GFP+ vs. GFP-</th>
</tr>
</thead>
<tbody>
<tr>
<td>136886_137007</td>
<td>Hypothetical protein</td>
<td>132</td>
<td>133c</td>
<td>1.73</td>
<td>2.73E-01</td>
<td>1.28</td>
<td>6.00E-01</td>
</tr>
<tr>
<td>17965_179873</td>
<td>cipB</td>
<td>181</td>
<td>182</td>
<td>1.69</td>
<td>7.64E-02</td>
<td>1.97</td>
<td>3.50E-02</td>
</tr>
<tr>
<td>367013_367066</td>
<td>390</td>
<td>391c</td>
<td>1.68</td>
<td>3.33E-01</td>
<td>2.63</td>
<td>1.97</td>
<td>6.80E-02</td>
</tr>
<tr>
<td>375868_376028</td>
<td>404c</td>
<td>402</td>
<td>1.55</td>
<td>5.32E-02</td>
<td>2.46E-02</td>
<td>2.26</td>
<td>5.85E-03</td>
</tr>
<tr>
<td>435807_435921</td>
<td>466</td>
<td>467</td>
<td>1.59</td>
<td>2.69E-01</td>
<td>2.53</td>
<td>2.06</td>
<td>7.70E-02</td>
</tr>
<tr>
<td>596121_596312</td>
<td>636</td>
<td>637c</td>
<td>1.77</td>
<td>1.13E-01</td>
<td>2.15</td>
<td>2.19</td>
<td>4.74E-02</td>
</tr>
<tr>
<td>71015_71084</td>
<td>765</td>
<td>766</td>
<td>2.34</td>
<td>9.05E-02</td>
<td>2.99</td>
<td>2.43</td>
<td>1.46E-01</td>
</tr>
<tr>
<td>71867_719212</td>
<td>770c</td>
<td>771c</td>
<td>4.53</td>
<td>4.54E-03</td>
<td>3.03</td>
<td>1.51E-02</td>
<td>1.74E-04</td>
</tr>
<tr>
<td>745624_745833</td>
<td>799c</td>
<td>800</td>
<td>1.46</td>
<td>1.28E-01</td>
<td>2.14</td>
<td>1.52E-02</td>
<td>1.37E-01</td>
</tr>
<tr>
<td>755296_755491</td>
<td>807</td>
<td>809</td>
<td>2.42</td>
<td>1.11E-01</td>
<td>3.18</td>
<td>5.39E-02</td>
<td>3.61E-02</td>
</tr>
<tr>
<td>593942_759398</td>
<td>813</td>
<td>814</td>
<td>1.67</td>
<td>5.65E-02</td>
<td>2.08</td>
<td>1.68E-02</td>
<td>1.92E-02</td>
</tr>
<tr>
<td>837975_838127</td>
<td>883</td>
<td>885</td>
<td>3.64</td>
<td>3.47E-02</td>
<td>3.71</td>
<td>3.29E-02</td>
<td>1.86E-01</td>
</tr>
<tr>
<td>844419_844621</td>
<td>889</td>
<td>890</td>
<td>2.06</td>
<td>4.32E-02</td>
<td>2.50</td>
<td>1.95E-02</td>
<td>1.60E-01</td>
</tr>
<tr>
<td>845165_845295</td>
<td>890</td>
<td>891</td>
<td>1.67</td>
<td>5.46E-02</td>
<td>2.01</td>
<td>1.96E-02</td>
<td>1.48E-01</td>
</tr>
<tr>
<td>878575_878770</td>
<td>923</td>
<td>924</td>
<td>1.94</td>
<td>2.86E-02</td>
<td>2.59</td>
<td>7.54E-03</td>
<td>2.08E-02</td>
</tr>
<tr>
<td>896354_896469</td>
<td>944</td>
<td>946</td>
<td>1.66</td>
<td>1.18E-01</td>
<td>2.60</td>
<td>4.36E-02</td>
<td>1.43E-01</td>
</tr>
<tr>
<td>938355_938638</td>
<td>991</td>
<td>992</td>
<td>2.15</td>
<td>1.06E-01</td>
<td>4.15</td>
<td>1.54E-02</td>
<td>2.61E-02</td>
</tr>
<tr>
<td>996212_996313</td>
<td>1048</td>
<td>1050</td>
<td>2.13</td>
<td>6.85E-02</td>
<td>2.95</td>
<td>2.21E-02</td>
<td>1.91E-01</td>
</tr>
<tr>
<td>998766_999060</td>
<td>1052</td>
<td>1053</td>
<td>1.51</td>
<td>2.21E-01</td>
<td>2.25</td>
<td>4.24E-02</td>
<td>1.69E-01</td>
</tr>
<tr>
<td>1008295_1008566</td>
<td>1063</td>
<td>1064c</td>
<td>2.37</td>
<td>4.82E-02</td>
<td>3.70</td>
<td>1.14E-02</td>
<td>3.90E-03</td>
</tr>
<tr>
<td>1019708_1019801</td>
<td>1076</td>
<td>1077</td>
<td>2.03</td>
<td>5.05E-02</td>
<td>2.05</td>
<td>4.91E-02</td>
<td>1.56E-01</td>
</tr>
<tr>
<td>1021518_1021647</td>
<td>1077</td>
<td>1078c</td>
<td>1.99</td>
<td>3.50E-02</td>
<td>2.29</td>
<td>1.88E-02</td>
<td>2.00E-02</td>
</tr>
<tr>
<td>1138619_1138670</td>
<td>1195</td>
<td>1196c</td>
<td>2.31</td>
<td>7.34E-02</td>
<td>3.46</td>
<td>1.22E-02</td>
<td>3.40E-02</td>
</tr>
<tr>
<td>1139665_1139763</td>
<td>R0092</td>
<td>1197</td>
<td>2.50</td>
<td>1.02E-02</td>
<td>2.73</td>
<td>7.10E-03</td>
<td>2.73E-03</td>
</tr>
<tr>
<td>1307188_1307342</td>
<td>1378</td>
<td>1379</td>
<td>2.07</td>
<td>6.15E-02</td>
<td>2.20</td>
<td>4.77E-02</td>
<td>1.61E-01</td>
</tr>
<tr>
<td>1381695_1381764</td>
<td>1430</td>
<td>1451</td>
<td>1.76</td>
<td>1.39E-01</td>
<td>2.57</td>
<td>3.34E-02</td>
<td>2.95E-02</td>
</tr>
<tr>
<td>1759753_1759939</td>
<td>1865</td>
<td>1867c</td>
<td>2.13</td>
<td>6.08E-02</td>
<td>3.12</td>
<td>1.57E-02</td>
<td>1.82E-01</td>
</tr>
<tr>
<td>1889423_1889587</td>
<td>2016</td>
<td>2017</td>
<td>1.29</td>
<td>3.09E-01</td>
<td>2.04</td>
<td>2.78E-02</td>
<td>1.66E-02</td>
</tr>
<tr>
<td>1892786_1893019</td>
<td>2025</td>
<td>2026c</td>
<td>1.51</td>
<td>5.55E-02</td>
<td>1.40</td>
<td>9.90E-02</td>
<td>2.14E-02</td>
</tr>
<tr>
<td>1906075_1906218</td>
<td>2035</td>
<td>2036</td>
<td>1.74</td>
<td>1.98E-01</td>
<td>2.76</td>
<td>4.33E-02</td>
<td>2.20E-02</td>
</tr>
</tbody>
</table>

Table 5: Intergenic regions that were significantly changed after cytometric sorting and whose adjacent genes were not affected.

IGR that are significantly changed (fold change > 2.0 and P-value < 0.05) are in bold.
Figures

Figure 1. Current model of competence development in *S. mutans*.

Five different systems were shown to influence competence development: The serine protease HtrA, the HdrRM and BsrRM regulatory systems and the two component systems CiaHR and ComDE. The *comC* gene product is the precursor of the CSP (Competence Stimulating Peptide) peptide which is transported and processed via the CslAB (ComAB) transporter (not shown). Accumulation of CSP activates the histidine kinase ComD, which activates its response regulator ComE by phosphorylation. Activated ComE regulates the expression of several mutacin related genes, including the self-acting bacteriocin CipB and its cognate immunity protein CipI. The HdrMR and BsrMR regulatory systems were shown to regulate each other. Furthermore, they were shown to be involved in the regulation of bacteriocin regulated genes (not shown). The signals that are sensed by HtrA, CiaHR, HdrMR and
BsrMR are unknown. All signals, including CSP, are integrated by the transcriptional regulator ComR through yet unknown mechanisms. ComR activates the expression of comS. The gene product of comS is the precursor of the XIP (sigma factor X Inducing Peptide) peptide. ComS is transported into the extracellular environment through an unknown transporter and is processed outside of the cell (not shown). The mature XIP peptide is transported back by the peptide transporter Opp/Ami. ComR in conjunction with XIP activates the expression of ComS, leading to an auto-catalytic positive feedback, and of the alternate sigma factor ComX. ComX further activates gene expression from late competence promoters and thus drives the expression of DNA-uptake and processing genes leading to genetic competence. (Modified after Mashburn-Warren et al. 2010, Perry et al. 2009 and Xie et al. 2011 (26, 33, 44))
Figure 2. Transient induction of *comE* and *comX* by CSP.

Growth characteristics and luminescence development of the *comE* (A) and *comX* (B) promoter-luciferase fusion in the absence and presence of 0.2 µM CSP. CSP was added after one hour of incubation. Figures show the mean and standard deviation of at least two independent experiments.
Figure 3. Subpopulation specific induction of SMGFPComX in the presence of CSP.

Growth curves and development of GFP fluorescence (A). CSP (0.2 µM) was added 1h after inoculation. Figures show the mean and standard deviation of two independent experiments. Fluorescence microscopy of CSP induced (sonicated) cells 2h (B) and 6h (C) post induction.
**Figure 4.** Scatter plots and GFP fluorescence distribution of *S. mutans* cells analysed by flow cytometry.

A: Side- and forward scatter plots of uninoculated media, with the corresponding gates for cell sorting. B: Side- and forward scatter plots of inoculated media, with the corresponding gates for cell sorting.

C: GFP intensity from media and not induced cells of the gate P3. Left, intensity of uninoculated media; right, intensity of uninduced SMGFPComX cells.

D: GFP intensity of induced SMGFPComX cells of gate P3 and re-analysis after sorting.
Figure 5. Propidium iodide (PI) counterstaining of CSP induced SMGFPcomX cells.

Cells were 1:10 diluted in fresh media and incubated for one hour before addition of 0.2 µM (A) or 2 µM (B) CSP. Images were recorded two hours post induction. The large images left are overlays of phase contrast, GFP and PI fluorescence. The highlighted area represent cells for every phenotype and is magnified in the right: image 1, overlay of phase contrast and GFP; image 2, PI channel; image 3, increased brightness of the GFP channel; image 4, overlay of bright GFP and PI. White arrows mark cells with a very weak GFP expression that became visible, when the brightness of GFP was increased. Importantly, the majority of dark cells did not show enhanced green fluorescence after brightness modification, confirming that the weak green fluorescence was due to GFP expression and not to auto fluorescence. The majority of red cells showed an enhanced green fluorescence after increasing the brightness.
Figure 6. Single cell analysis of comE, cipB and comX expression.

A: Microscopic pictures of CSP induced cells of SMGFPcomE (left), SMGFPcipB (middle) and SMGFPcomX (right). Images of phase contrast, GFP (green, gene expression) and Cy3 (yellow, DNA uptake) fluorescence were merged. B: GFP intensity distribution from 100,000 cells of strains SMGFPcomE (left), SMGFPcipB (middle) and SMGFPcomX (right) determined by flow cytometry. The mean GFP intensity and standard deviation of two biological replicates of the GFP+ fraction is shown. C: Gene expression analysis with quantitative real time PCR using both subpopulations of FACS separated, CSP induced SMGFPcomX cells and the non-induced mixed population.
**Figure 7.** Working model for bifurcation steps during CSP mediated competence development in *S. mutans*.

The CSP pheromone activates the histidine kinase ComD (not shown) which phosphorylates its cognate response regulator ComE. Phosphorylated ComE induces population wide bacteriocin gene expression. Cells having a high level of ComE, probably due to noise in ComE expression, induce a positive feedback loop for ComE through an unknown mechanism and activate further ComR and the alternate sigma factor ComX. The population segregates into autolysing cells and cells that become competent due to a different expression of the immunity protein CipI. Finally, at least three different subpopulations have developed from the clonal culture due to the presence of a high CSP concentration.